Annals of Mathematics, 117 (1983), 553–607

A proof of the Smale Conjecture, $\mathrm{Diff}(S^3) \simeq O(4)$

By ALLEN E. HATCHER

The Smale Conjecture [9] is the assertion that the inclusion of the orthogonal group $O(4)$ into $\mathrm{Diff}(S^3)$, the diffeomorphism group of the 3-sphere with the C^∞ topology, is a homotopy equivalence. There are many equivalent forms of this conjecture, some of which are listed in the appendix to this paper. We shall prove the following one:

THEOREM. *A smooth family of C^∞ embeddings $g_t\colon S^2 \to \mathbf{R}^3$, $t \in S^k$, extends to $\bar{g}_t\colon B^3 \to \mathbf{R}^3$, a smooth family of C^∞ embeddings of the 3-ball, for any $k \geq 0$.*

The case $k = 0$ is a strong form of the Schoenflies theorem in dimension three, essentially due to Alexander [1], while the case $k = 1$ is Cerf's theorem $\pi_0 \mathrm{Diff}(S^3) \approx \pi_0 O(4)$ [2] (which has the important corollary $\Gamma_4 = 0$).

Our general approach is the same as Alexander's and Cerf's: to cut $g_t(S^2)$ into a number of "simpler" 2-spheres by a sequence of surgeries in horizonal planes. There are two sorts of complications which arise when k is large. First, the "simple" 2-spheres produced by the surgery process, which we call *primitive* spheres, are considerably more complicated than when $k = 0$ or 1. In particular, an explicit enumeration as in [2] is not feasible. To analyze primitives effectively, we take a viewpoint somewhat different from that of Alexander and Cerf. Instead of projecting primitives onto a vertical line via the height function, we project them onto a horizontal plane. This leads to the notion of *contours* which is central to the proof. The second sort of complication which comes with larger values of k is that when constructing the extensions $\bar{g}_t$ it is no longer sufficient to make discrete choices, as it was in [2]. One is forced to make continuous choices, and to organize these choices a certain amount of machinery must be developed.

Reflecting these two sorts of complications, the paper is divided into two parts. The first contains mostly the machinery reducing the proof to the study of primitives, which is carried out in the second part. Each part has its own little introduction, so we will say no more here about the proof.

We refer also to [4], where the theorem was announced, for a few general remarks on the proof.

PART I: Reduction to Primitives

PART II: Primitive Spheres

PART I: Reduction to Primitives

A quick outline of this part runs as follows. After two easy preliminary normalizations of g_t in Section 1, the surgery process is formalized in Section 2 into the construction of a family of spaces Σ_{tu} for $(t, u) \in S^k \times [0,1]$. When $u = 1$, $\Sigma_{t1} = g_t(S^2)$. As u decreases to 0, Σ_{tu} changes by surgeries on $g_t(S^2)$ in horizontal planes at discrete times u (depending on t). When $u = 0$, Σ_{t0} consists entirely of primitive spheres. We extend Σ_{tu} to a family of spaces $\bar{\Sigma}_{tu}$ by adjoining disjoint copies of the open balls in $\mathbf{R}^3$ bounded by the sphere factors of Σ_{tu}. In Section 3, nice spherical models S_{tu} and $\bar{S}_{tu}$ for Σ_{tu} and $\bar{\Sigma}_{tu}$ are introduced, using the spherical geometry of S^3, with $\bar{S}_{tu}$ to serve eventually as domain of a family of diffeomorphisms $\bar{g}_{tu}$: $\bar{S}_{tu} \to \bar{\Sigma}_{tu}$. Since $\Sigma_{t1} = g_t(S^2)$ and $S_{t1} = S^2$, the desired extension $\bar{g}_t$ of the original g_t: $S^2 \to \mathbf{R}^3$ will be $\bar{g}_{t1}$. This $\bar{g}_{t1}$ is constructed by first constructing $\bar{g}_{t0}$, then extending to $\bar{g}_{tu}$, then restricting to $u = 1$. The construction of $\bar{g}_{t0}$ is the hard part, and this is what Part II achieves. Extending $\bar{g}_{t0}$ to $\bar{g}_{tu}$ is relatively straightforward, and is done in Section 4 of

Part I. (As a technical point, though, it is convenient to have first modified $\bar{S}_{tu}$ and $\bar{\Sigma}_{t0}$ to "continuous" versions of themselves, $\bar{S}^c_{tu}$ and $\bar{\Sigma}^c_{t0}$.)

1. Preliminary normalization

We operate in the C^∞ category. In particular, families of smooth objects are to vary continuously in the C^∞ topology.

Let $S_t \subset \mathbf{R}^3$ be a family of compact embedded surfaces, parametrized by a compact submanifold of S^k, and with a chosen orientation of their normal bundles (varying continuously with t). We say S_t is a *family of elementary surfaces* if the following conditions are satisfied:

(1) For each t, each component of ∂S_t lies in a horizontal plane (different components lying perhaps in different planes).

(2) For each t, each vertical line in $\mathbf{R}^3$ meets S_t in a connected set (which may be empty).

(3) If S_t^+ (S_t^-) denotes the subset of S_t where the positively oriented unit normal vector to S_t has strictly positive (negative) z-coordinate, then:

(a) The closures in $\mathbf{R}^2 \times S^k$ of $\bigcup_t \pi(S_t^+)$ and $\bigcup_t \pi(S_t^-)$ are disjoint, where π: $\mathbf{R}^3 \to \mathbf{R}^2$ is vertical projection to a horizontal $\mathbf{R}^2$.

(b) The closures of $\bigcup_t S_t^+$ and $\bigcup_t S_t^-$ in $\mathbf{R}^3 \times S^k$ are disjoint from $\bigcup_t \partial S_t$.

PROPOSITION 1.1. *If g_t: $S^2 \to \mathbf{R}^3$, $t \in S^k$, is a family of embeddings, then there exists a family of diffeomorphisms f_t: $\mathbf{R}^3 \to \mathbf{R}^3$ and a finite collection of closed k-balls $B_i \subset S^k$, each provided with a finite set of horizontal planes*

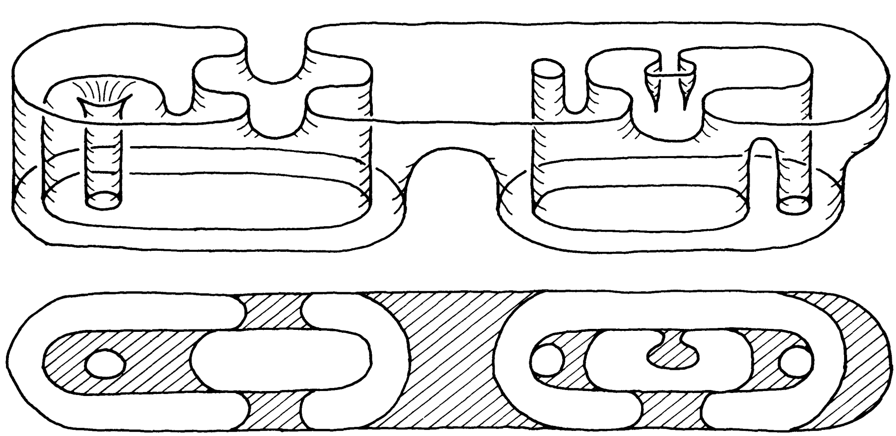


FIGURE 1.1 AN ELEMENTARY SURFACE: SIDE AND TOP VIEWS

$P_{ij} = \{(x, y, z) | z = z_{ij}\}$, $j = 1, \ldots, n_i$, *with* $z_{i1} < \cdots < z_{in_i}$, *such that:*

(i) $\bigcup_i \mathrm{int}(B_i) = S^k$.

(ii) $P_{ij} \neq P_{i'j'}$ *if* $(i, j) \neq (i', j')$.

(iii) *For each pair* (i, j) *with* $1 \leq j < n_i$, $f_t g_t(S^2) \cap \{(x, y, z) | z_{ij} \leq z \leq z_{i, j+1}\}$ *is a family of elementary surfaces as t ranges over* B_i.

(iv) *For each i and* $t \in B_i$, $f_t g_t(S^2)$ *lies between the planes* P_{i1} *and* P_{in_i}.

Proof. Let $U_t^{\pm} \subset g_t(S^2)$ be the set of points where the outer normal to $g_t(S^2)$ makes an angle $\leq \pi/4$ with the vector $\pm(0, 0, 1)$. Let ρ_t: $g_t(S^2) \to [0, 1]$ be a family of functions supported in $U_t^+ \cup U_t^-$, with $\rho_t > 0$ at points having a horizontal tangent plane. Define the vector field v_t on $g_t(S^2)$ to be $\mathrm{grad}(h_t) + \rho_t \cdot (0, 0, 1)$, where h_t is the height function on $g_t(S^2)$. Thus v_t has positive z-coordinate on $g_t(S^2)$ and can be extended to a vector field v_t on $\mathbf{R}^3$ with the same property. Let f_t be a family of level-preserving diffeomorphisms of $\mathbf{R}^3$ taking the trajectories of v_t to vertical lines in $\mathbf{R}^3$.

There exists a $\delta > 0$, independent of t, such that every vertical line segment in $\mathbf{R}^3$ of length $\leq \delta$ meets $f_t g_t(S^2)$ in a connected set (perhaps empty). This follows because the vector field $(0, 0, 1)$ on $f_t g_t(S^2)$ can point to only one side of $f_t g_t(S^2)$, locally. Namely, $(0, 0, 1)$ points outward on $f_t(U_t^+)$ and inward on $f_t(U_t^-)$, these two sets being separated by some distance $d > 0$, independent of t. We may suppose $\delta < \frac{1}{2}d$.

For each t we may choose a finite number of horizontal planes P_{tj} transverse to $f_t g_t(S^2)$, adjacent planes being of distance $< \delta$ apart and $f_t g_t(S^2)$ lying between the extreme P_{tj}'s. These planes remain transverse to $f_t g_t(S^2)$ in some ball $B_t \subset S^k$ about t. By compactness of S^k, a finite number of these balls have interiors covering S^k. Relabel these balls B_i and their associated planes P_{ij}. We may suppose condition (ii) holds since each P_{ij} may be chosen from an open set of horizontal planes.

By construction, for each i and $t \in B_i$, the part of $f_t g_t(S^2)$ between adjacent planes P_{ij} and $P_{ij'}$ satisfies the condition (2) for elementary surfaces. It also satisfies (3a) since we have chosen $\delta < \frac{1}{2}d$. Finally, we make $f_t g_t(S^2)$ vertical near each plane P_{ij} for $t \in B_i$ by a deformation supported near P_{ij}. This can clearly be done without destroying properties (2) and (3a). □

From now on, "g_t" will denote the family "$f_t g_t$" in the conclusion of Proposition 1.1.

Let C_t^{ij} be the collection of circles (components) of $g_t(S^2) \cap P_{ij}$ for $t \in B_i$, and let $C_t^i = \bigcup_{j=1}^{n_i} C_t^{ij}$ and $C_t = \bigcup_i C_t^i$, the union over all i such that $t \in B_i$.

PROPOSITION 1.2. *The family* g_t *may be isotoped, keeping the image spheres* $g_t(S^2)$ *setwise fixed, so that the circles of* $g_t^{-1}(C_t)$ *are actual geometric circles on* S^2 *for all* t.

Proof. Let $S^2(n)$ be S^2 with any n disjoint closed discs removed, the boundary of each disc being a geometric circle on S^2. Let $C(m,n)$ be the space of m-tuples of disjoint smooth circles $C_1,\ldots,C_m$ in $S^2(n)$ each of which bounds a disc in $S^2(n)$, and let $C_0(m,n) \subset C(m,n)$ be the subspace in which all circles C_i are geometric circles in $S^2(n)$. We claim the inclusion $C_0(m,n) \hookrightarrow C(m,n)$ is a homotopy equivalence. To see this, first consider the fibrations

$$\begin{array}{ccccc} C(p,q) \times C(r,s) & \to & C(m,n) & \to & C(1,n) \\ \cup & & \cup & & \cup \\ C_0(p,q) \times C_0(r,s) & \to & C_0(m,n) & \to & C_0(1,n) \end{array}$$

obtained by restriction to one circle C_i which (when $n > 0$) is outermost among $C_1,\ldots,C_m$; i.e, the disc C_i bounds in $S^2(n)$ is not contained in the disc bounded by any other C_j. Since $p < m$ and $r < m$, induction on m reduces us to the case $m = 1$. From Smale's theorem $\mathrm{Diff}(D^2 \mathrm{rel}\, \partial D^2) \simeq *$ it follows by standard arguments (as in the appendix) that $C(1,n)$ deformation retracts onto the subspace of small geometric circles in $S^2(n)$ if $n > 0$, and onto the subspace of great circles on S^2 if $n = 0$. The same is obviously true for $C_0(1,n)$, so the claimed equivalence $C_0(m,n) \simeq C(m,n)$ holds.

To prove the proposition we construct a family of isotopies of S^2 which "round out" the circles of $g_t^{-1}(C_t)$. This is done inductively over the multiple intersections $B_{i_1} \cap \cdots \cap B_{i_n}$, proceeding from larger to smaller values of n. The equivalences $C_0(m,0) \simeq C(m,0)$ assure that no obstructions to extending roundings from n-fold intersections $B_{i_1} \cap \cdots \cap B_{i_n}$ to $(n-1)$-fold intersections can occur. □

2. Factorization into primitives

We may suppose the given family of embeddings $g_t\colon S^2 \to \mathbf{R}^3$ has been normalized as in Propositions 1.1 and 1.2. Recall also from Section 1 the definitions of the sets C_t^{ij}, C_t^i, and C_t. Choose a (smooth) family of functions $\varphi_t\colon C_t \to (-1,1)$, $t \in S^k$, satisfying:

(1) $\varphi_t | C_t^{ij}$ is injective, giving a linear ordering of C_t^{ij} such that $\varphi_t(c) > \varphi_t(c')$ if c lies inside c' in P_{ij}.

(2) $\varphi_t(C_t^i) > 0$ for $t \in B_i'$, where B_i' is a closed ball in $\mathrm{int}(B_i)$ and $\bigcup_i \mathrm{int}(B_i') = S^k$.

(3) $\varphi_t(C_t^i) < 0$ for $t \in \partial B_i$.

We may assume all the graphs $Z'(c) = \{(t, \varphi_t(c)) | t \in S^k\} \subset S^k \times (-1,1)$ have general position intersections with each other and with $S^k \times \{0\}$. Let $Z(c) = Z'(c) \cap S^k \times [0,1]$ and $Z_0(c) = Z(c) \cap S^k \times \{0\}$. The intersections of the various $Z(c)$'s give a stratification $\mathfrak{S}$ of $S^k \times [0,1]$, which intersects $S^k \times \{0\}$ in a stratification $\mathfrak{S}_0$ of $S^k \times \{0\}$. Relative product neighborhoods $(Z(c), Z_0(c)) \times$

$[-1,1]$ of $(Z(c), Z_0(c)) = (Z(c), Z_0(c)) \times \{0\}$ in $(S^k \times [0,1], S^k \times \{0\})$ can be chosen so that

$$Z(c) \times \{s\} = \{(t,u) \in S^k \times [0,1] | u = \varphi_t(c) + s\varepsilon\}$$

for some constant $\varepsilon > 0$, chosen small enough that different $Z(c) \times [-1,1]$'s intersect only near the intersections of the corresponding $Z(c)$'s.

Let $\delta > 0$ be chosen so that $g_t(S^2)$ is vertical within distance δ of each P_{ij} for which $t \in B_i$, and so that any two P_{ij}'s are of distance greater than 2δ apart. For each $c \in C_t$ choose $\delta(c) \in (0, \delta)$, independent of t, so that $\delta(c) > \delta(c')$ for each pair $c, c' \in C_t^{ij}$ with c inside c' in P_{ij}. Define the collection C_{tu} of level circles on $g_t(S^2)$ for $(t,u) \in S^k \times [0,1]$ by:

(4) A circle $c \in C_t$ belongs to C_{tu} for $u = \varphi_t(c)$.

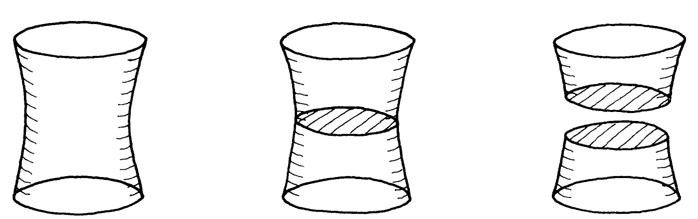
(5) The two circles parallel to $c \in C_t$, above and below at distance $\delta(c) \cdot \min\{s, 1\}$, belong to C_{tu} for $u = \varphi_t(c) - s\varepsilon$, $s > 0$.

For $(t,u) \in S^k \times [0,1]$, let Σ_{tu} be the family of spaces obtained from $g_t(S^2)$ by first removing the open vertical annuli between pairs of parallel circles of C_{tu}, in (5), and then adjoining to each circle of C_{tu} the horizontal disc it bounds. (Note that these discs are all disjoint for fixed t and u.)

Thus as u decreases from $u > \varphi_t(c)$ to $u < \varphi_t(c)$, $g_t(S^2)$ changes by surgery along c using the horizontal disc bounded by c. See Figure 2.1.

We define a *factor* of Σ_{tu} to be a 2-sphere (with corners) contained in Σ_{tu} obtained from the closure of a component of $g_t(S^2) - C_{tu}$, other than an annulus thrown away when doing surgery, by capping off its boundary circles by the horizontal discs they bound in Σ_{tu}. As (t,u) varies over a (connected) stratum of $\mathcal{S}$, each factor of Σ_{tu} varies only by isotopy. Restricting to $u = 0$, we see by condition (2) above that as t varies over a stratum of $\mathcal{S}_0$, the factors of Σ_{t0} form families of primitive 2-spheres, according to the following:

Definition. A family of 2-spheres with corners $\Sigma_t \subset \mathbf{R}^3$ is called a *family of primitive 2-spheres* if there is a family of elementary surfaces $S_t \subset \Sigma_t$ such that



$u > \varphi_t(c)$ $u > \varphi_t(c)$ $u > \varphi_t(c)$

FIGURE 2.1

$\mathrm{cl}(\Sigma_t - S_t)$, the closure of $\Sigma_t - S_t$, consists of finitely many disjoint horizontal discs, called *faces* of Σ_t, for each t. The only corners of Σ_t occur at the boundaries of these faces, i.e., at ∂S_t.

Let $\overline{\Sigma}_{tu}$ be the space obtained from the disjoint union of the balls in $\mathbf{R}^3$ bounded by the factors of Σ_{tu} by identifying their horizontal faces as they are identified in Σ_{tu}. Thus $\Sigma_{tu} \subset \overline{\Sigma}_{tu}$.

Define Γ_{tu} to be the graph whose vertices correspond to the factors of Σ_{tu}, and whose (open) edges correspond to the common horizontal faces between factors. An edge e corresponding to a common face Δ belongs to Γ_{tu} for $(t, u) \in Z(e) \equiv Z(\partial\Delta)$. For $(t, u) \in Z(e) \times \{s\}$ with $s > 0$, the edge e collapses to a single vertex of Γ_{tu}, while if $s < 0$, e is deleted from Γ_{tu}. On strata of $\mathcal{S}$ (which by definition are connected), Γ_{tu} is constant.

The components of Γ_{tu} are trees, since every circle on a 2-sphere separates. A component γ of Γ_{tu} corresponds to a connected component $\Sigma_{tu}(\gamma)$ of Σ_{tu}, varying continuously with (t, u) in the given stratum of $\mathcal{S}$ over which γ is constant. The factors of $\Sigma_{tu}(\gamma)$ are partially ordered by the inclusion relations among the balls they bound in $\mathbf{R}^3$. The maximal factors in this ordering correspond to the vertices of a subtree of γ, clearly. A common face Δ of two factors Σ_1 and Σ_2 of $\Sigma_{tu}(\gamma)$ is a *sum* face if Σ_1 and Σ_2 bound balls in $\mathbf{R}^3$ meeting only in Δ. In the opposite case, Δ is a *difference* face, and there is an inclusion relation between the two balls bounded by Σ_1 and Σ_2.

It will be convenient to extend the domain of definition of $\Sigma_{tu}(\gamma)$ from (t, u) in the given stratum of $\mathcal{S}$ over which γ is constant to (t, u) in the closure of this stratum, by continuity. With this done, certain inclusion relations $\Sigma_{tu}(\gamma) \subset \Sigma_{tu}(\gamma')$ result.

3. Spherical models

In this section we shall be doing some constructions using 3-dimensional spherical geometry. Thus the ambient space is S^3, the unit sphere in $\mathbf{R}^4$, and "circles" and "spheres" in S^3 mean intersections with affine subspaces of $\mathbf{R}^4$ of the appropriate dimension. In particular, we fix $S^2 \subset S^3$ to be the unit sphere in $\mathbf{R}^3 = \{(x, y, z, 0)\} \subset \mathbf{R}^4$, and we let B^3_+ and B^3_- be the balls bounded by S^2 in S^3. For any circle c on S^2, we let S^2_c be the 2-sphere orthogonal to S^2 with $S^2_c \cap S^2 = c$.

By Proposition 1.2, we have the collection $g_t^{-1}(C_t)$ of circles on S^2. The components of $g_t^{-1}(C_{tu})$ are near the circles of $g_t^{-1}(C_t)$, so we may take them to be actual circles too. For each circle $c \in g_t^{-1}(C_{tu})$, we choose the spherical disc D_c to be $S^2_c \cap B^3_+$ or $S^2_c \cap B^3_-$, according to whether the horizontal disc of Σ_{tu} bounded by $g_t(c)$ points inside or outside $g_t(S^2)$ at its boundary. For fixed (t, u), these discs D_c are disjoint since their boundaries are disjoint. In analogy with the definition of Σ_{tu}, we let S_{tu} be obtained from S^2 by adjoining all the discs D_c for $c \in g_t^{-1}(C_{tu})$, and by deleting the open annuli in S^2 whose g_t-images were deleted in the construction of Σ_{tu}.

Factors of S_{tu} are defined analogously to factors of Σ_{tu} and correspond bijectively via g_t. For each factor S of S_{tu}, let $\bar{S} \subset S^3$ be the topological 3-ball bounded by S, chosen so that the normals to $S \cap S^2$ pointing into $\bar{S}$ point into B^3_+ or B^3_- according to whether, for the corresponding factor Σ of Σ_{tu}, the normals to $\Sigma \cap g_t(S^2)$ pointing into $\bar{\Sigma}$ point into the ball in $\mathbf{R}^3$ bounded by $g_t(S^2)$ or not. We let $\bar{S}_{tu}$ be the disjoint union of all such 3-balls $\bar{S}$, with the discs D_c in their boundaries identified as they are in S_{tu}.

Let S be a factor of S_{tu}, having corners at the circles $c_1, \ldots, c_n \subset S^2$. Thus $\partial(S \cap S^2) = c_1 \cup \cdots \cup c_n$, c_i being capped off by the disc $D_i \subset S$ which we call a *face* of S. Let $B_i \subset S^3$ be the 3-ball bounded by $S^2_{c_i}$, with $B_i \cap S = D_i$. These B_i's are disjoint since they intersect S^2 in disjoint discs, the components of $\mathrm{cl}(S^2 - S)$. Define the *core* of S to be $\mathrm{cl}(\bar{S} - \bigcup_i B_i) - S^2$.

Components $S_{tu}(\gamma)$ of S_{tu} correspond via g_t to components $\Sigma_{tu}(\gamma)$ of Σ_{tu}. The factors of $S_{tu}(\gamma)$ are partially ordered: $S_1 < S_2$ if $\bar{S}_1 \subset \bar{S}_2$. This ordering corresponds under g_t to the ordering on factors of $\Sigma_{tu}(\gamma)$ described in Section 2. As with $\Sigma_{tu}(\gamma)$, we extend the domain of definition of $S_{tu}(\gamma)$ to be a closed subset of $S^k \times [0, 1]$, yielding some inclusion relations $S_{tu}(\gamma) \subset S_{tu}(\gamma')$.

The goal of the rest of this section is to construct families S^c_{tu} and $\bar{S}^c_{tu}$ which are continuous versions of S_{tu} and $\bar{S}_{tu}$, together with a family of one-dimensional foliations F_{tu} on $\bar{S}^c_{tu}$. Roughly, the idea is that discontinuities in S_{tu} occur when a common face is deleted, as in Figure 2.1 when u increases from $u = \varphi_t(c)$ to $u > \varphi_t(c)$. To remedy this, we wish to isotope the common face across an adjacent factor, rather than simply delete it. (Notice that only the deletion of *difference* faces causes discontinuities in $\bar{S}_{tu}$.) The foliations F_{tu} will actually be defined on $\bar{S}_{tu}$ as well as $\bar{S}^c_{tu}$, and will be used to construct S^c_{tu} and also $\bar{g}_{tu}$: $\bar{S}^c_{tu} \to \bar{\Sigma}^c_{tu}$, eventually.

A point $p \in S^3 - S^2$ determines a *polar foliation* F on S^3 whose leaves are arcs of circles through p orthogonal to S^2. There are two "pole" singularities of F, one at p and one at a dual point p'. Our next task is to construct families of polar foliations $F_{tu}(\gamma)$ for γ a component of Γ_{tu}, $F_{tu}(\gamma)$ being defined and varying continuously for (t, u) in the closure of the stratum of $\mathfrak{S}$ over which γ is defined. We shall require the foliations $F_{tu}(\gamma)$ to satisfy:

(1) $F_{tu}(\gamma) = F_{tu}(\gamma')$ if $S_{tu}(\gamma) \subset S_{tu}(\gamma')$.

(2) $F_{tu}(\gamma)$ has a pole in the core of a maximal factor of $S_{tu}(\gamma)$ unless (t, u) is near the boundary of the stratum where γ is defined.

(Here $F_{tu}(\gamma)$ is a foliation of S^3. The foliation F_{tu} mentioned in the preceding paragraph will be the pullback of all the $F_{tu}(\gamma)$'s by the natural projection $\bar{S}_{tu} \to S^3$.)

Such foliations $F_{tu}(\gamma)$ can be constructed inductively over the strata of $\mathcal{S}$. For an i-stratum X and a component γ of Γ_{tu} defined over X, condition (1) determines $F_{tu}(\gamma)$ over ∂X. The union U of the cores of the maximal factors of $S_{tu}(\gamma)$ is a deformation retract of $\text{int}(B^3_+)$ or $\text{int}(B^3_-)$, since these factors correspond to the vertices of a subtree γ_0 of γ whose edges correspond to common sum faces between factors. Using a monotone deformation retraction of $\text{int}(B^3_\pm)$ onto U we can push the poles of $F_{tu}(\gamma)$ into $\text{int}(U)$ as (t, u) goes from ∂X into X monotonically with respect to the bicollars $Z(e) \times [-1, 1]$. Then we extend $F_{tu}(\gamma)$ over the rest of X, keeping a pole in $\text{int}(U)$, which is possible since $\text{int}(U)$ is contractible.

LEMMA 3.1. *For each (t, u), one of the following holds:*

(a) *$F_{tu}(\gamma)$ has a pole in the core of a factor $\bar{S}_0$ of $\bar{S}_{tu}(\gamma)$ corresponding to a vertex of γ_0, and no other factor of $\bar{S}_{tu}(\gamma)$ contains a pole of $F_{tu}(\gamma)$.*

(b) *$F_{tu}(\gamma)$ has a pole in a common sum face of $S_{tu}(\gamma)$ corresponding to an edge of γ_0, and the two factors of $\bar{S}_{tu}(\gamma)$ containing this common face are the only factors of $\bar{S}_{tu}(\gamma)$ containing a pole of $F_{tu}(\gamma)$.*

(c) *No factor of $\bar{S}_{tu}(\gamma)$ contains a pole of $F_{tu}(\gamma)$, but a factor S_0 corresponding to a vertex of γ_0 has a preferred face D_i such that the poles of $F_{tu}(\gamma)$ lie in $\text{int}(B_i)$, B_i being the 3-ball such that $B_i \cap S_0 = D_i$.* □

These three possibilities are shown in Figure 3.1, in which $S_{tu}(\gamma)$ is represented by the solid lines and U is shaded. (In the limiting case of (c), a pole of $F_{tu}(\gamma)$ could lie in $D_i - \partial D_i$. But by definition, $D_i - \partial D_i$ is contained in the core of $\bar{S}_0$, so this is really case (a).)

We may assume by transversality that case (b) of Lemma 3.1 holds for the common face corresponding to a given edge e for (t, u) in a codimension one

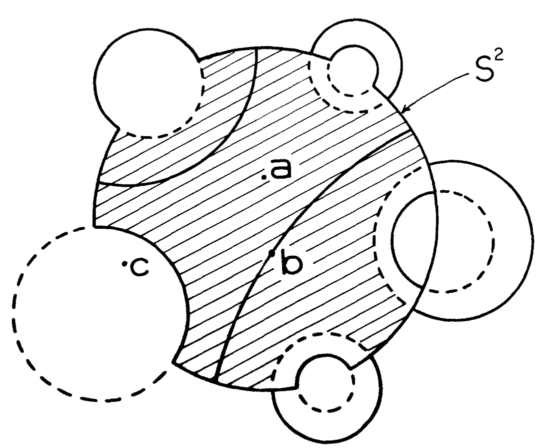


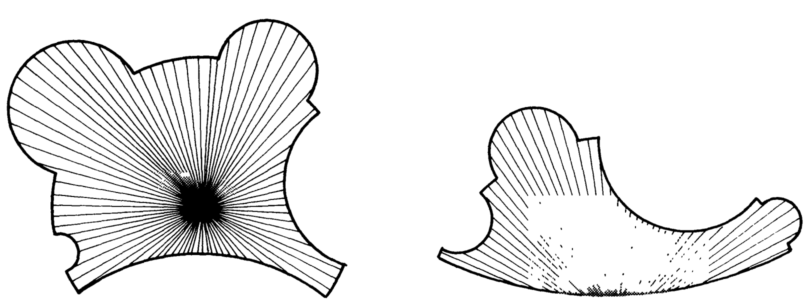
FIGURE 3.1

submanifold $Y(e)$ of $Z(e)$, and that the pole moves across this common face as (t, u) crosses a bicollar neighborhood $Y(e) \times [-1,1]$ of $Y(e) = Y(e) \times \{0\}$ in $Z(e)$. Via the projection $Z(e) \times [-1,1] \to Z(e)$, we can pull back $Y(e)$ to a codimension one submanifold $\tilde{Y}(e) = Y(e) \times [-1,1]$ of $Z(e) \times [-1,1]$ with bicollar neighborhood $\tilde{Y}(e) \times [-1,1]$ in $Z(e) \times [-1,1]$. In general position $Y(e)$ will be transverse to the stratification of $Z(e)$ induced by $\mathbb{S}$, and we may assume different $Y(e)$'s have general position intersections with each other and with $S^k \times \{0\}$. So the various $Y(e)$'s and $Z(e)$'s together determine stratifications $\mathbb{S}'$ subdividing $\mathbb{S}$ and $\mathbb{S}_0'$ subdividing $\mathbb{S}_0$.

Case (c) of Lemma 3.1 can hold only near the boundary of a given stratum of $\mathbb{S}'$. As (t, u) moves away from the boundary of the stratum, a pole of $F_{tu}(\gamma)$ moves inside the factor S_0, and we are in case (a). The vertex of γ corresponding to this factor S_0 we call the *base vertex* of γ; it does not change as (t, u) ranges over the given stratum of $\mathbb{S}'$. The only other possibility for a stratum of $\mathbb{S}'$ is that case (b) holds for all (t, u) in the stratum. Then the edge of γ corresponding to the common face containing a pole of $F_{tu}(\gamma)$ we call the *base edge* of γ.

For a factor S of $S_{tu}(\gamma)$, if the core of S contains a pole of $F_{tu}(\gamma)$, then the leaves of $F_{tu}(\gamma)$ in $\bar{S}$ are the trajectories of a flow on $\bar{S}$ in which all points move to the pole, staying in $\bar{S}$. To see this, apply a circle and sphere preserving map of S^3 leaving S^2 invariant and taking the pole of $F_{tu}(\gamma)$ in $\bar{S}$ to $(0,0,0,-1)$. Then under stereographic projection from $(0,0,0,1)$, $F_{tu}(\gamma)$ becomes the foliation of $\mathbf{R}^3$ by lines through the origin, and the assertion is then obvious. Similarly, if $\bar{S}$ contains neither pole of $F_{tu}(\gamma)$, then the leaves of $F_{tu}(\gamma)$ in $\bar{S}$ are arcs with one endpoint on a preferred face of $\bar{S}$. The latter type of foliation of $\bar{S}$ we call *facial*, the former type *polar*. By Lemma 3.1, the restriction of $F_{tu}(\gamma)$ to each factor of $\bar{S}_{tu}(\gamma)$ is either polar or facial.

Let $V^i \subset S^k \times [0,1]$ be the set of points lying in exactly i subsets $Z(e) \times [0,1]$. Each component V^{ij} of V^i corresponds to a unique stratum S^{ij} of $\mathbb{S}$, if we replace the sets $Z(e) \times [0,1]$ which define V^{ij} by their subsets $Z(e) = Z(e) \times \{0\}$. Let γ be a component of Γ_{tu} for $(t, u) \in S^{ij}$. We modify $S_{tu}(\gamma)$ to a family $S_{tu}^{ij}(\gamma)$



polar facial

FIGURE 3.2

by having its common faces lie over V^{ij} rather than S^{ij}. This is certainly possible if the collars $Z(e) \times [0,1]$ are chosen small enough, i.e., if the ε in Section 2 is small. The foliations $F_{tu}(\gamma)$ extend naturally to foliations $F_{tu}^{ij}(\gamma)$ on $\bar{S}_{tu}^{ij}(\gamma)$.

Using these foliations $F_{tu}^{ij}(\gamma)$ we form, by an inductive procedure, families of deformations $S_l(t_1,\ldots,t_n)$, $0 \le t_m \le 1$, of the factors S_l of $S_{tu}^{ij}(\gamma)$, where $D_1,\ldots,D_n$ are the common faces of $S_{tu}^{ij}(\gamma)$. To start, suppose S_n is a minimal factor such that $F_{tu}^{ij}(\gamma)$ is facial on $\bar{S}_n$, with preferred common face D_n. As t_n goes from 0 to 1, let D_n move monotonically across $\bar{S}_n$ to $S_n - \mathrm{int}(D_n)$, keeping ∂D_n fixed and staying transverse to $F_{tu}^{ij}(\gamma)$ (except at the end of the isotopy, when D_n has moved to $S_n - \mathrm{int}(D_n)$, which is not transverse to $F_{tu}^{ij}(\gamma)$ at its corners). This gives the deformation $S_n(0,\ldots,0,t_n)$ of S_n and also a deformation $S_m(0,\ldots,0,t_n)$ of the other factor S_m of $S_{tu}^{ij}(\gamma)$ having D_n as a face. For all other factors S_l of $S_{tu}^{ij}(\gamma)$, let $S_l(0,\ldots,0,t_n)$ be independent of t_n. Next, one proceeds to a factor S_{n-1} minimal among the remaining factors with $F_{tu}^{ij}(\gamma)$ facial on $\bar{S}_{n-1}$, and constructs the families $S_l(0,\ldots,0,t_{n-1},t_n)$, and so on. The process can be iterated since at each time the property that $F_{tu}^{ij}(\gamma)$ is facial on a factor $\bar{S}_l$ is evidently preserved. For the final step of deforming the common face D_1 of $S_{tu}^{ij}(\gamma)$, we note there is a two-fold ambiguity in which factor to push D_1 across in case $F_{tu}^{ij}(\gamma)$ has a pole in D_1.

Over V^{ij}, we form $\bar{S}_{tu}^{c}$ from $\bar{S}_{tu}$ by replacing each factor S_l of each component $S_{tu}^{ij}(\gamma)$ by its deformation $S_l(t_1,\ldots,t_n)$, where $(t,u) \in Z(e_m) \times \{t_m\} \subset Z(e_m) \times [0,1]$, for each edge e_m of γ, $m = 1,\ldots,n$. The two-fold ambiguity mentioned in the preceding paragraph concerns a sum face, so $\bar{S}_{tu}^{c}$ is unaffected by which choice is made. To form $\bar{S}_{tu}^{c}$ globally, for all V^{ij}'s, we proceed by induction on i. For the induction step we can extend families of deformations $S_l(t_1,\ldots,t_n)$ over V^{ij} since at each step in the formation of $S_l(t_1,\ldots,t_n)$, the space of permitted deformations of a face D_m transverse to $F_{tu}^{ij}(\gamma)$ is contractible.

We note that the foliations $F_{tu}^{ij}(\gamma)$ on the components $\bar{S}_{tu}^{ij}(\gamma)$ determine automatically a foliation F_{tu} defined on $\bar{S}_{tu}^{c}$, which is also either polar or facial on each factor of $\bar{S}_{tu}^{c}$.

The definition of $\bar{S}_{tu}^{c}$ also defines a family S_{tu}^{c}, modulo the two-fold ambiguity of how to push a common face D_1 which contains a pole of F_{tu}. If D_1 corresponds to the edge e_1 of Γ_{tu}, this ambiguity occurs for

$$(t,u) \in Y(e_1) \times \{0\} \times [0,1] \subset Z(e_1) \times [0,1].$$

We resolve the ambiguity by splitting D_1 into two copies of itself and taking *both* deformations of these for $(t,u) \in Y(e_1) \times \{0\} \times [0,1]$.

The resulting family S_{tu}^{c} has discontinuities at such sets $Y(e_1) \times \{0\} \times [0,1]$, as one of the two deformed copies of D_1 is deleted in passing to $Y(e_1) \times \{s\} \times [0,1]$ with $s \neq 0$. For later purposes, in Section 13, it will be convenient not to have such discontinuities, mild as they are. So we enlarge S^c_{tu} to a continuous family, which will still be called S^c_{tu}, as follows. Rather than delete one of the deformed copies of D_1 as s goes from $s = 0$ to $s > 0$, say, in $Y(e_1) \times \{s\} \times [0, 1]$, we adjoin to S^c_{tu} the stage $t_1 - s$ deformation of D_1 across the appropriate adjacent factor of (the unenlarged) S^c_{tu}, for $(t, u) \in Y(e_1) \times \{s\} \times \{t_1\}$. See the upper right quadrant of Figure 3.3, where this adjoined deformation of D_1 is indicated by the dashed line; and similarly for the other copy of D_1 in $Y(e_1) \times [-1, 0] \times [0, 1]$.

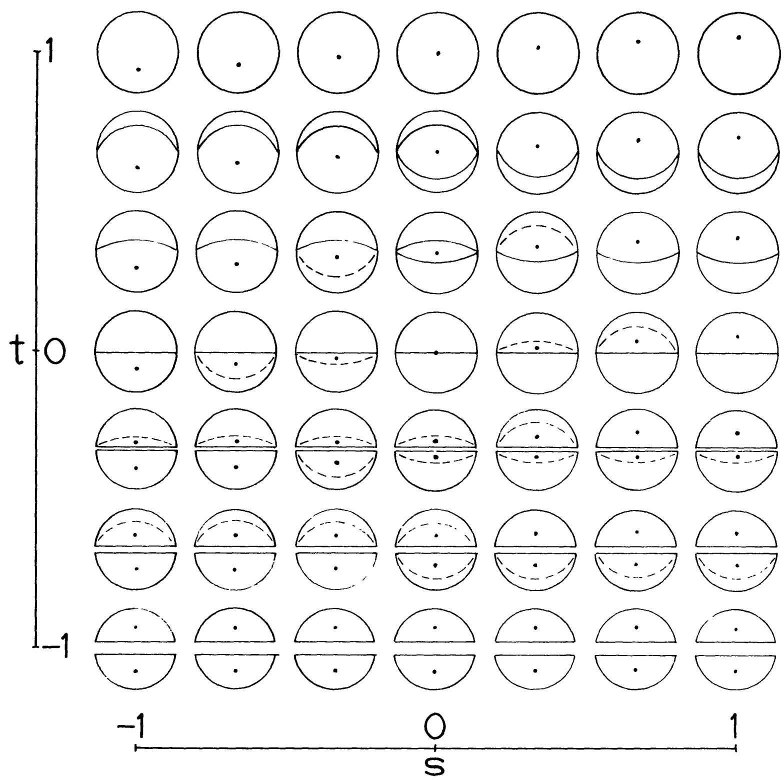


FIGURE 3.3

This still leaves a discontinuity in the (enlarged) family S^c_{tu}, as one of the deformed copies of D_1 is deleted as t_1 goes from $t_1 = 0$ to $t_1 < 0$. To eliminate this discontinuity we continue the deformation of the two copies of D_1 for $t_1 < 0$, as shown in the lower half of Figure 3.3. This can be described as follows. As t_1 goes from 0 to -1, a pole of F_{tu} crosses the face D_1. When this happens, we deform D_1 across the factor into which the pole is moving, the deformed D_1 staying ahead of the pole, like a shock wave. This deformation of D_1

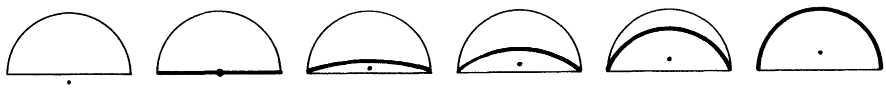


FIGURE 3.4

starts the instant the pole hits D_1 and finishes before $t_1 = -1$. See Figure 3.4. We adjoin this deformation of D_1 to S^c_{tu}. To be continuous, we then adjoin this "shock wave" deformation of a face D_1 being punctured by a pole, not just in $Y(e_1) \times [-1,1] \times [-1,0]$, but in all of $Z(e_1) \times [-1,0]$ in fact. There are no problems with interactions of such deformations of different D_1's since they occur in separate factors of (the unenlarged) S^c_{tu}.

4. Extending $\bar{g}_{t0}$ to $\bar{g}_{tu}$

PROPOSITION 4.1. *Given a family of maps $\bar{g}_{t0}$: $\bar{S}^c_{t0} \to \mathbf{R}^3$ which restrict to embeddings on the factors of $\bar{S}^c_{t0}$ and which agree with g_t on $S^2 \cap S_{t0} \subset \bar{S}^c_{t0}$, there exists a family $\bar{g}_{tu}$: $\bar{S}^c_{tu} \to \mathbf{R}^3$, $0 \leq u \leq 1$, extending $\bar{g}_{t0}$, which also restricts to embeddings on factors and agrees with g_t on $S^2 \cap S_{tu} \subset \bar{S}^c_{tu}$.*

Since $S^c_{t1} = S^2$ and hence $\bar{S}^c_{t1} = B^3$, this reduces the proof of the theorem to the construction of $\bar{g}_{t0}$.

Note also that given $\bar{g}_{t0}$: $\bar{S}^c_{t0} \to \mathbf{R}^3$, an embedding on each factor, we can construct a family of spaces $\bar{\Sigma}^c_{t0}$ by taking the disjoint union of the images of the factors of $\bar{S}^c_{t0}$ and then identifying points in their boundaries whenever their pre-images in $\bar{S}^c_{t0}$ are identified. Then $\bar{g}_{t0}$ gives a family of maps $\bar{g}_{t0}$: $\bar{S}^c_{t0} \to \bar{\Sigma}^c_{t0}$ which are diffeomorphisms on factors. Conversely, given such a family $\bar{g}_{t0}$: $\bar{S}^c_{t0} \to \bar{\Sigma}^c_{t0}$, we can recover $\bar{g}_{t0}$: $\bar{S}^c_{t0} \to \mathbf{R}^3$ by composing with the natural projections $\bar{\Sigma}^c_{t0} \to \mathbf{R}^3$.

For the proof of the proposition we shall use the following elementary result.

LEMMA 4.2. *Given maps $f_1, \dots, f_m$: $X \to [0,1]$, there exist maps $u_1, \dots, u_n$: $X \to [0,1]$ such that for all $x \in X$,*

$$u_1(x) \leq \cdots \leq u_n(x) \quad \textit{and} \quad \{u_i(x) | i = 1, \dots, n\} = \{f_j(x) | j = 1, \dots, m\}.$$

Proof. For $j = 1, \dots, m-1$, let $f'_j = \min\{f_j, f_m\}$ and $f''_j = \max\{f_j, f_m\}$. By induction on m, we apply the lemma to the collections $\{f'_j\}$ and $\{f''_j\}$ separately to produce functions $u'_1 \leq \cdots \leq u'_r$ and $u''_1 \leq \cdots \leq u''_s$ with

$$\{u'_i(x) | i = 1, \dots, r\} = \{f'_j(x) | j = 1, \dots, m-1\}$$

and

$$\{u_i''(x) | i = 1, \ldots, s\} = \{f_j''(x) | j = 1, \ldots, m-1\}.$$

Then the desired u_i's are the functions $u_1' \leq \cdots \leq u_r' \leq f_m \leq u_1'' \leq \cdots \leq u_s''$. □

Proof of Proposition 4.1. We apply the lemma to the collection of all functions $\varphi_t'(c)$ and $\varphi_t'(c) \pm \varepsilon$ from S^k to $[0, 1]$, where $\varphi_t' = \max\{\varphi_t, 0\}$ and ε is as defined in Section 2. Using the resulting functions u_i: $S^k \to [0, 1]$, we construct $\bar{g}_{tu}$ inductively, the induction step being to extend from $u \leq u_i(t)$ to $u \leq u_{i+1}(t)$. So we suppose $\bar{g}_{tu}$ has already been constructed for $u = u_i(t)$. As u increases from $u_i(t)$ to $u_{i+1}(t)$, $\bar{S}^c_{tu}$ changes in two ways. First, if $[u_i(t), u_{i+1}(t)]$ is contained in an interval $[\varphi_t(c) - \varepsilon, \varphi_t(c)]$, then there are two parallel horizontal faces of $\Sigma^c_{tu_i(t)}$ with boundary circles near $c \in C_t$, and in Σ_{tu} these two faces move toward each other as u goes from $u_i(t)$ to $u_{i+1}(t)$. There is a corresponding moving pair of faces in S_{tu}. It is easy to extend $g_{tu_i(t)}$ over $[u_i(t), u_{i+1}(t)]$ near these faces. If $u_{i+1}(t) = \varphi_t(c)$, then the moving pair of faces converges to a common face when $u = u_{i+1}(t)$, and we can use contractibility of $\mathrm{Diff}(D^2 \operatorname{rel} \partial D^2)$ to assure that $\bar{g}_{tu_{i+1}(t)}$ is well-defined on the common face. And we can arrange that $\bar{g}_{tu_{i+1}(t)}$ be smooth across such a common sum face. Doing this for all intervals $[\varphi_t(c) - \varepsilon, \varphi_t(c)] \supset [u_i(t), u_{i+1}(t)]$, call the resulting family $\bar{g}^{(i)}_{tu}$: $\bar{S}^{(i)}_{tu} \to \bar{\Sigma}^{(i)}_{tu}$, for $u \in [u_i(t), u_{i+1}(t)]$. Thus $\bar{S}^{(i)}_{tu}$ differs from $\bar{S}^c_{tu_i}$ only to the extent that such pairs of parallel faces move together, and similarly for $\bar{\Sigma}^{(i)}_{tu}$ and $\bar{\Sigma}^c_{tu_i(t)}$.

The second way in which $\bar{S}^c_{tu}$ changes as u goes from $u_i(t)$ to $u_{i+1}(t)$ occurs when $[u_i(t), u_{i+1}(t)]$ is contained in intervals $[\varphi_t(c), \varphi_t(c) + \varepsilon]$ corresponding to common difference faces of $\bar{S}^c_{tu_i(t)}$ moving across adjacent factors. (Sum faces are invisible in $\bar{S}^c_{tu}$, and $\bar{g}^{(i)}_{tu}$ is already defined across sum faces.) To handle these difference faces we operate on each component of $\bar{\Sigma}^c_{tu_i(t)}$ separately. (With varying t, the splitting of $\bar{\Sigma}^c_{tu_i(t)}$ into its components can change only when $u_i(t) = u_{i+1}(t)$, where the induction step is vacuous.) The common difference faces $\Delta_1(t, u), \ldots, \Delta_m(t, u)$ of a component of $\bar{\Sigma}^{(i)}_{tu}$ split it into factors $\bar{\Sigma}_0(t, u), \ldots, \bar{\Sigma}_m(t, u)$, where the numbering is chosen so that if the natural projection of $\bar{\Sigma}_p(t, u)$ to $\mathbf{R}^3$ contains the projection of $\bar{\Sigma}_q(t, u)$ then $p \leq q$, and $\Delta_j(t, u)$ is the face of $\bar{\Sigma}_j(t, u)$ separating it from the adjacent larger factor $\bar{\Sigma}_{j'}(t, u)$, $j' < j$.

We shall by an inductive procedure construct for each (t, u) with $u \in [u_i(t), u_{i+1}(t)]$ a family

$$\bar{g}_{tu}(s_1, \ldots, s_m): \bar{S}_{tu}(s_1, \ldots, s_m) \to \bar{\Sigma}_{tu}(s_1, \ldots, s_m)$$

having common difference faces $\Delta_j(t, u, s_1, \ldots, s_m)$ and factors $\bar{\Sigma}_j(t, u, s_1, \ldots, s_m)$, where $s_l \in [s_l^0, s_l^1]$, and s_l^0, s_l^1 are defined by $u_i(t) = \varphi_t(\partial \Delta_l) + s_l^0 \varepsilon$ and

$u_{i+1}(t) = \varphi_t(\partial\Delta_l) + s_l^1\varepsilon$. We begin with the family $\bar{g}_{tu}^{(i)}$: $\bar{S}_{tu}^{(i)} \to \bar{\Sigma}_{tu}^{(i)}$ when $s_l = s_l^0$ for all l. For the induction step we wish to let s_j vary from s_j^0 to s_j^1, assuming we have already made the construction for $s_l \in [s_l^0, s_l^1]$, $l > j$. The family $\bar{S}_{tu}(s_1, \ldots, s_m)$ we have in fact already constructed in the process of forming the family $\bar{S}_{tu}^c$; namely, the j^{th} common difference face of $\bar{S}_{tu}^{(i)}$ moves across the smaller of the two adjacent factors, always staying transverse to the foliation F_{tu}, as s_j goes from s_j^0 to s_j^1. Via $\bar{g}_{tu}(s_1^0, \ldots, s_j^0, s_{j+1}, \ldots, s_m)$, this yields also an isotopy $\Delta_j(t, u, s_1^0, \ldots, s_{j-1}^0, s_j, \ldots, s_m)$ of $\Delta_j(t, u, s_1^0, \ldots, s_j^0, s_{j+1}, \ldots, s_m)$, $s_j \in [s_j^0, s_j^1]$, and so defines $\bar{\Sigma}_{tu}(s_1^0, \ldots, s_{j-1}^0, s_j, \ldots, s_m)$. It remains to extend $\bar{g}_{tu}(s_1^0, \ldots, s_j^0, s_{j+1}, \ldots, s_m)$ to $s_j \in [s_j^0, s_j^1]$. For the factor

$$\bar{\Sigma}_j\big(t, u, s_1^0, \ldots, s_{j-1}^0, s_j, \ldots, s_m\big) \subset \bar{\Sigma}_j\big(t, u, s_1^0, \ldots, s_j^0, s_{j+1}, \ldots, s_m\big),$$

$\bar{g}_{tu}(s_1^0, \ldots, s_{j-1}^0, s_j, \ldots, s_m)$ is defined by restriction. On the adjacent factor $\bar{\Sigma}_{j'}(t, u, s_1^0, \ldots, s_{j-1}^0, s_j, \ldots, s_m)$, as s_j goes from s_j^0 to s_j^1 this factor is shrinking by isotopy. On the boundary of this factor, $\bar{g}_{tu}(s_1^0, \ldots, s_{j-1}^0, s_j, \ldots, s_m)$ is already defined, and we can extend over the interior by isotopy extension.

Having the family

$$\bar{g}_{tu}(s_1, \ldots, s_m) \colon \bar{S}_{tu}(s_1, \ldots, s_m) \to \bar{\Sigma}_{tu}(s_1, \ldots, s_m),$$

we define $\bar{g}_{tu}$: $\bar{S}_{tu}^c \to \bar{\Sigma}_{tu}^c$ for $u \in [u_i(t), u_{i+1}(t)]$ to be

$$\bar{g}_{tu}(s_1^u, \ldots, s_m^u) \colon \bar{S}_{tu}(s_1^u, \ldots, s_m^u) \to \bar{\Sigma}_{tu}(s_1^u, \ldots, s_m^u)$$

where s_j^u is defined by $u = \varphi_t(\partial\Delta_j) + s_j^u\varepsilon$. This completes the induction step in the construction of $\bar{g}_{tu}$: $\bar{S}_{tu}^c \to \bar{\Sigma}_{tu}^c$ for $u \in [0, 1]$. □

On one later occasion the following "blowing up" operation will be useful. Let $\mathfrak{T}$ be a triangulation of S^k in which the closed strata of $\mathbb{S}_0'$ are subcomplexes. There is a well-known way of associating to $\mathfrak{T}$ a handle structure on S^k, in which i-handles $H^i = D^i \times D^{k-i}$ are ε_i-neighborhoods of i-simplices σ^i of $\mathfrak{T}$, minus points in previously constructed handles of smaller index, and $\varepsilon_0 \gg \varepsilon_1 \gg \cdots \gg \varepsilon_k$. Let h: $S^k \to S^k$ be a map which collapses each set $\{x\} \times D^{k-i} \subset H^i$ to a point $h(x) \in \sigma^i$. Since h is homotopic to the identity, the map $t \mapsto g_t$ is homotopic to $t \mapsto g_{h(t)}$. We can go back now and replace the family g_t by the family $g_{h(t)}$, keeping the same functions φ_t, stratifications $\mathbb{S}$, $\mathbb{S}'$, etc. The net result is that we may assume the new family g_t is constant on slices $\{x\} \times D^{k-i}$ of the handles H^i, which will prove to be a useful property in Section 12.

PART II: Primitive Spheres

In the spherical models S_{tu}, the geometry of S^3 provided naturally the foliations F_{tu} by means of which the family S_{tu}^c was easily formed. For the family

Σ_{t0}, constructing foliations Φ_{t0} with qualitative behavior like F_{t0} is not so simple or natural. To obtain a polar Φ_{t0} on a factor $\bar{\Sigma}$ of $\bar{\Sigma}_{t0}$ the idea is to construct a monotone shrinking isotopy $\bar{\Sigma}_s$ of the ball $\bar{\Sigma}$ down to a point in its interior, then take Φ_{t0} on $\bar{\Sigma}$ to be transverse to the boundary spheres Σ_s. To obtain a facial Φ_{t0} on $\bar{\Sigma}$, we do a similar thing using a monotone shrinking of $\bar{\Sigma}$ to its preferred face. So the problem becomes finding shrinkings $\bar{\Sigma}_s$ for the primitive factors Σ of Σ_{t0} (in a way which varies continuously with t, of course). This is done by considering a natural 2-dimensional quotient $C(\Sigma)$ of $\bar{\Sigma}$, called the contour of Σ. In Section 5 we show that the contours of a family of primitive spheres have a very simple structure. In particular, contours of primitive spheres are shrinkable, and in Section 6 we obtain shrinkings of factors $\bar{\Sigma}$ by lifting shrinkings of their contours $C(\Sigma)$.

A difficulty arises however when one tries to piece together the resulting foliations Φ_{t0} for different strata of $\mathbb{S}_0'$, because such Φ_{t0}'s are not automatically compatible with the surgeries which occur in Σ_{t0} with varying t. An example illustrating this problem is given in Section 7. The solution is to choose the shrinkings of contours with much greater care. This is worked out in the more technical Sections 8–12, before the final Section 13 where we construct $\bar{g}_{t0}$: $\bar{\mathbb{S}}^c_{t0} \to \bar{\Sigma}^c_{t0}$.

5. Contours

For the rest of the paper we shall be dealing with embedded 2-spheres in $\mathbf{R}^3$ which have corners. But the corners will always be of the simplest sort: the bending locus will be a one-dimensional submanifold along which there are exactly two tangent planes. Further, the surface will be vertical, i.e., have vertical tangent planes, at all nearby points on one side of the bending locus.

If $\Sigma \subset \mathbf{R}^3$ is a 2-sphere (with corners) bounding the closed ball $\bar{\Sigma} \subset \mathbf{R}^3$, we define its *contour* $C(\Sigma)$ to be the quotient space $\bar{\Sigma}/\sim$, where $x \sim y$ if there is a vertical line segment in $\bar{\Sigma}$ joining x and y. In other words, $C(\Sigma)$ is the space of leaves of the foliation of $\bar{\Sigma}$ by vertical lines. The quotient map is denoted C: $\bar{\Sigma} \to C(\Sigma)$. (Note that C restricted to Σ is surjective.)

Example 5.1. In Figure 5.1 a primitive sphere is shown, whose contour is a disc with a "tongue".

A *disc with tongues* is a space C which is expressible as the union of finitely many 2-discs, $C = D_0 \cup D_1 \cup \cdots \cup D_n$, where for each $i > 0$, $D_i \cap (\bigcup_{j<i} D_j))$ is a subdisc d_i of D_i meeting ∂D_i in at least an arc. The topology on C is the natural quotient topology. Further, there is assumed to be present, if not always explicitly given, a projection map π: $C \to \mathbf{R}^2$ which is an embedding on each D_i, such that the discs $\pi(D_i)$ and $\pi(d_i)$ are smooth subdiscs of $\mathbf{R}^2$. As a set, C is the

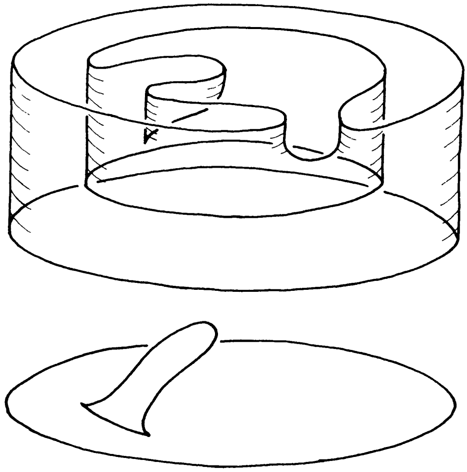


FIGURE 5.1

disjoint union of the *initial disc* D_0 with the *tongues* $T_i = D_i - d_i$. We call $\text{cl}(\partial D_i - \partial d_i)$ the *free edge* of T_i and $\text{cl}(\partial d_i - \partial D_i)$ the *attaching edge*. The union of the free and attaching edges of T_i is denoted ∂T_i. Points in both the free and attaching edges of T_i are *cusp points*. A disc-with-tongues *structure* on C is the decomposition of C into the initial disc D_0 and the (unordered) collection of tongues T_i. In particular, the discs D_i and d_i for $i > 0$ are not part of the structure, only their differences, the tongues. However, the existence of d_i in $D_0 \cup T_1 \cup \cdots \cup T_{i-1}$ imposes a non-trivial condition on how the tongue T_i can be attached.

We regard two disc-with-tongues structures as equivalent if one is obtained from the other by deleting empty tongues.

A tongue T_i need not be connected. In fact, ∂T_i can have infinitely many components. However, the finite character of a tongue is retained if it is regarded always as the difference between two discs. (Very likely, a preliminary normalization of the family $g_t(S^2)$ could be made to eliminate the necessity of considering tongues with infinitely many components, but there seems to be no particular advantage in doing this.)

A *family of discs with tongues* is a family of spaces each decomposed into an initial disc D_0 and finitely many tongues T_i, such that, locally in the parameter space:

(a) D_0 varies smoothly, when projected into $\mathbf{R}^2$.

(b) The discs D_i and d_i whose difference is a tongue T_i can be chosen to vary smoothly, when projected to $\mathbf{R}^2$, and $d_i \cap \partial D_i$ contains a smoothly varying arc.

For a family of tongues $T_t = D_t - d_t$, ∂T_t is defined by the condition that $\bigcup_t((\partial D_t - \partial d_t) \cup (\partial d_t - \partial D_t))$ have $\bigcup_t \partial T_t$ as its closure.

Locally in the parameter space the tongues can be attached in a definite order, but this order is not part of the data of a disc-with-tongues structure, and it need not be the case that one order of attaching tongues suffices globally in the parameter space.

If $\{D_0, T_i\}$ is a disc-with-tongues structure, we say a disc-with-tongues structure $\{D_{00}, T_{ij}\}$ on the same underlying space *subdivides* $\{D_0, T_i\}$ if, when we represent T_i as $D_i - d_i$, there are discs $d_i = D_{i0} \subset D_{i1} \subset \cdots \subset D_{in_i} = D_i$ such that $T_{ij} = D_{ij} - D_{i,j-1}$. When $i = 0$, d_0 is undefined, and $D_{00} \subset D_{01} \subset \cdots \subset D_{0n_0} = D_0$ gives a disc-with-tongues structure $\{D_{00}, T_{0j}\}$ on D_0.

A tongue T_i is of *type* I if:

(1) $\pi(\partial T_i) \cap \pi(\partial D_0) = \varnothing$, where D_0 is the initial disc.

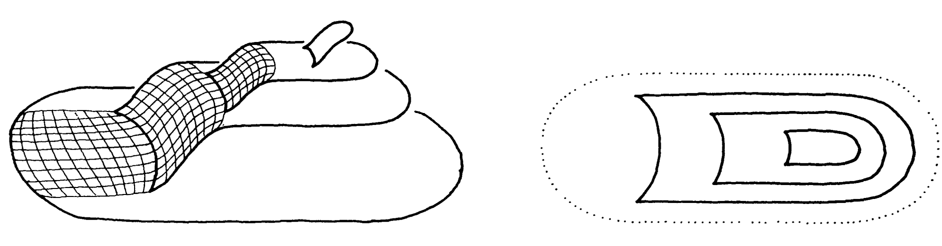
(2) $\pi(\partial T_i) \cap \pi(\partial T_j) = \varnothing$, for each tongue T_j, $j \neq i$.

Example 5.2. A disc with three type I tongues is shown in Figure 5.2a. (Ignore the shading for now.) The π-images of the boundaries of these tongues and of the initial disc are shown in Figure 5.2b.

A disc with type I tongues structure on a space is very nearly unique. The initial disc is certainly unique, and the only ambiguity arises from the possibility of regarding a tongue T_i for which ∂T_i is not connected as more than one tongue. In any case, two disc with type I tongues structures on the same space have a canonical common subdivision, whose tongues are the intersections of the tongues of one structure with the tongues of the other structure.

Here is the main result of this section:

PROPOSITION 5.1. *The contours of a family of primitive 2-spheres $\Sigma_t \subset \mathbf{R}^3$ have the structure of a family of discs with type* I *tongues, π: $C(\Sigma_t) \to \mathbf{R}^2$ being induced by vertical projection.*



(a) (b)

FIGURE 5.2

We will prove this only for a fixed value of the parameter t, leaving it to the reader to check what happens under (small) variations of t.

Before starting the proof, we make some preliminary observations. Let S be an elementary surface, with $\partial_- S$ denoting the components of ∂S near which S lies above ∂S, and $\partial_+ S$ the remaining components. Vertical projection $\pi: S \to \mathbf{R}^2$ restricts to an embedding on $\partial_+ S$ and on $\partial_- S$. Let $C_\pm = \pi(\partial_\pm S)$. These systems C_+ and C_- of disjoint simple closed curves in $\mathbf{R}^2$ satisfy the following conditions:

(i) C_+ meets C_- tangentially.

(ii) Locally, C_+ lies on only one side of C_-.

(iii) $C_+ \cup C_-$ is orientable; in fact, if C_- is oriented as the boundary of a compact subsurface of $\mathbf{R}^2$, this orientation extends over C_+.

Conditions (i) and (ii) follow immediately from condition (3) in the definition of elementary surface. To see (iii), note that $C_+ \cup C_-$ is contained on the π-image of the subset of S where S has vertical tangent planes, so a normal orientation for S in $\mathbf{R}^3$ projects to a normal orientation for $C_+ \cup C_-$. Hence $C_+ \cup C_-$ is also tangentially oriented. This orientation restricted to C_- bounds, in the sense of (iii), since we can enlarge S to a surface $S^* \subset \mathbf{R}^3$ by adding all points (x_0, y_0, z) such that either $(x_0, y_0, z_0) \in \partial_- S$ for some $z_0 \geq z$ or $(x_0, y_0, z_0) \in \partial_+ S$ for some $z_0 \leq z$. This S^* separates $\mathbf{R}^3$, and $C_- = S^* \cap (\mathbf{R}^2 \times \{z\})$ for $z \ll 0$.

Each component of $\mathbf{R}^2 - (C_+ \cup C_-)$ is bounded by a simple closed curve which is smooth except for a finite number (which is even, by (iii)) of outward-pointing cusps. For all but a finite number of these regions of $\mathbf{R}^2 - (C_+ \cup C_-)$, the number of cusps is two. The image $\pi(S)$ is the union of $C_+ \cup C_-$ with some of these regions of $\mathbf{R}^2 - (C_+ \cup C_-)$, including all regions whose boundaries have cusps (see Figure 1.1). In particular:

(iv) The unbounded component of $\mathbf{R}^2 - (C_+ \cup C_-)$ has smooth boundary.

(The reader may wish to convince himself that, conversely, given two systems C_+ and C_- of disjoint smooth circles in $\mathbf{R}^2$ which satisfy (i)–(iv), there is an elementary surface S with $C_\pm = \pi(\partial_\pm S)$.)

We label the closures of the components of $\mathbf{R}^2 - C_-$ with the symbols $+$ and $-$, the unbounded region being $+$ and adjacent regions having opposite signs. By condition (3) of Section 1, S can be expressed as the union of two compact subsurfaces S^+ and S^- which intersect each other in a finite number of vertical line segments, such that $\pi(S^+)$ is contained in the $+$ regions and $\pi(S^-)$ in the $-$ regions. Another way of saying this is that, if S is given the correct normal orientation, then normal vectors pointing in the direction of this orientation have z-coordinate ≥ 0 on S^+ and ≤ 0 on S^-. For example, for the elementary surface of Example 5.1, S^+ can be chosen to be the left half of S and S^- the right half; see Figure 5.3.

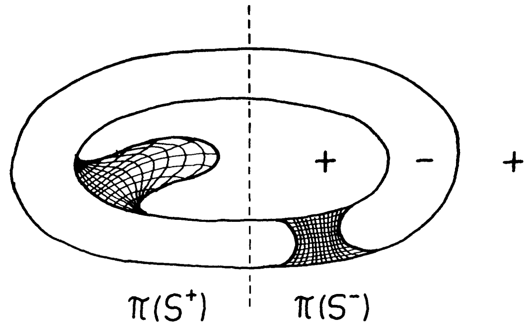
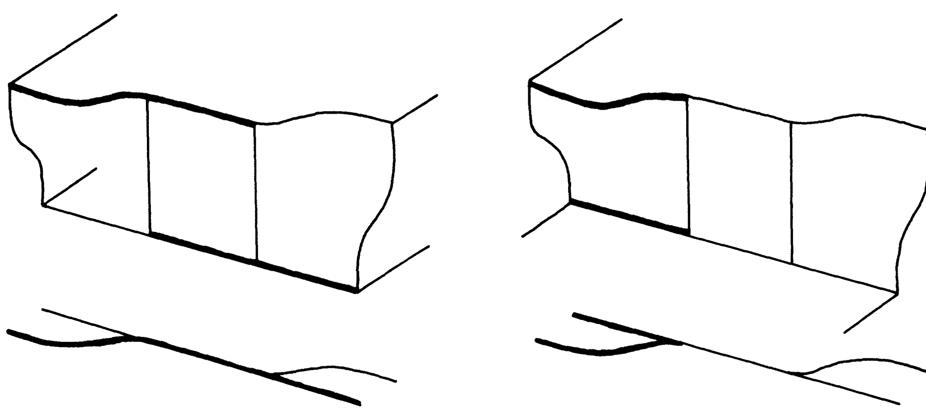


FIGURE 5.3

Now let Σ be a primitive surface, the union of an elementary surface S with horizontal discs, the faces of Σ. Such a face Δ is called *lower* if $\partial\Delta \subset \partial_- S$, *upper* if $\partial\Delta \subset \partial_+ S$. Further, Δ is *outer* if the corner angle of Σ at $\partial\Delta$, measured in $\overline{\Sigma}$, is 90°; Δ is *inner* if this angle is 270°.

Figure 5.4 illustrates the various possibilities for how an upper and lower face give rise to tangencies of C_+ and C_-.



(a) (b)

FIGURE 5.4

In (a), if the upper and lower faces shown are both outer faces, then the heavy line becomes a free edge of $C(\Sigma)$. In the opposite case that both faces in (a) are inner faces, then the heavy line becomes an attaching edge in $C(\Sigma)$. In both cases no cusp points are present. In (b), whether $\overline{\Sigma}$ lies on one side of the part of Σ shown or the other side, the heavy line becomes in $C(\Sigma)$ a free edge and an attaching edge meeting at a cusp; the attaching edge comes from the boundary of the inner face, the free edge from the boundary of the outer face. We conclude from this that $C(\Sigma)$ is locally a disc with type I tongues.

Let $C'(\Sigma) \subset C(\Sigma)$ be the set of non-regular (i.e., free edge, attaching edge, and cusp) points of $C(\Sigma)$. From Figure 5.4 we can also deduce that π restricts to an embedding of $C'(\Sigma)$ into $\mathbf{R}^2$, since above and below the part of S shown there can only be points of $\Sigma - S$.

Also, one can see from Figure 5.4 that free edge points of $C(\Sigma)$ project to $+$ regions of $\mathbf{R}^2$, attaching edge points to $-$ regions.

Proof of Proposition 5.1. The components of $S^{\pm} \cap \partial S$ we call *segments*. The proof will be by induction on the number of segments and faces of Σ.

Let Δ be a face of Σ such that $\pi(\partial\Delta)$ is an innermost circle of C_- if Δ is a lower face, or C_+ if Δ is an upper face; we call such a Δ *innermost*. For definiteness, let Δ be an innermost lower face. There is a well-defined finite set of disjoint discs $D_j \subset \pi(\Delta)$ such that:

(i) $\pi(S) \cap \pi(\Delta) = \pi(\Delta) - \bigcup_j \operatorname{int}(D_j)$.

(ii) Each ∂D_j is a (finite) union of π-images of segments, coming alternately from $\partial_+ S$ and $\partial_- S$ as one proceeds around ∂D_j.

A typical example is shown in Figure 5.5. We label the segments of $\partial\Delta$ alternately *inner* and *outer*, where if $\pi(\Delta)$ is a $\pm$ region, inner segments come from $S^{\pm}$ and outer segments from $S^{\mp}$. In Figure 5.5 the segment s_i is inner if i is even, outer if i is odd. Outer segments are contained in D_j's.

(1) If $\partial\Delta$ is a single outer segment, we can change Σ to another primitive Σ' by taking the elementary surface $S' \subset \Sigma'$ to be $S \cup \Delta$, with the corner at $\partial\Delta$ smoothed. Clearly $C(\Sigma) = C(\Sigma')$, and Σ' has one fewer segment than Σ. So we may assume $\partial\Delta$ has inner segments.

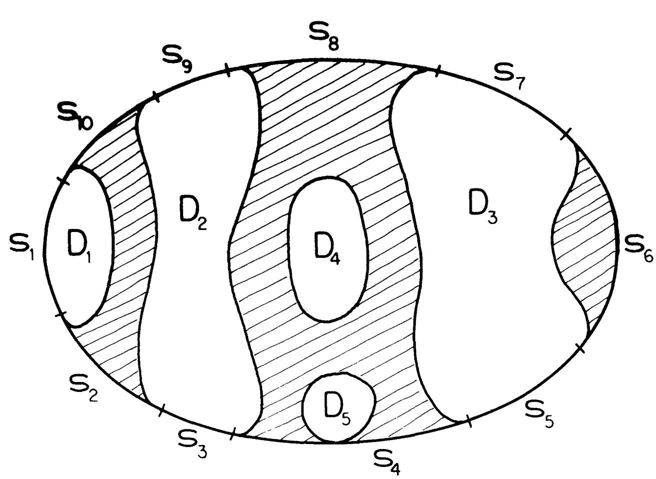
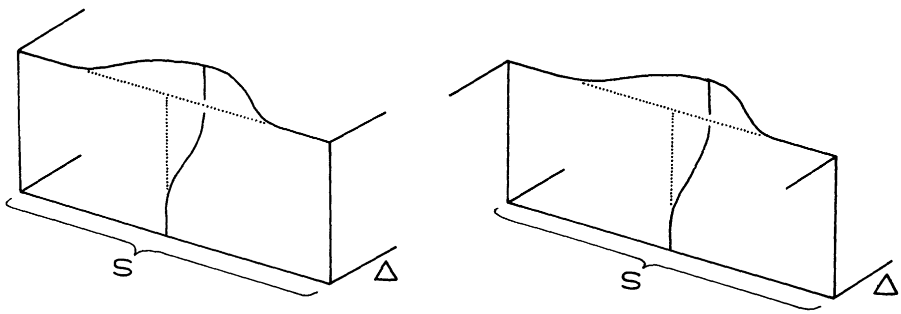


FIGURE 5.5

(2) If some D_j meets no outer segment, e.g., D_4 or D_5 in Figure 5.5, then $D_j = \pi(\Delta')$ for an upper face Δ' to which (1) is applicable. Hence we may assume each D_j meets, hence contains, an outer segment.

(3) If the number of D_j's in $\pi(\Delta)$ is zero (for example, if Σ has only one segment, $\partial\Delta$), then $\pi(\Delta) \subset \pi(S)$, and in fact $\pi(\Delta) = \pi(S)$ since $\partial\Delta$ is a single segment and $\pi(S)$ is connected (since S is). In this trivial case $C(\Sigma) = C(\Delta)$, a disc. So we may assume the number of D_j's is non-zero. By (2) this means $\partial\Delta$ has outer segments, as well as inner ones.

(4) Suppose there is an inner segment s of $\partial\Delta$ for which both adjacent outer segments (which may coincide) project into the same D_j, e.g., s_6 in Figure 5.5. There are two possible configurations, shown in Figure 5.6.



(a) (b)

FIGURE 5.6

We can change Σ to another primitive Σ' by replacing the component S_1 of $S^{\pm}$ containing s by a vertical strip S_1' above s (shown dotted). In Σ', S_1' can be taken out of $(S')^{\pm}$ and put in $(S')^{\mp}$, thereby amalgamating s with the adjacent segments of $\partial\Delta$. In case (a) $C(\Sigma) = C(\Sigma')$, while in (b) $C(\Sigma)$ is $C(\Sigma')$ with a tongue attached, of type I since we know π restricts to an embedding of the non-regular points of $C(\Sigma)$. So we may assume no segments s of the type just considered remain in $\partial\Delta$.

(5) If there is only one D_j in $\pi(\Delta)$, then (4) would apply. So we may assume there are at least two D_j's. Hence there are two D_j's, say D_1 and D_2, which contain the π-image of just one outer segment each, say s_1 and s_2. Let S_i $(i = 1, 2)$ be the component of $S^{\mp}$ containing s_i. Then S_i cannot meet any other (outer) segment of $\partial\Delta$, for if it did, ∂D_i would be the π-image of a non-separating circle $C \subset \Sigma$, as in Figure 5.7.

(6) If S_i meets another lower face Δ' with $\pi(\Delta) \cap \pi(\Delta') = \varnothing$, lying, we may assume, in the same horizonal plane as Δ, then Δ and Δ' can be amalgamated into a single lower face by pushing a neighborhood of an arc in S_i joining

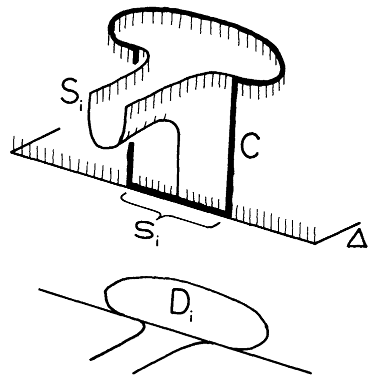
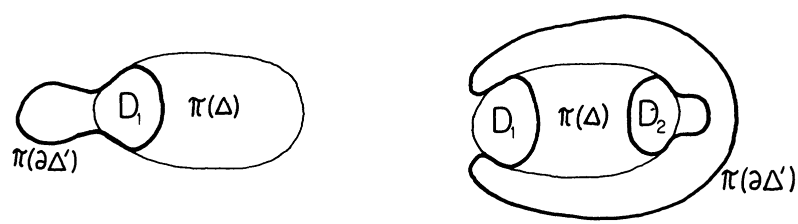


FIGURE 5.7

Δ and Δ' down to the level of Δ and Δ'. This yields a new primitive Σ' with the same contour as Σ but one fewer face.

(7) If both S_1 and S_2 meet a non-innermost lower face Δ', with $\pi(\Delta) \subset \pi(\Delta')$ necessarily, then just as in (5), Σ would contain a non-separating circle. Thus we may assume S_1, say, meets no other lower face besides Δ. For the upper face Δ' with $\partial D_1 - \pi(\partial\Delta) \subset \pi(\partial\Delta')$ there are two possibilities, shown in Figure 5.8:



(a) (b)

FIGURE 5.8

either (a) $D_1 \subset \pi(\Delta')$, or (b) $\pi(\Delta) - D_1 \subset \pi(\Delta')$. Case (b) reduces to case (a) by passing from S_1 to S_2. In case (a), by applying (1) to upper faces we may assume Δ' is innermost. Since D_1 is the only "D_j" for Δ', our previous steps apply to Δ' to reduce the number of segments of Σ. □

Remark. It seems likely that every disc with type I tongues occurs as the contour of some primitive 2-sphere.

We shall also need a relative form of Proposition 5.1, and this requires a new type of tongue. We say a tongue T_i in a disc with tongues structure

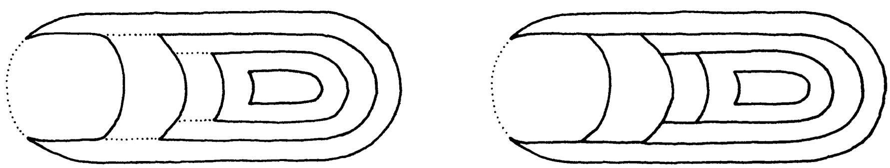
$\{D_0, T_1, \ldots, T_n\}$ is of *type* II if:

(1) The attaching edge of T_i lies in the initial disc D_0, and near its cusp points lies in ∂D_0.

(2) The free edge of T_i projects disjointly from $\pi(D_0)$ except for its cusp points.

(3) $\pi(\partial T_i) \cap \pi(\partial T_j) = \varnothing$ for $j \neq i$.

Example 5.3. For the disc with type I tongues shown in Figure 5.2a, if the shaded region is regarded as the initial disc, we obtain the structure of a disc with three type II tongues and one type I tongue. Figure 5.9a shows the π-images of these tongues. For comparison, Figure 5.9b superimposes Figures 5.2b and 5.9a.



(a) (b)

FIGURE 5.9

A disc with type I and II tongues structure enjoys the same uniqueness properties as a disc with type I tongues structure does, once the initial disc is specified.

A face Δ_t of a family of primitives Σ_t is called *large* if, locally in t, $\pi(\Delta_t) \cap \partial\pi(\Sigma_t)$ contains a smoothly varying arc. In other words, $C(\Delta_t)$ meets the boundary of the initial disc of $C(\Sigma_t)$ in at least an arc (which varies smoothly with t). A large face is necessarily an outer face.

PROPOSITION 5.2. *If Δ_t is a large face of the family of primitives Σ_t, then $C(\Sigma_t)$ has the structure of a disc with type* I *and* II *tongues, with $C(\Delta_t)$ as the initial disc.*

We shall refer to this disc-with-tongues structure as a structure on $C(\Sigma_t, \Delta_t)$, to distinguish it from the disc with type I tongues structure on $C(\Sigma_t)$ itself given by Proposition 5.1.

Proof. Consider the unparametrized case (the parametrized case will follow immediately). From Figure 5.4 and the fact that Δ is outer, $C(\partial\Delta)$ meets the attaching edges of $C(\Sigma)$ only in their cusp points, and near such a cusp point

$C(\partial\Delta)$ lies in the abutting free edge. Further, by condition (3) for elementary surfaces, $\pi(C(\partial\Delta)) \cap \pi(C'(\Sigma)) = \pi(C(\partial\Delta) \cap C'(\Sigma))$. The result follows by observing:

(1) If $C(\Delta)$ meets a type I tongue T of $C(\Sigma)$, then $T - C(\Delta)$ is a type II tongue attached to $C(\Delta)$.

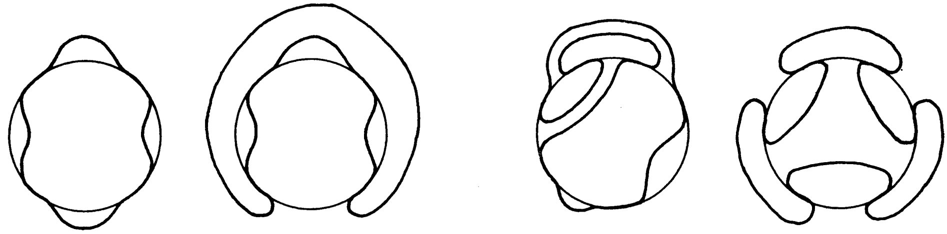
(2) For the initial disc D_0 of $C(\Sigma)$, $D_0 - C(\Delta)$ is a type II tongue attached to $C(\Delta)$.

(3) All type I tongues of $C(\Sigma)$ disjoint from $C(\Delta)$ remain type I tongues of $C(\Sigma, \Delta)$. □

For use in Section 9 we have the following:

LEMMA 5.3. *Let Σ be a primitive 2-sphere, with Δ_1 a lower face and Δ_2 an upper face. If $\pi(\partial\Delta_1) \cap \pi(\partial\Delta_2) \neq \varnothing$, then in the extended plane $\hat{\mathbf{R}} = \mathbf{R}^2 \cup \{\infty\}$, the components of $\hat{\mathbf{R}}^2 - (\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2))$ each have two cusps (necessarily pointing outside) except for two components whose boundaries are smooth circles.*

Thus two sorts of configurations for $\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2)$ are permitted, as shown in Figure 5.10a below. Excluded are configurations such as those in Figure 5.10b.



(a) (b)

FIGURE 5.10

Proof. Let α be a component arc of $\pi(\partial\Delta_1) \cap \text{int}(\pi(\Delta_2))$. (If no such arc exists, then $\pi(\Delta_2) \subset \pi(\Delta_1)$ or $\pi(\Delta_2) \subset \mathbf{R}^2 - \text{int}\,\pi(\Delta_1)$; in these cases the lemma is obvious.) Let α' be a copy of α pushed off $\pi(\partial\Delta_1)$ slightly, in the direction which makes the endpoints of α' lie outside $\pi(\Delta_2)$, so that $\alpha' \cap \pi(\Delta_2)$ is a closed subarc α'' of α'. There is a disc $D \subset \mathbf{R}^3$ with $\pi(D) = \alpha'$, ∂D consisting of four arcs: two vertical ones projecting to the endpoints of α' and two horizontal ones projecting to α', in levels just above the levels of Δ_1 and Δ_2. See Figure 5.11. By construction $\partial D \cap \Sigma = \varnothing$, and we may suppose D meets Σ transversely. One component of $D \cap \Sigma$ is a circle C with $\pi(C \cap \Delta_2) = \alpha''$. Since Σ is a 2-sphere, C

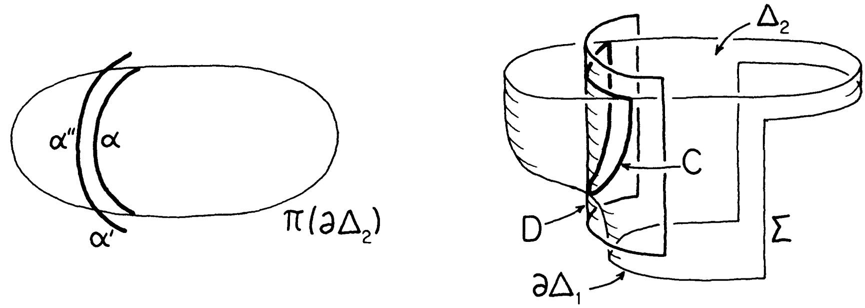


FIGURE 5.11

must separate Σ. It follows that one component of $\pi(\Delta_2) - \alpha''$ must be disjoint from $\pi(\partial\Delta_1)$, and that the component of $\mathbf{R}^2 - (\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2))$ containing α'' has just two cusps on its boundary.

To finish the proof there are two cases:

Case I. *Both* Δ_1 *and* Δ_2 *are outer faces, or both are inner. This means that* $\pi(\Delta_1)$ *and* $\pi(\Delta_2)$ *lie on the same side of* $\pi(\partial\Delta_1) \cap \pi(\partial\Delta_2)$. *In this case what we have shown is that the components of* $\pi(\Delta_2) - \pi(\Delta_1)$ *are 2-cusped. If we remove them from* $\pi(\Delta_2)$, *we are left with a smooth subdisc of* $\pi(\Delta_1)$. *In effect, this reduces us to the easy case* $\pi(\Delta_2) \subset \pi(\Delta_1)$.

Case II. *One of* Δ_1 *and* Δ_2 *is inner and the other is outer. Now the components of* $\pi(\Delta_2) \cap \pi(\Delta_1)$ *are 2-cusped, and removing them from* $\pi(\Delta_2)$ *reduces us to the easy case* $\pi(\Delta_2) \subset \mathbf{R}^2 - int\, \pi(\Delta_1)$. □

6. Shrinking

Let C_t be a family of discs with tongues. By a *shrinking* of C_t we mean a family C_{ts} of discs with tongues, $s \in [0, 1]$, such that:

(1) $C_{t0} = C_t$.
(2) $C_{ts} \subset C_{ts'}$ if $s > s'$.
(3) For each tongue T_t of C_t, $T_t \cap C_{ts}$ is a tongue of C_{ts}.
(4) If D_t is the initial disc of C_t, then $D_t \cap C_{ts}$ is the initial disc of C_{ts}.

LEMMA 6.1. *Let* $\Sigma_t \subset \mathbf{R}^3$ *be a family of 2-spheres with corners* (*of the type described in* § 5) *such that* $C(\Sigma_t)$ *is a family of discs with tongues. Then a shrinking* C_{ts} *of* $C(\Sigma_t) = C_{t0}$ *lifts to an isotopy* Σ_{ts} *of* $\Sigma_t = \Sigma_{t0}$ *such that* (i) $C(\Sigma_{ts}) = C_{ts}$, (ii) $\overline{\Sigma}_{ts} \subset \overline{\Sigma}_{ts'}$ *if* $s > s'$, *and* (iii) Σ_{ts} *is smooth for* $s > 0$.

Proof. It suffices to consider the case that the tongues of C_t attach in one order independent of t. For if $\{\psi_i\}$, $i = 1, 2, \ldots$, is a partition of unity subordinate to a cover of the parameter domain in each of whose sets a single order of attaching is possible, then we can lift the shrinking C_{ts} inductively over the intervals $\Sigma_{i<l}\psi_i \leq s \leq \Sigma_{i\leq l}\psi_i$; in this l^{th} interval we use the order associated to the support of ψ_l.

Suppressing t from the notation, let $T_1, \ldots, T_n$ be the tongues of C, attached in that order, and let D be the initial disc of C. From the shrinking C_s we define a family of discs with tongues $C(s_0, \ldots, s_n) \subset C$, for $0 \leq s_0 \leq \cdots \leq s_n \leq 1$, by the conditions $C(s_0, \ldots, s_n) \cap T_i = C_{s_i} \cap T_i$ and $C(s_0, \ldots, s_n) \cap D = C_{s_0} \cap D$. We view $C(0, \ldots, 0, s_i, s_{i+1}, \ldots, s_n)$ for $s_i \in [0, s_{i+1}]$ as a shrinking of $C(0, \ldots, 0, s_{i+1}, \ldots, s_n)$. If we assume inductively that a family $\Sigma(0, \ldots, 0, s_{i+1}, \ldots, s_n)$ has already been constructed with contour equal to $C(0, \ldots, 0, s_{i+1}, \ldots, s_n)$, it suffices to lift the shrinking $C(0, \ldots, 0, s_i, s_{i+1}, \ldots, s_n)$, $s_i \in [0, s_{i+1}]$, to a shrinking isotopy $\Sigma(0, \ldots, 0, s_i, s_{i+1}, \ldots, s_n)$ of $\Sigma(0, \ldots, 0, s_{i+1}, \ldots, s_n)$. For the desired shrinking Σ_s is then $\Sigma(s, \ldots, s)$.

Thus we have reduced to two cases: either only a single tongue $T = T_i - C_{s_{i+1}}$ is shrinking (and no other tongues attach to this tongue), or only the initial disc is shrinking. We consider the first case, the second case being similar. As a preliminary to lifting such a shrinking C_s, we construct a small shrinking isotopy Σ'_s of $\Sigma = \Sigma'_0$ such that:

(a) Σ'_s is smooth for $s > 0$.

(b) Σ'_s has no vertical tangents at points projecting to $\text{int}(T)$.

(c) $C(\Sigma'_s) = C(\Sigma)$.

Figure 6.1 indicates how to achieve (a), where the shaded region is $\bar{\Sigma}$.

FIGURE 6.1

To achieve (b), let V be the set of points of Σ having vertical tangent planes and projecting to points of $\text{int}(T)$. Let V_+ (V_-) be the subset of V where an upward (downward) directed vertical line passes from inside Σ to outside Σ, as in Figure 6.2. Notice that V_+ and V_- are disjoint, relatively closed sets in $C^{-1}(\text{int}(T))$, where $C\colon \Sigma \to C(\Sigma) = C$ is the quotient map. So we can find a vector field v on $\mathbf{R}^3$ (smooth, and varying smoothly with t) such that:

(i) $v|\Sigma$ is non-zero only on $C^{-1}(\text{int}(T))$.

(ii) $v|\Sigma$ is orthogonal to Σ, pointing inside where it is non-zero.

(iii) $\frac{\partial}{\partial z}|v| > 0$ on V_+ and $\frac{\partial}{\partial z}|v| < 0$ on V_-.

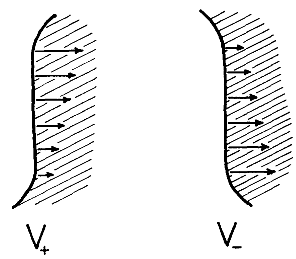


FIGURE 6.2

Having v, we obtain Σ_s' by flowing along v for a sufficiently short interval $s \in [0, \varepsilon]$, then standing still for $s \in [\varepsilon, 1]$.

The desired shrinking Σ_s of Σ is obtained by setting $\overline{\Sigma}_s = C^{-1}(C_s)$, where now $C\colon \overline{\Sigma}_s' \to C$ is the projection. This Σ_s has corner points which project to the free edge of $T \cap C_s$, but these can easily be smoothed without affecting the contour, as in the left half of Figure 6.1. □

Obvious modifications of the preceding proof yield:

Addendum 6.2. If Σ_t contains a horizontal disc Δ_t with $\partial\Delta_t$ a corner of Σ_t, then a shrinking of $C(\Sigma_t)$ to $C(\Delta_t)$ lifts to a monotone isotopy of $\Sigma_t - \text{int}(\Delta_t)$, rel $\partial\Delta_t$, across $\overline{\Sigma}_t$ to Δ_t.

If T_t is a family of tongues, the difference between discs $d_t \subset D_t$, then by a *shrinking of* T_t we mean a shrinking of the disc-with-tongues structure $D_t = d_t \cup T_t$ down to its initial disc d_t. Thus a shrinking of T_t amounts to a monotone family of subtongues of T_t whose attaching edges are contained in the attaching edge of T_t. Different shrinkings of T_t which yield the same monotone family of subtongues will be regarded as equivalent; they differ only in the rate at which the shrinking occurs.

LEMMA 6.3. *Given shrinkings of the individual tongues of a family of discs with tongues C_t, there is a shrinking of C_t down to its initial disc, whose restriction to each tongue of C_t is the given shrinking of that tongue.*

Proof. As in the proof of Lemma 6.1, a partition of unity argument easily reduces to the case that the tongues of C_t attach in a single order, independent of t. Then we can just shrink the tongues of C_t one at a time, in the reverse of the order of attaching, using the given shrinkings of these tongues. □

7. An example

Suppose a component $\Sigma(\gamma) = \Sigma_{t0}(\gamma)$ of Σ_{t0} has a single common horizontal face Δ_1, which factors the primitive $\Sigma \subset \Sigma(\gamma)$ as the difference of two primitives $\Sigma_0, \Sigma_1 \subset \Sigma(\gamma)$ (see Figure 7.1).

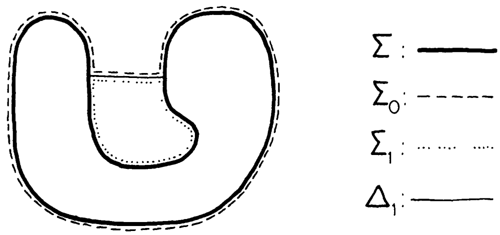
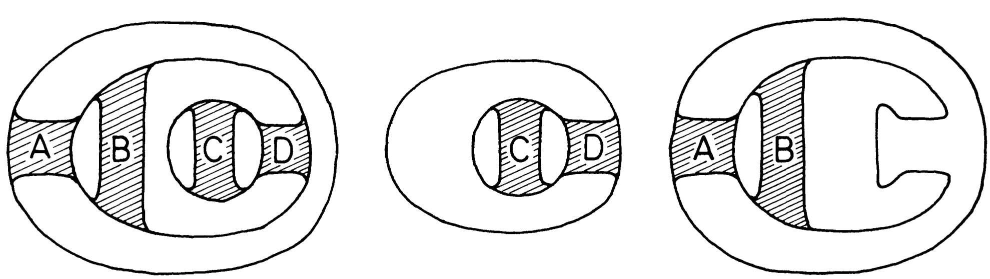


FIGURE 7.1

We might hope that our prescription for constructing the foliations Φ_{t0} on $\overline{\Sigma}$, $\overline{\Sigma}_0$, and $\overline{\Sigma}_1$ mentioned in the introduction to Part II, as transversals to shrinkings of $\overline{\Sigma}$ and $\overline{\Sigma}_0$ to a point and $\overline{\Sigma}_1$ to Δ_1, would have the property that the restriction to $\overline{\Sigma}$ of Φ_{t0} on $\overline{\Sigma}_0$ would equal Φ_{t0} on $\overline{\Sigma}$. This happens in the spherical models, and it would happen here if there were some sort of compatibility in the shrinkings of $\overline{\Sigma}$, $\overline{\Sigma}_0$, and $\overline{\Sigma}_1$. However, shrinking $\overline{\Sigma}_1$ to Δ_1 has the effect of *expanding* $\overline{\Sigma}$ rather than shrinking it. At the very least, one would want $C(\Sigma)$ to keep its structure of a disc with tongues as $\overline{\Sigma}$ expands. The point of this section is to show an example where $C(\Sigma)$ loses its disc-with-tongues structure as $\overline{\Sigma}$ expands, and to suggest the remedy: subdividing tongue structures before shrinking.

We remark that this problem does not arise in a generic family $g_t\colon S^2 \to \mathbf{R}^3$, $t \in S^k$ with $k \leq 2$, since the disc-with-tongue contours which one has to consider in these cases are very simple. (The projection $\pi\colon C(\Sigma) \to \mathbf{R}^2$ is at most 2-to-1.) Hence our proof could be considerably shortened if one were interested only in showing $\pi_i \mathrm{Diff}(S^3) \approx \pi_i O(4)$ for $i = 0, 1$.

The elementary surface part S of the primitive Σ of our example can be described, as in Section 5, by its image $\pi(S)$ shown in Figure 7.2a. Thus $\partial_- S$



(a) $\pi(S)$ (b) $\pi(S_1)$ (c) $\pi(S_0)$

FIGURE 7.2

consists of three circles C_1, C_2, C_3 with nested π-images. Four handles A, B, C, D are attached to the top edges of vertical collars on C_1, C_2, and C_3. These handles have the effect of doing surgery on $C_1 \cup C_2 \cup C_3$ to obtain three other circles. Adding vertical collars above these three circles, we reach the three circles C_4, C_5, C_6 of $\partial_+ S$, whose π-images are also nested. A full picture of S is attempted in Figure 7.3.

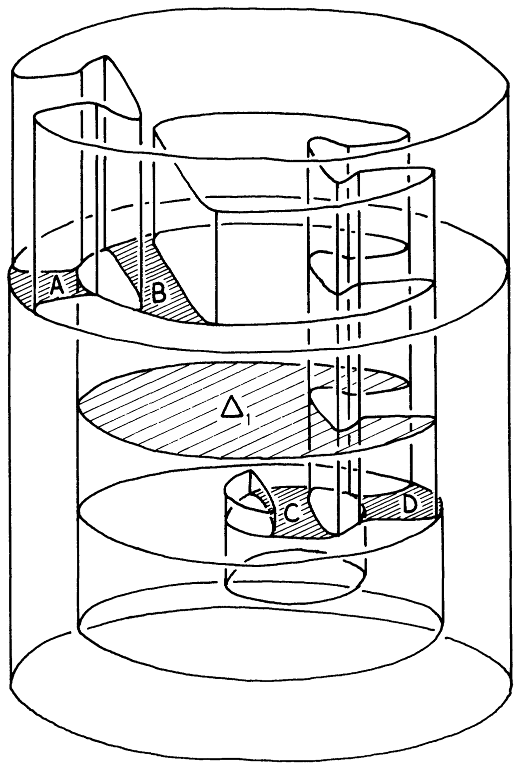
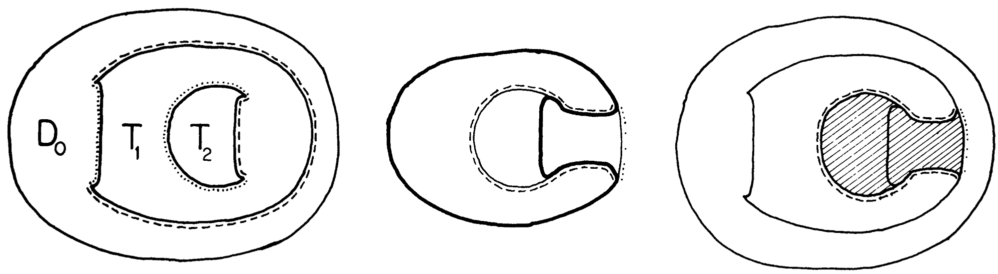


FIGURE 7.3

We obtain the factorization $\Sigma = \Sigma_0 - \Sigma_1$ by putting the handles A and B in a level above the level of C and D, and choosing $\partial\Delta_1 \subset S$ as a horizontal circle in an intermediate level, such that $\partial\Delta_1$ separates A and B from C and D on S. The elementary surface parts $S_0 \subset \Sigma_0$ and $S_1 \subset \Sigma_1$ are shown in Figure 7.2b, c. Note that S_0 and S_1 are essentially isomorphic.

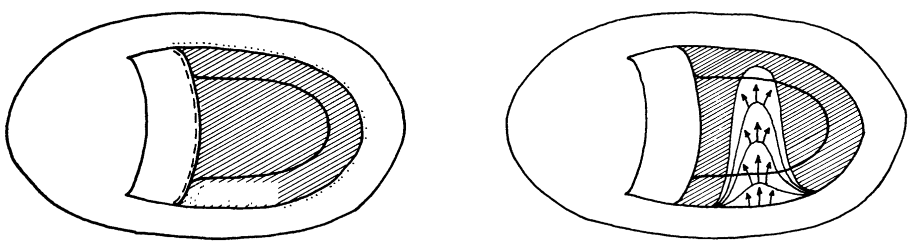
From Figure 7.3, or by more general considerations using Figure 7.2a, one can see that $C(\Sigma)$ is a disc D_0 with a type I tongue T_1 attached to the top side of D_0, then a type I tongue T_2 attached to the bottom side of T_1. Figure 7.4a shows the π-images, with free edge points of T_1 and T_2 paralleled by dotted lines and attaching edge points by dashed lines. Also we see that $C(\Sigma_1, \Delta_1)$ consists of a single type II tongue T_3 attached to the underside of the disc $C(\Delta_1)$. This is shown in Figure 7.4b, $\partial C(\Delta_1)$ being indicated by the heavy line.



(a) $C(\Sigma)$ (b) $C(\Sigma_1, \Delta_1)$ (c)

FIGURE 7.4

Figure 7.4c shows the π-images of D_0, T_1, and T_2, with $\pi(T_3)$ superimposed (shaded). A diffeomorphic but visually simpler form of this configuration is shown in Figure 7.5a.



(a) (b)

FIGURE 7.5

Suppose we begin to shrink T_3 in the way shown in Figure 7.5b. At first during this shrinking of T_3, the effect on $C(\Sigma)$ is to pinch T_1 to D_0, and what is left of T_1 remains a tongue. But then, when the part of T_3 being shrunk has π-image meeting $\pi(T_2)$, we are pinching T_2 to D_0. When all that is left of T_3 is the shaded part in Figure 7.5b, then what remains of T_2 has two components, which are both discs topologically. One of these (on the right in Figure 7.5b) has no free edge, so it cannot be a tongue. (A further difficulty is that what remains of T_1 is not even simply-connected, so it cannot be a tongue either.)

The way to avoid this problem is to use Figure 7.5a to subdivide T_1 into three tongues T_{11}, T_{12}, T_{13} and to subdivide T_3 into two tongues T_{31} and T_{32}, as indicated in Figure 7.6. The tongue T_2 we leave unsubdivided. Then shrinking T_{32} pinches T_{11} bit by bit to D_0, T_{11} remaining a tongue throughout the process. And similarly, shrinking T_{31} pinches T_2 to T_{12} so that it remains a tongue throughout too.

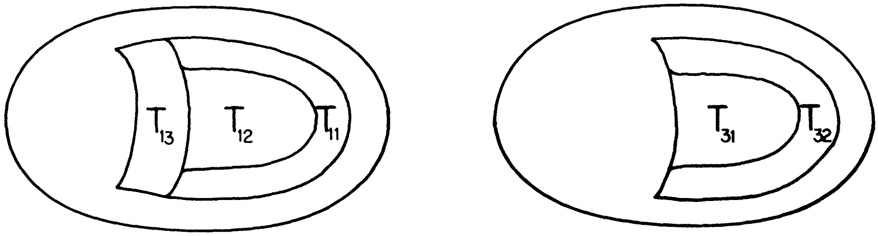


FIGURE 7.6

8. Large factors

Consider a family $\Sigma(\gamma) = \Sigma_t(\gamma) = \Sigma_{t0}(\gamma) \subset \Sigma_{t0}$, as defined in Section 2. Thus $\Sigma(\gamma)$ is the union of a primitive 2-sphere Σ with certain horizontal discs $\Delta_1, \ldots, \Delta_n$ which split Σ into primitive factors $\Sigma_0, \ldots, \Sigma_n$. The Δ_i's correspond to the edges e_i of γ, the Σ_i's to the vertices v_i. The maximal factors Σ_i correspond to the vertices of a subtree γ_0 of γ. The edges of γ_0 correspond to sum Δ_i's.

A family of subsets $X_t \subset \Sigma_t(\gamma)$ is called *large* if, locally in t, $\pi(X_t) \cap \partial\pi(\Sigma_t(\gamma))$ contains a smoothly varying arc. (Note that $\pi(\Sigma_t(\gamma))$ is a disc since Σ is primitive.)

LEMMA 8.1. *The large factors Σ_i of $\Sigma_t(\gamma)$ correspond to the vertices of a non-empty subtree $\lambda_t \subset \gamma_0 \subset \gamma$ (λ_t can vary with t).*

Proof. We shall use the notation and terminology introduced to prove Proposition 5.1. First we verify that large factors exist. The boundary of the disc $\pi(\Sigma)$ is the union of finitely many arcs lying in either $\pi(S^+)$ or $\pi(S^-)$. Let α be such an arc, say in $\pi(S^+)$. Then $\partial_+ S$ contains an arc $\tilde{\alpha}$ with $\pi(\tilde{\alpha}) = \alpha$. The component of $\partial_+ S$ containing $\tilde{\alpha}$ lies in some factor Σ_i, which is therefore large.

Large factors of $\Sigma(\gamma)$ are clearly maximal.

Let Σ_0 and Σ_1 be large factors and suppose Δ_1 is a (sum) face corresponding to an edge of γ which lies on the path in γ joining the vertices v_0 and v_1. We claim that Δ_1 is large. From this it follows that the two factors of $\Sigma(\gamma)$ having Δ_1 as their common face are also large. Hence all the vertices of γ on the path from v_0 to v_1 correspond to large factors, so the vertices of γ corresponding to large factors are the vertices of a subtree of γ.

To see that Δ_1 is large, we may as well assume that Σ_0 and Σ_1 are the only factors of $\Sigma(\gamma)$, since discarding the other faces $\Delta_2, \ldots, \Delta_n$ can only enlarge $\pi(\Sigma_0)$ and $\pi(\Sigma_1)$. Say Δ_1 is an upper face of Σ_0 and a lower face of Σ_1. Suppose $\pi(\partial\Delta_1)$ were not an outermost circle of $\pi(\partial_- S_1)$. Then $\pi(\Delta_1) \subset \operatorname{int}(\pi(\Sigma_1))$, and in the contour $C(\Sigma) = C(\Sigma_0) \cup C(\Sigma_1)$, with $C(\Sigma_0) \cap C(\Sigma_1) = C(\Delta_1)$, the initial disc of $C(\Sigma_1)$ would also be the initial disc of $C(\Sigma)$. Hence Σ_0 could not be large, or else $C(\Sigma)$ would not be a disc with type I tongues. So $\pi(\partial\Delta_1)$ is an outermost circle of $\pi(\partial_- S_1)$, and likewise of $\pi(\partial_+ S_0)$. If $\partial\Delta_1 \subset S^+$, then $\pi(\Sigma_0) \subset \pi(\Delta_1)$, so Δ_1 is large since Σ_0 is. Similarly, Δ_1 is large if $\partial\Delta_1 \subset S^-$. So we may assume $\partial\Delta_1$ meets both S^+ and S^-. Considering Σ_0 and Σ_1 separately, we see that

$$\pi(\partial\Delta_1 \cap S^+ \cap S^-) \subset \partial\pi(\Sigma_0) \cap \partial\pi(\Sigma_1) \subset \partial\pi(\Sigma).$$

Thus $\pi(\Delta_1)$ meets $\partial\pi(\Sigma)$, in at least an arc in fact, since we may take S^+ and S^- to be vertical in a neighborhood of $S^+ \cap S^-$. See Figure 8.1. □

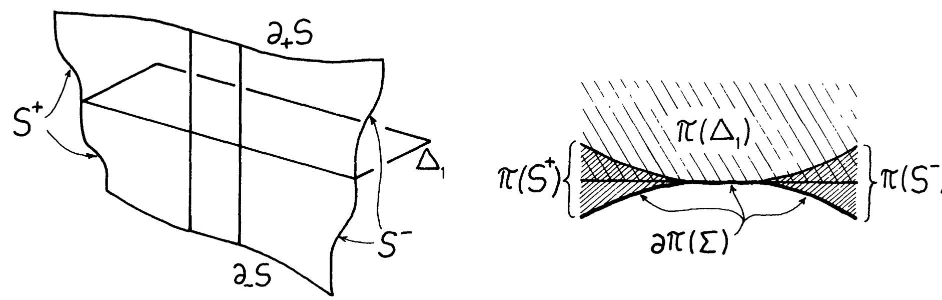


FIGURE 8.1

Let us restrict now to an open subset of the parameter domain over which some factor Σ_0 of $\Sigma(\gamma)$ is large. The other factors we relabel so that Δ_i is a face of Σ_i for $i = 1, \ldots, n$. Each Δ_i splits Σ as a sum or difference of two primitives Σ^i and ${}^i\Sigma$, where $\Sigma^i \cup {}^i\Sigma = \Sigma \cup \Delta_i$ and $\Sigma^i \cap {}^i\Sigma = \Delta_i$. One of Σ^i and ${}^i\Sigma$, say Σ^i, meets Σ_i in more than just the face Δ_i. We call Σ^i a *cofactor* of $\Sigma(\gamma)$. For convenience we regard Σ as the cofactor Σ^0. The cofactors Σ^i are partially ordered by defining $\Sigma^i < \Sigma^j$ to mean $\Sigma^i - \Delta_i \subset \Sigma^j$. In particular $\Sigma^i < \Sigma^0$ for all $i > 0$. (Cofactors turn out to be more useful than factors in constructing $\overline{\Sigma}^c_{t0}$ and $\bar{g}_{t0}\colon \bar{S}^c_{t0} \to \overline{\Sigma}^c_{t0}$.)

LEMMA 8.2. *Δ_i is a large face of Σ^i; that is, $\pi(\Delta_i) \cap \partial\pi(\Sigma^i)$ contains an arc which, locally in t, can be chosen to vary smoothly with t.*

Proof. Since the Δ_j's with $j \neq i$ are irrelevant here, we may as well assume $n = i = 1$, so that Δ_1 splits Σ into the factors Σ_0 and Σ_1 with Σ_0 large. Say Δ_1 is an upper face of Σ_0 and a lower face of Σ_1. Suppose $\pi(\partial\Delta_1)$ were not an outermost circle of $\pi(\partial_- S_1)$. If Δ_1 is a sum face, we reach a contradiction just as in the proof of Lemma 8.1. If Δ_1 is a difference face, then the boundary of the initial disc of $C(\Sigma_1)$ would create a circle of attaching edge points in $C(\Sigma)$, which is impossible in a disc with type I tongues. Hence $\pi(\partial\Delta_1)$ must be an outermost circle of $\pi(\partial_- S_1)$.

Suppose now that Δ_1 is a sum face. (The argument from here on in the difference case is just the same, but with S^+ and S^- interchanged.) If $\partial\Delta_1 \subset S^+$, then $\pi(\Sigma_0) \subset \pi(\Delta_1)$, so Δ_1 is large in $\Sigma(\gamma)$ since Σ_0 is, and *a fortiori* Δ_1 is a large face of Σ_1. So we may assume $\partial\Delta_1$ meets S^-; hence $\partial\Delta_1 \cap S^-$ contains an arc. Since $\pi(\partial\Delta_1 - S^+) \subset \partial\pi(\Sigma_1)$, Δ_1 is then a large face of Σ_1. □

As an application of these lemmas, we refine condition (2) in the construction of $F_{tu}(\gamma)$ in Section 3 to be:

(2′) $F_{tu}(\gamma)$ has a pole in the core of a maximal factor of $S_{tu}(\gamma)$ *(and moreover, a maximal factor which corresponds to a large factor of* $\Sigma_{t0}(\gamma)$ *if* $u = 0$) unless (t, u) is near the boundary of the stratum where γ is defined.

Lemma 8.2 then guarantees that the factor S_0 of $S_{t0}(\gamma)$ in (a) and (c) of Lemma 3.1 corresponds to a large factor of $\Sigma_{t0}(\gamma)$.

Thus for a stratum of $\mathbb{S}_0'$ over which γ has a base vertex (see § 3), this vertex determines a large factor Σ_0 of $\Sigma(\gamma)$. In the opposite case that γ has a base edge, both factors of $\Sigma(\gamma)$ containing the common face Δ_i corresponding to this edge are large, as is Δ_i itself, as we showed above. The most convenient thing to do in this case is to choose for "Σ_0" the cylinder between two infinitesimally displaced parallel copies of Δ_i.

9. Tongue patterns

We continue with the situation of Section 8.

Each subtree τ of γ determines a primitive 2-sphere $\Sigma_\tau \subset \Sigma(\gamma)$, which is the union of the factors Σ_i corresponding to the vertices of τ, minus the interiors of the faces Δ_i corresponding to the edges of τ. In case the base vertex or edge of γ does not lie in τ, Σ_τ has a preferred face Δ_τ, which is the Δ_i corresponding to the edge of $\gamma - \tau$ abutting τ in the direction of the base vertex or edge of γ. Note that $\Delta_\tau = \Delta_i$ is a large face of Σ_τ since $\pi(\Sigma_\tau) \subset \pi(\Sigma^i)$ and Δ_i is a large face of Σ^i by Lemma 8.2.

The main purpose of this section is to describe a certain compatibility relationship which holds among all the disc-with-tongue structures $C(\Sigma_\tau)$ and $C(\Sigma_\tau, \Delta_\tau)$ as τ ranges over subtrees of γ. This compatibility will be measured via the projections $\pi\colon C(\Sigma_\tau) \to \mathbf{R}^2$ and $\pi\colon C(\Sigma_\tau, \Delta_\tau) \to \mathbf{R}^2$.

Some definitions are needed first. A subset P of $\mathbf{R}^2$ is called a *tongue pattern* if it is the union of a finite number of disjoint subsets P_i called *tongue blocks* or just *blocks*, each of which has the form $P_i = \bigcup_j \partial T_{ij}$, where the T_{ij}'s are the tongues of a subdivision of a single tongue $T_i \subset \mathbf{R}^2$. (See the definition of subdivision in §5.) A *family of tongue patterns* is obtained by taking families of tongues T_i and T_{ij}, as in Section 5.

As a simple example, if Σ is a primitive 2-sphere and $T_1, \dots, T_n$ are the type I tongues of $C(\Sigma)$, then $P(\Sigma) = \bigcup_i \pi(\partial T_i)$ is a tongue pattern. Similarly, if Δ is a large face of Σ and $T_1, \dots, T_n$ are the type I and II tongues of $C(\Sigma, \Delta)$, then $P(\Sigma, \Delta) = \bigcup_i \pi(\partial T_i)$ is a tongue pattern. Less trivially, $P(\Sigma) \cup P(\Sigma, \Delta)$ is a tongue pattern. This can be deduced from Proposition 9.1 below, or seen directly. For example, the solid lines in Figure 5.9b show $P(\Sigma) \cup P(\Sigma, \Delta)$ where Σ has contour shown in Figure 5.2a, with $C(\Delta)$ the shaded disc.

A tongue block P_i can be characterized as follows. Each component of P_i is obtained from a component of the boundary ∂T_i of a tongue $T_i \subset \mathbf{R}^2$ by adding a bounded number n of arcs $\alpha_l \subset T_i \cup \partial T_i$, with $\partial \alpha_l \subset \partial T_i$, such that the components of $T_i - \bigcup_l \alpha_l$ are all 2-cusped. For, given such arcs α_l, each α_l subdivides T_i into two subtongues, and the various n-fold intersections of these subtongues form the tongues T_{ij} of a subdivision of T_i.

We note also that different subdivisions $\{T_{ij}\}$ and $\{T'_{ij'}\}$ of T_i can have $P_i = \bigcup_j \partial T_{ij} = \bigcup_{j'} \partial T'_{ij'}$. However, in this case $\{T_{ij} \cap T'_{ij'}\}$ is a common subdivision with $P_i = \bigcup_{j, j'} \partial(T_{ij} \cap T'_{ij'})$.

Returning now to the situation at the beginning of this section, we define

$$P(\gamma) = \bigcup_{\tau \subset \gamma} \left(P(\Sigma_\tau) \cup P(\Sigma_\tau, \Delta_\tau)\right),$$

the union over all subtrees τ of γ, it being understood that the term $P(\Sigma_\tau, \Delta_\tau)$ is omitted when it is undefined, namely when τ contains the base vertex or edge of γ.

Proposition 9.1. *$P(\gamma)$ is a family of tongue patterns.*

Proof. Consider first a fixed value of t. Let $P^+(\Sigma) = P(\Sigma) \cup \pi(\partial D)$ where D is the initial disc of $C(\Sigma)$. Intrinsically, $P^+(\Sigma)$ is the projection to $\mathbf{R}^2$ of all the non-regular points of $C(\Sigma)$, i.e., points where $C(\Sigma)$ is not locally homeomorphic to $\mathbf{R}^2$. Let $P^+(\gamma) = P(\gamma) \cup P^+(\Sigma)$. $P^+(\gamma)$ is orientable, via a continuous field of non-vanishing tangent vectors. Namely, $P^+(\gamma)$ is contained in the π-image of the subset of Σ where Σ has vertical tangent planes; so a normal orientation for Σ projects to a normal orientation for $P^+(\gamma)$.

Before beginning the proof proper, we make three preliminary observations. Let $S(\Sigma_\tau, \Delta_\tau) \subset \pi(\partial \Delta_\tau)$ consist of points of $\pi(\partial \Delta_\tau)$ lying on or inside blocks of $P(\Sigma_\tau, \Delta_\tau)$. Thus $S(\Sigma_\tau, \Delta_\tau) \subset P(\Sigma_\tau) \cup P(\Sigma_\tau, \Delta_\tau)$.

(1) $P^+(\gamma) - P^+(\Sigma) \subset \bigcup_i S(\Sigma^i, \Delta_i)$. To see this, consider $x \in P^+(\gamma) - P^+(\Sigma) \subset P(\gamma)$, say $x \in P(\Sigma_\tau) \cup P(\Sigma_\tau, \Delta_\tau)$. From Figure 5.4 we see that $x \in \pi(\partial \Delta)$ for some horizontal face Δ of Σ_τ. (There are at most two faces Δ of Σ_τ such that $x \in \pi(\partial \Delta)$, since Σ_τ is primitive.) If $x \notin \pi(\partial \Delta_i)$ for any i, then $x \in P(\Sigma_\tau)$

(since $P(\Sigma_\tau, \Delta_\tau) - P(\Sigma_\tau) \subset \pi(\partial\Delta_\tau)$ and $\Delta_\tau = \Delta_i$ for some i), and x lifts to a non-regular point of $C(\Sigma_\tau)$ in $\partial\Delta$ which remains non-regular in $C(\Sigma)$ (see Figure 5.4); hence $x \in P^+(\Sigma)$, a contradiction. So we may assume $x \in \pi(\partial\Delta_i)$ with Δ_i a face of Σ_τ; this can happen for at most two Δ_i's. If either of these Δ_i's is Δ_τ, then $x \in S(\Sigma_\tau, \Delta_\tau) \subset S(\Sigma^i, \Delta_i)$. If neither Δ_i is Δ_τ, then Δ_i is a common face between Σ_τ and Σ^i. From Figure 5.4 again we see that if $x \notin S(\Sigma^i, \Delta_i)$, x lifts to a non-regular point of $C(\Sigma_\tau)$ in $\partial\Delta_i$ which remains non-regular in $C(\Sigma_\tau \pm \Sigma^i)$, hence also in $C(\Sigma)$, a contradiction.

By a *shadow arc* we mean a maximal connected subset of $S(\Sigma^i, \Delta_i)$, for some i.

(2) A shadow arc of $S(\Sigma^i, \Delta_i)$ can enter a tongue of $P(\Sigma)$ from outside only at the cusp points of that tongue. In fact, from Figure 5.4 we see that $\pi(\partial\Delta_i)$ can meet an outer collar (see Figure 9.1) on the tongue of $P(\Sigma)$ only at points in the boundary of the tongue.

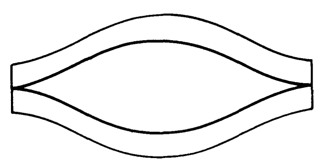


FIGURE 9.1

(3) A shadow arc $\alpha \subset S(\Sigma^i, \Delta_i)$ has both its endpoints on one component of $P^+(\Sigma)$, and:

(a) If the endpoints of α lie on $\pi(\partial D)$ for D the initial disc of $C(\Sigma)$, then α subdivides $\pi(D)$ into a disc and a tongue.

(b) If the endpoints of α lie on $\pi(\partial T)$ for T a tongue of $C(\Sigma)$, then α subdivides $\pi(T)$ into two tongues.

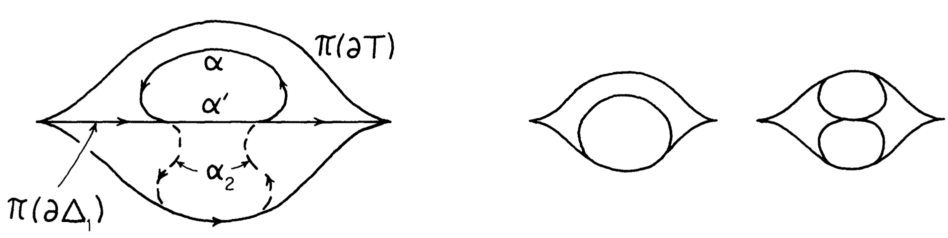
This is because the endpoints of α are also the endpoints of $\pi(A)$ for an arc A of free edge points of a type II tongue of $C(\Sigma^i, \Delta_i)$. The region between α and $\pi(A)$ is a tongue T', and $\pi(A) \subset P^+(\Sigma)$. If $\pi(A) \subset \pi(\partial D)$, then $\pi(D) - T'$ is a disc and (a) holds. If $\pi(A) \subset \pi(\partial T)$ for a tongue T of $C(\Sigma)$, then $T' \subset \pi(T)$ by (2), so $\pi(T) - T'$ is a tongue and (b) holds.

Consider now $\pi(\partial T)$ for a tongue T of $C(\Sigma)$. Let $\alpha'_1, \ldots, \alpha'_m$ be shadow arcs whose endpoints lie on or outside $\pi(\partial T)$, and let α_i be the part of α'_i on or inside $\pi(\partial T)$. (Thus $\alpha_i = \alpha'_i$ if the endpoints of α'_i are on $\pi(\partial T)$.) We claim that $\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_m$ is a tongue block. This will be shown by induction on m. Since $\pi(\partial T) \cup \cdots \cup \alpha_m$ is orientable, it could fail to be a tongue block only if it contained a smooth circle.

$m = 1$. If the ends of α_1' are outside $\pi(\partial T)$, then the ends of α_1 are at the cusps of $\pi(\partial T)$ and clearly $\pi(\partial T) \cup \alpha_1$ can contain no smooth circle. The other possibility, that the ends of $\alpha_1' = \alpha_1$ are on $\pi(\partial T)$, was handled in (3) above.

$m = 2$. If $\alpha_1 \cap \alpha_2 = \varnothing$, this case follows from the preceding one. If $\alpha_1 \cap \alpha_2 \neq \varnothing$, then α_1 and α_2 come from two different shadow sets, say $S(\Sigma^1, \Delta_1)$ and $S(\Sigma^2, \Delta_2)$. Applying Lemma 5.3 to a primitive Σ_τ having Δ_1 and Δ_2 as faces, we see that $\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2)$ is a configuration as in Figure 5.10a. As we saw in (2) above, $\pi(\partial\Delta_1)$ and $\pi(\partial\Delta_2)$ are disjoint from an outer collar on $\pi(\partial T)$. We distinguish two cases:

(a) If $\pi(\partial\Delta_1)$ or $\pi(\partial\Delta_2)$, say $\pi(\partial\Delta_1)$, goes outside $\pi(\partial T)$ at the cusps of $\pi(\partial T)$, then the claim we are trying to establish for $\pi(\partial T) \cup \alpha_1 \cup \alpha_2$ could fail only if there is a smooth circle in $\pi(T)$ consisting of arcs $\alpha \subset \alpha_2$ and $\alpha' \subset \pi(\partial\Delta_1)$ (see Figure 9.2a). Lemma 5.3 implies that if we orient $\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2)$ and use this orientation to assign a cyclic order to each triple of distinct points in $\pi(\partial\Delta_1)$, or each triple of distinct points in $\pi(\partial\Delta_2)$, then for triples in $\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2)$ these two cyclic orderings agree. As a result, $\pi(\partial\Delta_2)$ can meet $\pi(\partial\Delta_1)$ only in α'. Prolonging α to α_2, whose endpoints lie on $\pi(\partial T)$, and using the orientability of $\pi(\partial T) \cup \alpha_2$, we see that $\pi(\partial T) \cup \alpha_2$ contains a smooth circle, contrary to the case $m = 1$.



(a) (b)

FIGURE 9.2

(b) If $\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2) \subset \pi(T)$, we can without loss of generality pinch $\pi(\partial\Delta_1)$, $\pi(\partial\Delta_2)$, and $\pi(\partial T)$ together to eliminate any 2-cusped regions in the complement of their union. This leaves only the two configurations for $\pi(\partial T) \cup \pi(\partial\Delta_1) \cup \pi(\partial\Delta_2)$ shown in Figure 9.2b. In these cases the result for $m = 1$ clearly implies the result for $m = 2$.

$m > 2$. Suppose inductively that $\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_{m-1}$ is a tongue block. If there is a smooth circle in $\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_m$, then in some tongue of the tongue block $\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_{m-1}$ there is a configuration like that in Figure 9.3a: a smooth circle consisting of an arc $\alpha \subset \alpha_m$ and an arc $\alpha' \subset$

$\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_{m-1}$. We may assume the interior of the disc bounded by $\alpha \cup \alpha'$ is disjoint from $\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_{m-1}$.

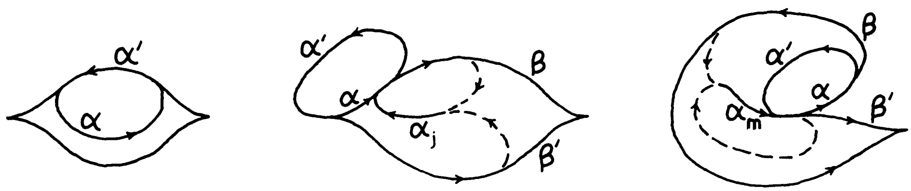


FIGURE 9.3

The endpoints of α cannot both lie on $\pi(\partial T)$, else $\pi(\partial T) \cup \alpha_m$ would contain a smooth circle. Orient $P^+(\gamma)$ so that the head of α is not on $\pi(\partial T)$. There are smooth arcs $\beta \subset \pi(\partial T) \cup \alpha_m$ and $\beta' \subset \pi(\partial T) \cup \alpha_i$, for some $i < m$, such that the tail of β is the head of α, the tail of β' is the tail of α, and the head of β is the head of β'. The two possible configurations for $\alpha \cup \alpha' \cup \beta \cup \beta'$ are shown in Figures 9.3b and 9.3c; the distinction is whether α' lies outside or inside $\alpha \cup \beta \cup \beta'$. In 9.3b there is some α_j, $j < m$, entering at the head of α from outside the disc bounded by $\alpha \cup \alpha'$. Following this α_j backward we must eventually meet either β or β'. If we meet β, then there is a smooth circle in $\alpha_j \cup \beta \subset \pi(\partial T) \cup \alpha_j \cup \alpha_m$, contradicting the case $m = 2$. If we meet β', then there is a smooth circle in $\alpha_j \cup \alpha' \cup \beta' \subset \pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_{m-1}$, contrary to the induction hypothesis. In 9.3c we follow α_m backwards from the tail of α until we meet either β or β'. Again a contradiction results in either case. Hence $\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_m$ is a tongue block.

Now let $\alpha_1, \ldots, \alpha_m$ be shadow arcs whose endpoints lie on $\pi(\partial D)$, where D is the initial disc of $C(\Sigma)$. Define a *disc block* to be a subset of $\mathbf{R}^2$ of the form $\partial D_0 \cup \partial T_1 \cup \cdots \cup \partial T_p$, where $\{D_0, T_1, \ldots, T_p\}$ is a disc-with-tongues structure on a disc $D_1 \subset \mathbf{R}^2$, such that $\partial D_0 \cap \partial D_1$ contains an arc (which can be chosen to vary smoothly with the parameter t, at least locally in t). The claim is that $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_m$ is a disc block.

If the arc α_i arises from Δ_j, so $\alpha_i \subset \pi(\partial \Delta_j)$, then Δ_j must be a sum face splitting off Σ^j, since α_i goes out to $\pi(\partial D)$. Let $T(\alpha_i) \subset \pi(D)$ be the tongue cut off $\pi(D)$ by α_i. ($T(\alpha_i)$ exists because $\pi(\partial D) \cup \alpha_i$ is orientable.) The components of $\pi(D) - \bigcup_i T(\alpha_i)$ are smooth discs, since they arise by intersecting the smooth discs $\pi(D) - T(\alpha_i)$ whose boundaries meet tangentially with consistent orientations. The intersection of one disc component of $\pi(D) - \bigcup_i T(\alpha_i)$ with $\pi(\partial D)$ must contain an arc A, since points of $\pi(\partial D) - \bigcup_i T(\alpha_i)$ are projections of free edge points of the initial disc of $C(\Sigma_0)$ which remain free edge points of the initial disc of $C(\Sigma)$, and Σ_0 was chosen to be a large factor of $\Sigma(\gamma)$.

Note also for future reference that $\pi(\partial D) - \bigcup_i T(\alpha_i) = P^+(\gamma) - P(\gamma)$, so that $P(\gamma)$ will be a tongue pattern if $P^+(\gamma)$ is a disc block with disjoint tongue blocks in its interior.

We check that $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_m$ is a disc block by induction on m. The case $m = 1$ is obvious. For $m = 2$, we suppose as before that α_1 and α_2 come from the common faces Δ_1 and Δ_2. Of the two types of configurations for $\pi(\partial\Delta_1) \cup \pi(\partial\Delta_2)$ shown in Figure 5.10a, the second is ruled out here since $\pi(\partial D) \cup \pi(\partial\Delta_1) \cup \pi(\partial\Delta_2)$ is orientable and both $\pi(\partial\Delta_1)$ and $\pi(\partial\Delta_2)$ lie in $\pi(D)$ and must meet $\pi(\partial D)$. For the first configuration the result is obvious. For $m > 2$ we assume $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_{m-1}$ is a disc block. To show that $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_m$ is a disc block we must show that every smooth circle in $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_m$ contains the arc $A \subset \pi(\partial D)$. Suppose not. Then there is a smooth circle disjoint from A, consisting of an arc $\alpha \subset \alpha_m$ and an arc $\alpha' \subset \pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_{m-1}$. The endpoints of α cannot both lie on $\pi(\partial D)$, since if they did, either $\alpha \cup \pi(\partial D)$ or $\alpha' \cup \pi(\partial D)$ would contain a smooth circle disjoint from A, contrary to the induction hypothesis (see Figure 9.4).

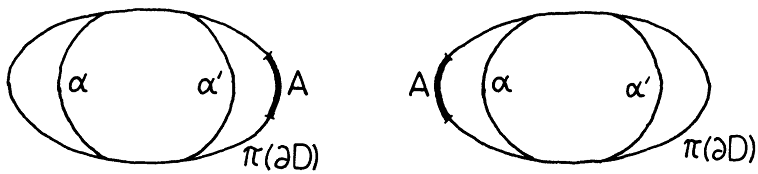


FIGURE 9.4

The rest of the argument that $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_m$ is a disc block proceeds just as in the earlier argument that $\pi(\partial T) \cup \alpha_1 \cup \cdots \cup \alpha_m$ is a tongue block, with $\pi(\partial T)$ replaced now by $\pi(\partial D) - A$.

The proposition now follows by an inductive argument. Upon the disc block $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_m$, where $\alpha_1, \ldots, \alpha_m$ are all the shadow arcs with endpoints on $\pi(\partial D)$, one first superimposes the largest tongue $\pi(\partial T)$ which meets $\pi(\partial D) \cup \alpha_1 \cup \cdots \cup \alpha_m$, together with all shadow arcs with endpoints on $\pi(\partial T)$. The result is again a disc block, since it decomposes $\pi(D)$ into regions all of which are 2-cusped except for one smooth disc. On this disc block one superimposes the next largest $\pi(\partial T)$ meeting this disc block, with its shadow arcs, etc. Thus the component of $P^+(\gamma)$ containing $\pi(\partial D)$ is a disc block. Then one repeats the argument for the largest remaining tongue $\pi(\partial T)$, and so on. Thus $P^+(\gamma)$ is a disc block with disjoint tongue blocks in its interior. As mentioned earlier this implies that $P(\gamma)$ is a tongue pattern—a disjoint union of tongue blocks.

For the preceding argument we have had in mind a fixed parameter value t_0. But letting t vary presents no problems; we are just adding shadow arcs to

$P^+(\Sigma)$, which have their endpoints on $P^+(\Sigma)$. Such arcs vary continuously with t, except when they are absorbed into $P^+(\Sigma)$. □

We shall use the notation $\Sigma_t(\gamma) < \Sigma_t(\gamma')$ to mean that $\Sigma_t(\gamma) \subset \Sigma_t(\gamma')$ and the closure of the stratum of $\mathcal{S}_0'$ over which $\Sigma_t(\gamma)$ is defined contains the stratum over which $\Sigma_t(\gamma')$ is defined.

LEMMA 9.2. *If* $\Sigma_t(\gamma) < \Sigma_t(\gamma')$, *then* $P_t(\gamma) \subset P_t(\gamma')$.

Proof. Since $\Sigma(\gamma) \subset \Sigma(\gamma')$, each term $P(\Sigma_\tau)$ appearing in the definition of $P(\gamma)$ appears also in $P(\gamma')$. It remains to check that this holds also for terms $P(\Sigma_\tau, \Delta_\tau)$. It suffices to consider the case that the strata of $\Sigma(\gamma)$ and $\Sigma(\gamma')$ have dimensions differing by one. There are three subcases:

(a) $\Sigma(\gamma)$ is obtained from $\Sigma(\gamma')$ by deleting a common face Δ_i.

(b) $\Sigma(\gamma)$ is obtained from $\Sigma(\gamma')$ by splitting into two components along a common face Δ_i.

(c) $\Sigma(\gamma) = \Sigma(\gamma')$, $\Sigma(\gamma')$ has a base edge, and $\Sigma(\gamma)$ has a base vertex at one end of this edge.

In each of these subcases the desired conclusion follows immediately from the definitions, as the reader can easily check. □

10. Shrinking tongue patterns

For a tongue $T \subset \mathbf{R}^2$ in a tongue block, there is a dual tongue $T^* \subset \mathbf{R}^2$ with $\partial T^* = \partial T$, obtained by interchanging the free and attaching edges of T. Shrinkings of T and T^* correspond bijectively, in the obvious way, and will in fact be identified, for convenience. A shrinking of T has an associated tangent line field, the lines tangent to the free edges of the various stages of the shrinking. The dual shrinking of T^* has the same tangent line field. Even for a C^∞ shrinking, the tangent line field will be only C^0, not C^1, since it has non-unique trajectories through the cusps of T, and possibly also through some other points.

Two shrinkings of a tongue with the same tangent line field may not be equivalent in the sense of Section 6. For example, if we start with one shrinking of the tongue shown in Figure 10.1, we can obtain inequivalent shrinkings by independently varying the rate of shrinking in the two halves of the tongue. (Less trivially, this same sort of thing could be happening in the interior of a tongue.)

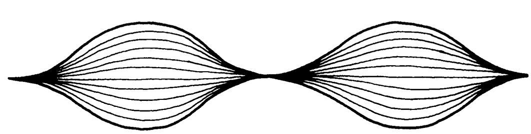


FIGURE 10.1

However, if T_s and T_s' are two shrinkings of a tongue T with the same tangent line field, there is a canonical path of shrinkings connecting T_s and T_s', obtained by first connecting T_s to the shrinking $T_s \cap T_s'$ via the path $T_s \cap T_{rs}'$, $0 \leq r \leq 1$, then connecting $T_s \cap T_s'$ to T_s' in the same way. The tangent line field is constant along this path from T_s to T_s'.

PROPOSITION 10.1. *There is a triangulation $\mathfrak{T}$ of S^k in which closed strata of $\mathfrak{S}_0'$ are subcomplexes, such that for each simplex σ of $\mathfrak{T}$ and each family $\Sigma_t(\gamma)$ defined for $t \in \sigma$, there exist:*

(i) *families of tongue blocks $Q_r(\gamma)$, parametrized by $t \in \sigma$,*

(ii) *inclusion maps $P_q(\gamma) \hookrightarrow Q_r(\gamma)$, $r = r(q)$, for the tongue blocks $P_q(\gamma)$ whose disjoint union is the tongue pattern $P(\gamma)$,*

(iii) *families of shrinkings of the tongues of $Q_r(\gamma)$ such that, if $r_1 \neq r_2$, the associated tangent line fields for the tongues of $Q_{r_1}(\gamma)$ meet those for the tongues of $Q_{r_2}(\gamma)$ transversely for all parameter values $t \in \sigma - \partial\sigma$.*

Further, if σ' is a face of σ, $\Sigma_t(\gamma')$ is defined for $t \in \sigma'$, and $\Sigma_t(\gamma) < \Sigma_t(\gamma')$ for $t \in \sigma'$, then we have a diagram

$$\begin{array}{ccc} P_q(\gamma) & \hookrightarrow & Q_r(\gamma) \\ \downarrow & & \downarrow \\ P_{q'}(\gamma') & \hookrightarrow & Q_{r'}(\gamma') \end{array}$$

and the tangent line fields associated to the chosen shrinkings of the tongues of $Q_{r'}(\gamma')$ restrict to those for the tongues of $Q_r(\gamma)$.

Proof. To begin we choose a smooth triangulation $\mathfrak{T}_1$ of S^k with the following properties:

(1) The closed strata of $\mathfrak{S}_0'$ are subcomplexes.

(2) For each simplex σ of $\mathfrak{T}_1$ and each $\Sigma_t(\gamma)$ defined for $t \in \sigma$, the tongue pattern $P(\gamma) = P_t(\gamma)$ is decomposed into well-defined disjoint blocks $P_q(\gamma)$ (depending continuously on t) such that, if $\sigma \supset \sigma'$ and $\Sigma_t(\gamma')$ is defined for $t \in \sigma'$, with $\Sigma_t(\gamma) < \Sigma_t(\gamma')$, then each $P_q(\gamma)$ is contained in a (unique) $P_{q'}(\gamma')$. (Recall that blocks need not be connected, and this is why (2) is not automatic.) We can achieve (1) and (2) as follows. For each $\Sigma_t(\gamma)$ defined over a closed stratum X of $\mathfrak{S}_0'$ we can certainly choose well-defined block decompositions $\{P_q(\gamma)\}_j$ of $P(\gamma)$ locally in X, say in the intersections of X with finitely many PL k-balls $B_j(X) \subset S^k$ whose interiors cover X. Let $\mathfrak{T}_1$ be such that (1) holds and all these $B_j(X)$'s are subcomplexes. Then over a simplex σ of $\mathfrak{T}_1$ in X, we take as a first approximation to $\{P_q(\gamma)\}$ the block decomposition of $P(\gamma)$ obtained by intersecting all blocks of the decompositions $\{P_q(\gamma)\}_j$ such that $\sigma - \partial\sigma \subset \operatorname{int}(B_j(X))$. Further, if some face σ' of σ lies in a closed stratum X' and

$\Sigma(\gamma) < \Sigma(\gamma')$ for $\Sigma(\gamma')$ defined over X', then the blocks of the decomposition of $P(\gamma')$ can be used to further decompose (by intersection) the blocks of $\{P_q(\gamma)\}$, certainly for $t \in \sigma'$, but also for $t \in \sigma$ if the balls are chosen small enough. With this refined definition of $\{P_q(\gamma)\}$, it is clear that (2) holds.

For a tongue block $P_q(\gamma)$, let $\bar{P}_q(\gamma)$ be the union of $P_q(\gamma)$ with the bounded components of $\mathbf{R}^2 - P_q(\gamma)$. The blocks $P_q(\gamma)$ are partially ordered by setting $P_q(\gamma) < P_r(\gamma)$ if $\bar{P}_q(\gamma) \subset \bar{P}_r(\gamma)$. If $P_q(\gamma)$ is defined over the simplex σ of $\mathfrak{T}_1$, let $V_q(\gamma)$ be a disc neighborhood of $P_q(\gamma)$, varying smoothly with $t \in \sigma$, and let $\nu_q(\gamma)$ be a nonvanishing vector field on $V_q(\gamma)$ transverse to $P_q(\gamma)$, also varying smoothly with $t \in \sigma$, and satisfying the compatibility condition that if $P_q(\gamma) \subset P_{q'}(\gamma')$ for $t \in \sigma' \subset \partial\sigma$ then $\nu_q(\gamma) = \nu_{q'}(\gamma')$ on $V_q(\gamma)$. (We may assume $V_q(\gamma) \subset V_{q'}(\gamma')$ for $t \in \sigma'$.) Such vector fields $\nu_q(\gamma)$ can be constructed inductively over skeletons of $\mathfrak{T}_1$. For convenience, $V_q(\gamma)$ may be chosen so that $\partial V_q(\gamma)$ consists of two transversals to $\nu_q(\gamma)$, meeting in two points $x_q^+(\gamma)$ and $x_q^-(\gamma)$.

By flowing along trajectories of $\nu_q(\gamma)$, we can construct a shrinking of the tongue $\bar{P}_q(\gamma)$ which restricts to shrinkings of the tongues of $P_q(\gamma)$. By doing this construction inductively over skeletons of $\mathfrak{T}_1$, we may arrange that the shrinking of $\bar{P}_{q'}(\gamma')$ restricts to the shrinking of $\bar{P}_q(\gamma)$ when $P_q(\gamma) \subset P_{q'}(\gamma')$ and $t \in \sigma' \subset \partial\sigma$. Let $\tau_q(\gamma)$ be the unit tangent vector field of the chosen shrinking of $\bar{P}_q(\gamma)$, determined by orienting $P_q(\gamma)$ as in Section 9. We may extend $\tau_q(\gamma)$ to be defined on $V_q(\gamma)$, not just on $\bar{P}_q(\gamma)$, with $\tau_q(\gamma) = \tau_{q'}(\gamma')$ on $V_q(\gamma)$ if $P_q(\gamma) \subset P_{q'}(\gamma')$ and $t \in \sigma' \subset \partial\sigma$. By choosing carefully the shrinkings of tongues of $P_q(\gamma)$, i.e., the rates of flow along trajectories of $\nu_q(\gamma)$, we may assume:

(3) For $t \in \sigma - \partial\sigma$, $\tau_q(\gamma)$ is C^∞ except at cusps of tongues of $P_q(\gamma)$. We may assume also that for $t \in \sigma - \partial\sigma$, all integral curves of $\tau_q(\gamma)$ in $V_q(\gamma)$ which meet $P_q(\gamma)$ enter $V_q(\gamma)$ at $x_q^-(\gamma)$ and exit at $x_q^+(\gamma)$.

Let $\theta_q(\gamma)$ be the vector $\tau_q(\gamma)$ at $x_q^-(\gamma)$. We may suppose this depends piecewise linearly on the parameter t. We now deform the $\tau_q(\gamma)$'s so that:

(4) $\tau_q(\gamma) - \tau_r(\gamma) = \theta_q(\gamma) - \theta_r(\gamma)$ on $V_r(\gamma)$ if $P_r(\gamma) < P_q(\gamma)$. Here the subtraction of unit vectors in $\mathbf{R}^2$ is carried out by regarding them as angles formed with the x-axis. To achieve (4) we proceed inductively over skeletons of $\mathfrak{T}_1$. Over a given simplex of $\mathfrak{T}_1$ we modify $\tau_q(\gamma)$, assuming $\tau_r(\gamma)$ has already been modified for all $P_r(\gamma) < P_q(\gamma)$ with $r \neq q$, by deforming $\tau_q(\gamma)$ near the maximal $V_r(\gamma)$'s in $\bar{P}_q(\gamma)$ so that (4) holds not just at the points $x_r^-(\gamma)$, but on the discs $V_r(\gamma)$. After doing this, (3) has to be weakened to the statement that $\tau_q(\gamma)$ is C^∞ except at cusps of tongues of $P_q(\gamma)$ and of blocks $P_r(\gamma) < P_q(\gamma)$. However, the following is still true for $t \in \sigma - \partial\sigma$:

(5) $\tau_q(\gamma)$ has unique local trajectories through all points of $\bar{P}_q(\gamma)$ except:

(a) cusps of tongues of $P_q(\gamma)$,

(b) cusps of tongues of $P_r(\gamma)$ with $P_r(\gamma) < P_q(\gamma)$ and $2\theta_q(\gamma) = 2\theta_r(\gamma)$.

For $P_r(\gamma) < P_q(\gamma)$, let $X_{qr} = \{t | 2\theta_q(\gamma) = 2\theta_r(\gamma)\}$ and let $\mathcal{T}_2$ be a subdivision of $\mathcal{T}_1$ such that all these sets X_{qr} are subcomplexes. Fix a simplex σ of $\mathcal{T}_2$, and consider $t \in \sigma - \partial\sigma$. For $P_q(\gamma)$ defined over σ, let $\mathfrak{M}_q(\gamma)$ be the set of maximal $P_r(\gamma)$'s, $r \neq q$, such that $P_r(\gamma) < P_q(\gamma)$ and $2\theta_r(\gamma) = 2\theta_q(\gamma)$. Let T be a (closed) tongue of $P_q(\gamma)$. Define $\hat{T}$ to be T minus the interiors of those $\bar{P}_r(\gamma)$'s with $P_r(\gamma) \in \mathfrak{M}_q(\gamma)$. Let $\tilde{T}$ be the quotient of $\hat{T}$ obtained by identifying, for each tongue $\bar{P}_r(\gamma) \subset T$ with $P_r(\gamma) \in \mathfrak{M}_q(\gamma)$, pairs of free and attaching edge points of $\bar{P}_r(\gamma)$ which lie on the same trajectory of $\nu_r(\gamma)$. Thus $\tilde{T}$ is a tongue, and $\tau_q(\gamma)$ defines a vector field $\tilde{\tau}_q(\gamma)$ on $\tilde{T}$ tangent to $\partial\tilde{T}$ and with unique local trajectories through non-cusp points of $\tilde{T}$. It is not hard to see that $\tilde{T}$ has a shrinking with $\tilde{\tau}_q(\gamma)$ as tangent field. That is, the free-edge curves of this shrinking are made up of free-edge curves of the earlier-chosen shrinking of T near ∂T, and integral curves of $\tilde{\tau}_q(\gamma)$ away from ∂T. (The modification of $\tau_q(\gamma)$ to obtain (4) occurred only in the interior of T.) The free edges of the various stages of this shrinking of $\tilde{T}$ we shall call *trajectories of* $\tilde{\tau}_q(\gamma)$ in $\tilde{T}$, even though they may not be connected if $\tilde{T}$ is not connected. The shrinking of $\tilde{T}$ we may assume depends continuously on $t \in \sigma - \partial\sigma$, and may be extended by continuity to a shrinking for $t \in \partial\sigma$ as well: The trajectories of $\tilde{\tau}_q(\gamma)$ extend continuously over $\partial\sigma$ since $\tilde{\tau}_q(\gamma)$ extends continuously, and we can obtain shrinkings of $\tilde{T}$ from these trajectories of $\tilde{\tau}_q(\gamma)$ by specifying that the area of $\tilde{T}$ decrease at a given rate, depending continuously on $t \in \sigma$. So at $t \in \partial\sigma$, the limiting monotone family of trajectories of $\tilde{\tau}_q(\gamma)$ has no gaps, hence forms a shrinking of $\tilde{T}$.

Now we restrict again to $t \in \sigma - \partial\sigma$. The trajectories of $\tilde{\tau}_q(\gamma)$ in $\tilde{T}$ pull back naturally to trajectories of $\tau_q(\gamma)$ in $\hat{T}$; the finitely many trajectories of $\tilde{\tau}_q(\gamma)$ containing the images of the $\partial\bar{P}_r(\gamma)$'s for $P_r(\gamma) \in \mathfrak{M}_q(\gamma)$ pull back to pairs of trajectories of $\tau_q(\gamma)$ in $\hat{T}$, which bifurcate at cusps of the $\bar{P}_r(\gamma)$'s. By continuity, we can again extend the trajectories of $\tau_q(\gamma)$ to be defined over $\partial\sigma$. The union of the bifurcating trajectories of $\tau_q(\gamma)$ determines a subdivision of T, at least for fixed $t \in \sigma$. However, as t varies over σ, these subtongues of T may not form a family of tongue structures on T because, locally in σ, it may not be possible to attach these subtongues in a single order. An example of this is indicated in Figure 10.2, with $\sigma = [0, 1]$.

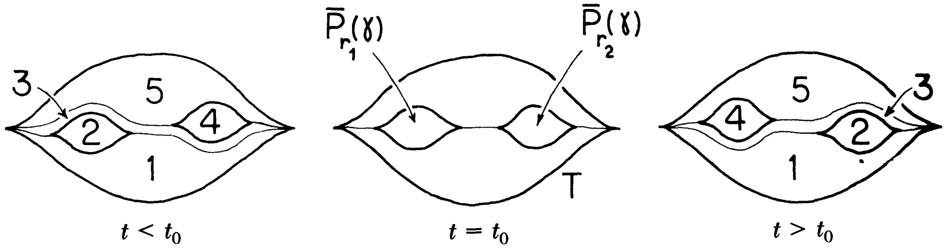


FIGURE 10.2

To resolve this problem, consider the sets $Y_{r_1 r_2}$ of $t \in \sigma$ such that, for $P_{r_1}(\gamma)$ and $P_{r_2}(\gamma)$ in $\mathfrak{M}_q(\gamma)$ with $r_1 \neq r_2$, both $P_{r_1}(\gamma)$ and $P_{r_2}(\gamma)$ meet the same trajectory of $\tau_q(\gamma)$ in $\hat{T}$. We may assume all such sets $Y_{r_1 r_2}$ are subcomplexes of σ (for all $\sigma \in \mathfrak{T}_2$) and subdivide $\mathfrak{T}_2$ to a triangulation $\mathfrak{T}$ in which the $Y_{r_1 r_2}$'s are subcomplexes. Then for σ a simplex of $\mathfrak{T}$, the bifurcating trajectories of $\tau_q(\gamma)$ in $\hat{T}$ do determine a family of subdivisions of T over σ. Further, the shrinkings of the tongues $\tilde{T}$ yield naturally shrinkings of the tongues in this family of subdivisions of T, except for the $\bar{P}_r(\gamma)$'s with $P_r(\gamma) \in \mathfrak{M}_q(\gamma)$.

Define inductively families of tongue blocks $P_q'(\gamma)$ over the simplices of $\mathfrak{T}$, by letting $P_q'(\gamma)$ be the union of:

(a) the boundaries of all the subtongues (just constructed) of the tongues T of $P_q(\gamma)$,

(b) all $P_r'(\gamma)$'s with $P_r(\gamma) \in \mathfrak{M}_q(\gamma)$.

Each $P_q(\gamma)$ is contained in a maximal $P_r'(\gamma)$, which we define to be $Q_r(\gamma)$. We have constructed shrinkings of the tongues of all the $Q_r(\gamma)$'s. These satisfy the conditions of the proposition. □

The shrinkings of the tongues of the blocks $Q_r(\gamma)$ which we have constructed may be only C^1, not C^∞, since their tangent fields $\tau_r(\gamma)$ were only C^0. Later we will observe how to return easily to the C^∞ category.

11. The main construction: Local form

In this section we consider a family $\Sigma(\gamma)$ with common faces $\Delta_1, \ldots, \Delta_n$, as in Section 8, defined over a simplex of $\mathfrak{T}$. The main goal is a procedure for constructing n-parameter deformations $\Sigma^i(s_1, \ldots, s_n)$, $0 \leq s_j \leq 1$, of the cofactors $\Sigma^i = \Sigma^i(0, \ldots, 0)$ of $\Sigma(\gamma)$. Since Σ^i itself depends on a parameter t, we will really be constructing families $\Sigma_t^i(s_1, \ldots, s_n)$. But for convenience we continue to drop the parameter t from the notation. Also for notational convenience, let us label the Σ^i's so that $\Sigma^i > \Sigma^j$ implies $i \leq j$. (Order of cofactors was defined in § 8.) It will turn out that $\Sigma^i(s_1, \ldots, s_n)$ will depend only on variables s_j with $\Sigma^j < \Sigma^i$; in particular, $\Sigma^i(s_1, \ldots, s_n)$ will be independent of $s_1, \ldots, s_{i-1}$.

The construction of the families $\Sigma^i(s_1, \ldots, s_n)$ is made by an inductive process, which we now describe formally. First, $\Sigma^n(0, \ldots, 0, s_n)$ is constructed as a shrinking of Σ^n to its face Δ_n. This means:

(a) $\Sigma^n(0, \ldots, 0) = \Sigma^n$.

(b) $\Sigma^n(0, \ldots, 0, 1) = \Delta_n$.

(c) $\Delta_n \subset \Sigma^n(0, \ldots, 0, s_n)$ for $0 \leq s_n \leq 1$.

(d) $\Sigma^n(0, \ldots, 0, s_n) - \operatorname{int}(\Delta_n)$ for $0 < s_n < 1$ is a smooth disc bounded by $\partial\Delta_n$, which moves across $\bar{\Sigma}^n$ by monotone isotopy (rel $\partial\Delta_n$) from $\Sigma^n - \operatorname{int}(\Delta_n)$ to Δ_n as s_n goes from 0 to 1.

If $\Sigma^n < \Sigma^i$, $i \neq n$, then $\Sigma^n - \mathrm{int}(\Delta_n) \subset \Sigma^i$, and $\Sigma^i(0,\ldots,0,s_n)$ is specified by the conditions $\Sigma^n(0,\ldots,0,s_n) - \mathrm{int}(\Delta_n) \subset \Sigma^i(0,\ldots,0,s_n)$ and $\Sigma^i - \Sigma^n \subset \Sigma^i(0,\ldots,0,s_n)$. If $\Sigma^n \not< \Sigma^i$, we set $\Sigma^i(0,\ldots,0,s_n) = \Sigma^i$.

The induction step is to extend families $\Sigma^i(0,\ldots,0,s_{j+1},\ldots,s_n)$ to families $\Sigma^i(0,\ldots,0,s_j,\ldots,s_n)$. To simplify notation we abbreviate $(0,\ldots,0,s_j,s_{j+1},\ldots,s_n)$ to (s_j). Assume inductively that $\Sigma^j(0)$ is an embedded sphere containing Δ_j, which is smooth except possibly at corners of Σ^j and at circles $\partial\Delta_i \subset \Sigma^j$, and assume also:

(1) If $\Sigma^j < \Sigma^i$, $j \neq i$, then $\overline{\Sigma}^j(0) \cap \Sigma^i(0) = \Sigma^j(0) - \mathrm{int}(\Delta_j)$.

Then $\Sigma^j(s_j)$, $0 \leq s_j \leq 1$, is constructed as a shrinking of $\Sigma^j(0)$ to Δ_j. As in the case $j = n$, this shrinking determines, via (1), an isotopy $\Sigma^i(s_j)$ of $\Sigma^i(0)$ for each $\Sigma^i > \Sigma^j$, $i \neq j$. If $\Sigma^i \not> \Sigma^j$, we let $\Sigma^i(s_j)$ be independent of s_j. It should be clear that the induction hypotheses continue to hold, with j replaced by $j-1$.

To carry out the induction step we need $C(\Sigma^j(0))$ to be a disc with tongues, with $C(\Delta_j)$ as initial disc, and we need a shrinking of $C(\Sigma^j(0))$ to $C(\Delta_j)$, so that the Addendum to Lemma 6.1 will apply to give a shrinking of $\Sigma^j(0)$ to Δ_j. As the example in Section 7 shows, even at the first step of the induction process, if we choose the shrinking of $C(\Sigma^n)$ to $C(\Delta_n)$ carelessly, the disc with tongues structure on $C(\Sigma^{n-1})$ may be destroyed, so we would be unable to continue the construction.

By Proposition 9.1 we have the tongue pattern $P(\gamma)$, which is the disjoint union of tongue blocks $P_q(\gamma)$. After Proposition 10.1, there are inclusions $P_q(\gamma) \hookrightarrow Q_r(\gamma)$, $r = r(q)$. Each type I or II tongue T of a contour $C(\Sigma^j, \Delta_j)$ (or $C(\Sigma^0)$ if $j = 0$) has $\pi(\partial T)$ contained in a unique tongue block $P_q(\gamma)$, hence also in a well-defined $Q_r(\gamma)$. The tongues of $Q_r(\gamma)$ induce naturally a subdivision of T into tongues which project homeomorphically to tongues of $Q_r(\gamma)$. We say these subtongues of T lie *in the q^{th} block*. Doing this for all tongues T, we obtain a disc-with-tongues structure which we call $C(\Sigma^j(0,\ldots,0))$, subdividing $C(\Sigma^j, \Delta_j)$. The tongues of $C(\Sigma^j(0,\ldots,0))$ are partitioned into *blocks*, one block for each $P_q(\gamma)$. These blocks are partially ordered according to the nesting relations among the corresponding $P_q(\gamma)$'s. Since different $P_q(\gamma)$'s are disjoint, we see that the tongues of $C(\Sigma^j(0,\ldots,0))$ can be attached in an order in which tongues in smaller blocks attach after tongues in larger blocks.

At the same time that we construct inductively the families $\Sigma^i(s_1,\ldots,s_n)$, we shall also construct inductively families of tongue blocks $Q_r^i(s_1,\ldots,s_n) \supset Q_r^i(0,\ldots,0) = Q_r(\gamma)$ for $0 \leq i \leq n$, whose tongues are *segments* of the tongues of $Q_r(\gamma)$, that is, differences between stages in the shrinkings of the tongues of $Q_r(\gamma)$ chosen in Proposition 10.1. The inductive hypotheses are (again abbreviating $(0,\ldots,0,s_j,s_{j+1},\ldots,s_n)$ to (s_j)):

(2) For each $i \geq 0$, $C(\Sigma^i(0))$ has a disc-with-tongues structure, with $C(\Delta_i)$ as initial disc if $i > 0$, the tongues being partitioned into blocks, one block for

each $P_q(\gamma)$, such that:

(a) For each tongue T of this structure in the q^{th} block, $\pi(T)$ is a tongue in $Q_r^i(0)$, $r = r(q)$.

(b) The tongues of $C(\Sigma^i(0))$ can be attached in an order in which tongues in smaller blocks attach after tongues in larger blocks.

(3) $Q_r^j(0) \subset Q_r^i(0) \subset Q_r^l(0)$ if $\Sigma^j < \Sigma^i < \Sigma^l$.

LEMMA 11.1. *A disc-with-tongues structure satisfying* (2a) *is unique, provided its initial disc is prescribed. In particular, its partition into blocks is unique.*

Proof. By induction on the number of tongues, it suffices to show that a last-attached tongue T in one structure satisfying (2a) is also a tongue in any other structure satisfying (2a). By the transversality condition of Proposition 10.1, ∂T projects into a unique $Q_r^i(0)$. A small arc of the free edge of T must also be an arc of free edge points in a tongue T' in any other tongue structure. By (2a) and the transversality condition, $\partial T' \subset Q_r^i(0)$, hence $\pi(\partial T') = \pi(\partial T)$. Since T was a last-attached tongue, this implies $T' = T$. □

With a disc-with-tongues structure on $C(\Sigma^j(0))$ satisfying (2), the shrinkings of the tongues of blocks $Q_r(\gamma)$ restrict to shrinkings of the tongues of $Q_r^j(0)$, which lift to shrinkings of the tongues of $C(\Sigma^j(0))$. We shrink $C(\Sigma^j(0))$ to $C(\Delta_j)$ via Lemma 6.3, then lift this to the shrinking $\Sigma^j(s_j)$ of $\Sigma^j(0)$ to Δ_j, using Addendum 6.2. Define $Q_r^j(s_j)$ by adding to $Q_r^j(0)$ the projections of the free edges of the tongues of $C(\Sigma^j(s_j))$ in the q^{th} block (i.e., tongues of $C(\Sigma^j(s_j))$ which are subtongues of tongues of $C(\Sigma^j(0))$ in the q^{th} block), for all q such that $r = r(q)$. If $\Sigma^i > \Sigma^j$ we set $Q_r^i(s_j) = Q_r^i(0) \cup Q_r^j(s_j)$, and if $\Sigma^i \not> \Sigma^j$ we set $Q_r^i(s_j) = Q_r^i(0)$. Thus (3) continues to hold with $j - 1$ in place of j.

It remains to check that the induced deformation $\Sigma^i(s_j)$ of $\Sigma^i(0)$ is such that (2) continues to hold with $j - 1$ in place of j. If $\Sigma^i \not> \Sigma^j$, then $\Sigma^i(s_j)$ and each $Q_r^i(s_j)$ are independent of s_j; so this is automatic. If $\Sigma^i > \Sigma^j$ there are two cases according to whether Δ_j splits Σ^i as a sum or a difference. In the sum case, the tongues of $C(\Sigma^j(0))$ are also tongues of $C(\Sigma^i(0))$. The shrinking $C(\Sigma^j(s_j))$ induces a shrinking of $C(\Sigma^i(0))$, and the result is clear.

In the difference case it suffices to consider, as in the proof of Lemma 6.1, the case that a single tongue T of $C(\Sigma^j(0))$, in the q^{th} block, say, is shrinking during the interval $[0, s_j]$, and that no other tongue of $C(\Sigma^j(0))$ attaches to the subtongue $T' \subset T$ which shrinks away during $[0, s_j]$. In $C(\Sigma^i(0))$ there are then two continuous sheets T^+ and T^- adjacent to T', above and below respectively, with $\pi(T^+) = \pi(T^-) = \text{cl}(\pi(T'))$. T^+ and T^- meet only at points with the same π-image as the free edge of T'.

LEMMA 11.2. *One of the following two alternatives holds*:

(a) $\mathrm{Int}(T^+)$ *and* $\mathrm{int}(T^-)$ *are the interiors of tongues* T_1 *and* T_2 *of* $C(\Sigma^i(0))$ *in a block labelled* q' *with* $r(q') = r(q)$.

(b) *One of* $\mathrm{int}(T^+), \mathrm{int}(T^-)$ *is the interior of a tongue* T_1 *of* $C(\Sigma^i(0))$ *in a block* q' *with* $r(q') = r(q)$, *and the other is disjoint from all tongues of* $C(\Sigma^i(0))$ *in the* q'^{th} *or smaller blocks.*

In both cases, T_1 *attaches along* $T^+ \cap T^-$.

Proof. It suffices by continuity to consider a t-value in the interior of the given simplex of $\mathfrak{T}$, so that the transversality condition of Proposition 10.1 holds. Consider the process of building $C(\Sigma^i(0))$ by attaching tongues as in condition (2b). By transversality, the configuration of the two sheets T^+ and T^- meeting along $T^+ \cap T^-$ can arise only when some tongue T_1 in a q'^{th} block with $r(q') = r(q)$ is attached, with $T^+ \cap T^-$ as its attaching edge. No tongues attached after T_1 can contribute to T^+ or T^-. This is clear for tongues projecting homeomorphically to tongues of $Q_r^i(0)$. And for tongues projecting to other $Q_{r'}^i(0)$'s, the transversality condition implies that the last-attached such tongue contributing to T would have to have a cusp in $\mathrm{int}(T)$, violating the continuity of the sheet T.

Thus T_1 coincides with $T^+ - T^-$ or $T^- - T^+$, say $T^+ - T^-$, and T^- is disjoint from all tongues in blocks smaller than the q'^{th} block. If $\mathrm{int}(T^-)$ meets a tongue T_2 in the q'^{th} block, then $\mathrm{int}(T_2)$ must coincide with $\mathrm{int}(T^-)$ since no tongue in the q'^{th} or any larger block can attach to $\mathrm{int}(T_2)$, by (2). □

As $\Sigma^j(0)$ shrinks to $\Sigma^j(s_j)$, $\Sigma^i(0)$ expands, changing $C(\Sigma^i(0))$ by pinching together successively larger segments of T^+ and T^- containing $T^+ \cap T^-$. In case (b) of Lemma 11.2, we are pinching T_1 to the union of the initial disc of $C(\Sigma^i(0))$ with tongues of larger blocks, and clearly the given disc-with-tongues structure on $C(\Sigma^i(0))$ passes down to a disc-with-tongues structure on the quotient $C(\Sigma^i(s_j))$, all tongues projecting homeomorphically except T_1, and attaching in the same order (satisfying (2b)) as in $C(\Sigma^i(0))$. This quotient tongue structure on $C(\Sigma^i(s_j))$ then subdivides to a tongue structure satisfying (2a) also.

In case (a) of Lemma 11.2, if $T^+ \cap T^-$ is the free edge of T_2, then T_1 must attach after T_2. If $T^+ \cap T^-$ is the attaching edge of T_2, then there is no difference between T_1 and T_2 for our purposes, so that we may as well assume that in an order of attaching as in (2b), T_1 attaches after T_2. Now the rest of the argument in case (a) of Lemma 11.2 is just the same as in case (b).

This completes the inductive step in the construction of the families $\Sigma^i(s_1, \ldots, s_n)$. We note that these families $\Sigma^i(s_1, \ldots, s_n)$ are independent of the particular labelling of the cofactors Σ^i we chose, satisfying the condition that $\Sigma^i > \Sigma^j$ implies $i \leq j$, by the uniqueness property in Lemma 11.1.

A mild extension of the process of constructing the families $\Sigma^i(s_1,\ldots,s_n)$ will be needed, in which each s_j ranges over $[0,2]$ rather than $[0,1]$. At the inductive step of constructing the families $\Sigma^i(s_j)$, we have $\Sigma^j(1) = \Delta_j$. As s_j goes from 1 to 2 we replace Δ_j by two parallel copies of itself, above and below at distance $(s_j - 1)\cdot\delta(\partial\Delta_j)$, for $\delta(-)$ as defined in Section 2. If Δ_j is, say, a lower face of Σ^j, then the lower of the two diverging copies of Δ_j gives an isotopy $\Sigma^i(s_j)$, $1 \le s_j \le 2$, of each $\Sigma^i(1)$ with $\Sigma^i > \Sigma^j$, and the upper of the two copies of Δ_j gives an isotopy $\Sigma^j(s_j)$ of the disc $\Sigma^j(1) = \Delta_j$. These isotopies are vertical motion, so contours are unaffected, and by the arguments already given we can continue the construction with j replaced by $j-1$.

For $0 \le s_j \le 1$, the shrinking $\Sigma^j(s_j)$ of $\Sigma^j(0)$ gives an isotopy $\Delta_j(s_j)$ of the disc $\Delta_j(0) = \Sigma^j(0) - \operatorname{int}(\Delta_j)$ across $\overline{\Sigma}^j(0)$ to $\Delta_j = \Delta_j(1)$, rel $\partial\Delta_j$. By a small perturbation of this isotopy, using collars, we may assume $\Delta_j(s_j) \cap \Sigma^j(0) = \partial\Delta_j$ for $0 < s_j < 1$. The disc $\Delta_j(s_j)$ splits $\Sigma^j(0)$ into the sum of $\Sigma^j(s_j)$ and a complementary factor which we call $\Sigma_j(s_j)$. When $s_j = 0$, $\Sigma_j(0)$ is just the disc $\Delta_j(0)$; but for $s_j > 0$, $\Sigma_j(s_j)$ is an embedded sphere. We have $\Sigma_j(1) = \Sigma^j(0)$, and we can easily extend $\Sigma_j(s_j)$ to be defined also for $1 \le s_j \le 2$ by just replacing $\Delta_j \subset \Sigma_j(1)$ with a parallel copy of itself, above or below according to whether Δ_j is a lower or upper face of Σ_j, at distance $(s_j - 1)\cdot\delta(\partial\Delta_j)$. We define $\Sigma_j(s_1,\ldots, s_n)$ to be $\Sigma_j(s_j)$, independent of $s_1,\ldots,s_{j-1}$.

Finally, for use in Section 13, define $\Sigma_i(t_1,\ldots,t_n)$ for $t_j \in [-1,1]$, $j = 1,\ldots,n$, to be the family $\Sigma_i(s_1,\ldots,s_n)$, where $t_j = 1 - s_j$. (This notation is of course ambiguous if specific values are assigned to the variables. But only the *new* families $\Sigma_i(t_1,\ldots,t_n)$ will be used hereafter, so there should be no confusion.)

12. The main construction: Global form

We apply the "blowing-up" operation described at the end of Section 4 to the triangulation $\mathcal{T}$ of Proposition 10.1. Thus to each i-simplex σ^{ij} of $\mathcal{T}$ is associated a handle $H^{ij} = D^i \times D^{k-i}$. Via the map $h\colon S^k \to S^k$ collapsing each ball $\{x\} \times D^{k-i} \subset H^{ij}$ to a point $h(x) \in \sigma^{ij}$, each family $\Sigma_t(\gamma)$ associated to the g_t before blowing up and defined over a simplex σ^{ij} pulls back to a family $\Sigma_t^{ij}(\gamma)$ defined for $t \in H^{ij} \subset h^{-1}(\sigma^{ij})$. Similarly, the tongue blocks $Q_r(\gamma)$ and their shrinkings obtained in Proposition 10.1 pull back to tongue blocks $Q_r^{ij}(\gamma)$ with chosen shrinkings, defined for $t \in H^{ij}$.

A handle structure associated to $\mathcal{T}$ in the same way as $\{H^{ij}\}$ but with smaller ε_i's we refer to as a *contraction* of $\{H^{ij}\}$.

LEMMA 12.1. *After contracting $\{H^{ij}\}$, we may assume the shrinkings of the tongues of the $Q_r^{ij}(\gamma)$'s satisfy the following compatibility condition: If σ^{ij} is a* *face of* $\sigma^{i'j'}$, *then for each inclusion* $Q_{r'}^{i'j'}(\gamma') \hookrightarrow Q_r^{ij}(\gamma)$, *the chosen shrinkings of the tongues of* $Q_{r'}^{i'j'}(\gamma')$ *restrict to the chosen shrinkings of tongues of* $Q_r^{ij}(\gamma)$ (*for* $t \in H^{ij} \cap H^{i'j'}$).

Proof. Let $\{V^{ij}\}$ be a system of small open neighborhoods of a contraction $\{h^{ij}\}$ of $\{H^{ij}\}$. We may suppose each $Q_r^{ij}(\gamma)$, with the given shrinkings of its tongues, is defined over V^{ij}. By induction on i we may assume compatible shrinkings have been constructed over all $V^{i'j'}$'s with $i' < i$. These compatible shrinkings can be composed sequentially as in Lemma 6.3 to give shrinkings of the tongues of each $Q_r^{ij}(\gamma)$ over the $V^{i'j'}$'s, $i' < i$. These new shrinkings for tongues of $Q_r^{ij}(\gamma)$ have the same tangent line fields as the given shrinkings, by Proposition 10.1 and induction. So by the remarks preceding Proposition 10.1, we can deform the given shrinkings so that they agree with the new ones on slightly smaller neighborhoods $V^{i'j'}$. This is the induction step. Finally, we restrict the compatible shrinkings obtained over the neighborhoods V^{ij} to the contracted handles $h^{ij} \subset V^{ij}$. □

We note that the compatible shrinkings of the tongues of each $Q_r^{ij}(\gamma)$ given by Lemma 12.1 may be assumed still to be independent of the second coordinate in the handle $H^{ij} = D^i \times D^{k-i}$. Using these shrinkings, we apply the procedure of Section 11 to construct deformations $\Sigma_t^l(s_1, \ldots, s_n)$ of the cofactors Σ_t^l of all families $\Sigma_t^{ij}(\gamma)$, $t \in H^{ij}$. If $t \in H^{ij} \cap H^{i'j'}$ with $\sigma^{ij} \subset \partial\sigma^{i'j'}$, so that each factor $\Sigma_t^{l'} \subset \Sigma_t^{i'j'}(\gamma')$ corresponds to a factor $\Sigma_t^l \subset \Sigma_t^{ij}(\gamma)$, then we obtain, at least formally, a deformation $\Sigma_t^{l'}(s_1, \ldots, s_{n'})$ from the deformation $\Sigma_t^l(s_1, \ldots, s_n)$ by *specialization*—setting the appropriate variables s_m equal to 0 or 2, according to whether in passing from $\Sigma_t^{ij}(\gamma)$ to $\Sigma_t^{i'j'}(\gamma')$ the corresponding common face Δ_m of $\Sigma_t^{ij}(\gamma)$ is deleted from $\Sigma_t^{ij}(\gamma)$, or splits $\Sigma_t^{ij}(\gamma)$ into two components, respectively.

Proposition 12.2. *After contracting* $\{H^{ij}\}$, *we may construct deformations* $\Sigma_t^l(s_1, \ldots, s_n)$ *for the cofactors* Σ_t^l *of all families* $\Sigma_t^{ij}(\gamma)$, $t \in H^{ij}$, *satisfying*:

(∗) *If* $t \in H^{ij} \cap H^{i'j'}$ *with* $\sigma^{ij} \subset \partial\sigma^{i'j'}$, *and* $\Sigma_t^{l'} \subset \Sigma_t^{i'j'}(\gamma')$ *corresponds to* $\Sigma_t^l \subset \Sigma_t^{ij}(\gamma)$, *then the deformation* $\Sigma_t^{l'}(s_1, \ldots, s_{n'})$ *is obtained from* $\Sigma_t^l(s_1, \ldots, s_n)$ *by specialization.*

Proof. This is done inductively over the handles H^{ij}, with decreasing i. For a given $H^{ij} = D^i \times D^{k-i}$, suppose inductively that compatible deformations $\Sigma_t^{l'}(s_1, \ldots, s_{n'})$ of cofactors $\Sigma_t^{l'} \subset \Sigma_t^{i'j'}(\gamma')$ have already been constructed for $t \in D^i \times \partial D^{k-i}$ near each $\sigma^{i'j'} \supset \sigma^{ij}$. It is important to observe that such deformations $\Sigma_t^{l'}(s_1, \ldots, s_{n'})$ are in fact specialized deformations $\Sigma_t^l(s_1, \ldots, s_n)$ constructed *with respect to* H^{ij}. For, by Lemma 12.1, the only effect of passing from a higher-index handle $H^{i'j'}$ to H^{ij} is to subdivide tongue structures. (The reader should recall the procedure of §11 to be convinced of this. In fact, this is the reason why we must use decreasing i in the induction, rather than increasing i, which is more natural for specialization.) We may suppose these specialized deformations $\Sigma_t^l(s_1,\ldots, s_n)$ are defined not just for $t \in D^i \times \partial D^{k-i}$, but for t in a radial collar on $D^i \times \partial D^{k-i}$ in $D^i \times D^{k-i}$, simply by making them independent of the radial parameter in this collar. This is possible since we have blown up the original family g_t, making it independent of this radial parameter, and also the shrinkings in Lemma 12.1 are independent of this radial parameter.

To extend over H^{ij} itself we make a secondary induction. First, near $\sigma^{ij} \cap H^{ij}$, we construct deformations $\Sigma_t^l(s_1,\ldots, s_n)$ of cofactors $\Sigma_t^l \subset \Sigma_t^{ij}(\gamma)$, as in Section 11; there is no problem with compatibility here since σ^{ij} is disjoint from $D^i \times \partial D^{k-i}$. Next, consider a simplex $\sigma^{i'j'}$ meeting H^{ij}, with $i' = i + 1$. Near $\sigma^{i'j'} \cap [\sigma^{ij} \cup (D^i \times \partial D^{k-i})]$ we already have deformations $\Sigma_t^l(s_1,\ldots, s_n)$ specialized according to $\sigma^{i'j'}$. To extend these specialized deformations over a neighborhood of $\sigma^{i'j'} \cap H^{ij}$ in H^{ij}, we could carry out the procedure of Section 11 if we had relative forms of Lemmas 6.1 and 6.3. But it is easier to use a simple tapering process, as follows.

Constructing the deformations $\Sigma_t^l(s_1,\ldots, s_n)$ in Section 11 was an inductive process: first the families $\Sigma_t^l(0,\ldots, 0, s_n)$, then the families $\Sigma_t^l(0,\ldots, 0, s_{n-1}, s_n)$, etc. We can view this process as taking place during a time interval $s \in [0, 1]$, by choosing a function $\psi_t^\gamma\colon \{1, 2,\ldots, n\} \to (0, 1)$ such that $\psi_t^\gamma(l) < \psi_t^\gamma(m)$ if $\Sigma_t^l < \Sigma_t^m$, then extending each $\Sigma_t^l(0,\ldots, 0, s_{j+1},\ldots, s_n)$ to $\Sigma_t^l(0,\ldots, 0, s_j,\ldots, s_n)$ by shrinking $\Sigma_t^j(0,\ldots, 0, s_{j+1},\ldots, s_n)$ to Δ_j during the time interval $[\psi_t^\gamma(j), \psi_t^\gamma(j) + \varepsilon]$, for small enough ε. The tapering process referred to above consists of restricting the previously chosen construction of the deformations $\Sigma_t^l(s_1,\ldots, s_n)$ (specialized according to $\sigma^{i'j'}$) to a time interval $[0, s(t)]$, where $s(t)$ has support near $\sigma^{ij} \cup (D^i \times \partial D^{k-i})$ and $s(t) = 1$ in a smaller neighborhood of $\sigma^{ij} \cup (D^i \times \partial D^{k-i})$. Then on the interval $[s(t), 1]$ we continue afresh the construction of the specialized deformations $\Sigma_t^l(s_1,\ldots, s_n)$ for t near $\sigma^{i'j'}$ in H^{ij}. This has the effect of eliminating the need for relative forms of Lemmas 6.1 and 6.3.

Having taken care of $(i + 1)$-simplices meeting H^{ij}, we proceed in the same way with $(i + 2)$-simplices, etc. Thus the process can be completed over H^{ij}. When all handles H^{ij} have been treated, we replace $\{H^{ij}\}$ by a sufficiently small contraction of itself, and we obtain the compatibility condition $(*)$. □

The number ε in Section 2 can be chosen small enough so that a set $Z_0(e) \times [-1, 1]$ meets a handle H^{ij} only if the corresponding simplex σ^{ij} is contained in $Z_0(e)$. Likewise we may assume $H^{ij} \cap (\tilde{Y}(e) \times [-1, 1]) \neq \emptyset$ only if $\sigma^{ij} \subset Y(e)$.

13. Construction of $\bar{g}_{t0}$: $\bar{S}^c_{t0} \to \bar{\Sigma}^c_{t0}$

The machinery is all ready now to complete the proof. The first step is to construct the family Σ^c_{t0}, mimicking S^c_{t0}. For a handle H^{ij} as in Section 12, consider a family $\Sigma^{ij}_t(\gamma)$ defined for $t \in H^{ij}$. For each edge e_l of γ we have $\sigma^{ij} \subset Z_0(e_l) \times [-1, 1]$, and we can extend the product structure $Z_0(e_l) \times [-1, 1]$ to $Z_0(e_l) \times \mathbf{R} \supset H^{ij}$. For $t \in Z_0(e_l) \times \{r\}$, let:

$$t_l = \begin{cases} r & \text{if } \; |r| \le 1 \\ 1 & \text{if } \; r > 1 \\ -1 & \text{if } \; r < -1. \end{cases}$$

In case γ has a base vertex, we obtain Σ^c_{t0} in two steps:

(1) Take the disjoint union of all $\Sigma_l(t_1, \ldots, t_n)$ as e_l ranges over edges of γ and identify their common faces $\Delta_m(t_1, \ldots, t_n)$, for $t_m \ge 0$. (In general these factors $\Sigma_l(t_1, \ldots, t_n)$ can have intersections in $\mathbf{R}^3$ besides the obvious ones consisting of a common face $\Delta_m(t_1, \ldots, t_n)$ which is the intersection of the two factors containing it.)

(2) For the "shock wave" effect of Figure 3.4, subdivide $\Sigma_l(t_1, \ldots, t_n)$ for $t_l \le 0$ by adjoining the disc $\Delta_l(t_1, \ldots, \tau_l(t_l), t_{l+1}, \ldots, t_n)$, where τ_l: $[-1, 0] \to [0, 1]$ is chosen appropriately. Or more precisely, take the small vertical deformation of this disc obtained from the small vertical (ambient) isotopy of the disc $\Delta_l \subset \Sigma_l(t_1, \ldots, t_n)$ as t_l goes from 0 to -1.

In case γ has a base edge, the procedure is just the same except that Figure 3.3 serves as the model for choosing the deformations $\Delta_1(t_1, \ldots, t_n)$ of the face Δ_1 corresponding to the base edge of γ.

This defines Σ^c_{t0} for $t \in H^{ij}$. By the constructions of Section 12, the result depends only on t, not H^{ij}. So we obtain Σ^c_{t0}, well-defined and varying continuously with $t \in S^k$.

Next we construct foliations Φ_{t0} on $\bar{\Sigma}^c_{t0}$. For factors of Σ^c_{t0} not corresponding to polar factors of S^c_{t0}, we have already at hand shrinkings of these factors $\Sigma_l(t_1, \ldots, t_n)$ to their preferred faces $\Delta_l(t_1, \ldots, t_n)$ as part of the process of constructing $\Sigma_l(t_1, \ldots, t_n)$. So for such factors we let Φ_{t0} be transverse to the stages of the shrinking. And for factors $\Sigma_0(t_1, \ldots, t_n)$ of Σ^c_{t0} corresponding to polar factors of S^c_{t0}, we know from Section 11 that their contours retain disc-with-tongues structures with varying $(t_1, \ldots, t_n)$; so by the same methods we can obtain shrinkings of these factors to points, lifting shrinkings of their contours to points in their initial discs. And again we take Φ_{t0} transverse to these shrinkings. We can take Φ_{t0} to be a standard pole at its singularity in a factor $\bar{\Sigma}_0(t_1, \ldots, t_n)$ since the shrinking of this factor can be chosen to end with a standard family of concentric spheres.

We remark that the degrees of differentiability lost in Section 10, where we got only C^1 shrinkings, can now be recovered. For Φ_{t0} can easily be chosen to be C^∞, since it is required only to be transverse to a continuous family of tangent planes. Then having $\Phi_{t0} C^\infty$ we can perturb the family Σ_{t0}^c to be C^∞, staying transverse to Φ_{t0}.

Since $\mathrm{Diff}(D^2 \operatorname{rel} \partial D^2)$ is contractible, there are no obstructions to constructing a family of diffeomorphisms $S_{t0} \to \Sigma_{t0}$ agreeing with g_t on $S_{t0} \cap S^2$, and then extending to a family of diffeomorphisms g_{t0}: $S_{t0}^c \to \Sigma_{t0}^c$. We may suppose that when a common face D between factors of S_{t0}^c contains a pole of F_{t0}, and hence is a union of leaves of F_{t0}, then Φ_{t0} agrees with $g_{t0}(F_{t0})$ on $g_{t0}(D)$. Further, by modifying our choice of Φ_{t0} if necessary (still keeping it transverse to the same shrinkings) we may use contractibility of $\mathrm{Diff}(D^2 \operatorname{rel} \partial D^2)$ again to assure that whenever two points of a factor of S_{t0}^c are joined by a leaf of F_{t0}, then their images under g_{t0} are joined by a leaf of Φ_{t0}.

With these relations between F_{t0} and Φ_{t0} there is, clearly, a family of homeomorphisms $\bar{g}_{t0}$: $\bar{S}_{t0}^c \to \bar{\Sigma}_{t0}^c$ extending g_{t0} and taking F_{t0} to Φ_{t0}, which are diffeomorphisms on factors (smoothness at corners can easily be arranged by modeling Φ_{t0} on F_{t0} there) and vary continuously with t in the C^∞ topology. Thus the hypotheses of Proposition 4.1 are satisfied, and the theorem is proved.

Appendix. Some equivalent forms of the Smale Conjecture

As is well-known and easily shown, $\mathrm{Diff}(S^n) \simeq O(n+1) \times \mathrm{Diff}(D^n \operatorname{rel} \partial D^n)$ for any n. Hence the Smale Conjecture $\mathrm{Diff}(S^3) \simeq O(4)$ is equivalent to:

(1) $\mathrm{Diff}(D^3 \operatorname{rel} \partial D^3) \simeq *$.

We can see that (1) is equivalent to the theorem of this paper, as follows. The theorem can be restated as asserting that for the (smooth) embedding spaces $\mathrm{Emb}(D^3, \mathbf{R}^3)$ and $\mathrm{Emb}(S^2, \mathbf{R}^3)$, the restriction map ρ: $\mathrm{Emb}(D^3, \mathbf{R}^3) \to \mathrm{Emb}(S^2, \mathbf{R}^3)$ induces a surjection on π_k for all k. From the diagram

$$\begin{array}{ccc} \mathrm{Emb}(D^n, \mathbf{R}^n) & \xrightarrow{\ \rho\ } & \mathrm{Emb}(S^{n-1}, \mathbf{R}^n) \\ & \searrow^{\simeq} \quad \swarrow & \\ & \mathrm{GL}(n, \mathbf{R}) & \end{array}$$

we see that ρ is injective on π_k (for any n). Here the lower left arrow is "evaluate derivative at a point", clearly a homotopy equivalence, and the lower right arrow is "evaluate derivative at a point and adjoin the normal vector." Also, ρ is a fibration whose fiber is just $\mathrm{Diff}(D^n \operatorname{rel} \partial D^n)$. So $\mathrm{Diff}(D^3 \operatorname{rel} \partial D^3) \simeq *$ if and only if ρ is surjective on all π_k's.

According to a theorem of Morlet (see [6] for discussion and references), $\mathrm{Diff}(D^n \operatorname{rel} \partial D^n) \simeq \Omega^{n+1}(\mathrm{PL}(n)/O(n))$ for any n. So (1) is equivalent to:

(2) $\mathrm{PL}(3) \simeq O(3)$.

A consequence of this (see [6] again) is:

(3) $\mathrm{Diff}(M^3 \operatorname{rel} \partial M^3) \simeq \mathrm{PL}(M^3 \operatorname{rel} \partial M^3)$ for any 3-manifold M^3, where $\mathrm{PL}(M \operatorname{rel} \partial M)$ is the simplicial group of PL homeomorphisms $M \to M$ fixed on ∂M.

Of course, (3) implies (1) when $M^3 = D^3$. (In (2) and (3) one could just as well take the category TOP instead of PL. The implication (1) $\Rightarrow \mathrm{Diff}(M^3 \operatorname{rel} \partial M^3) \simeq \mathrm{TOP}(M^3 \operatorname{rel} \partial M^3)$ goes back to Cerf [3], and was the starting point for Morlet's work.)

We return to the C^∞ category for the next several forms of the Smale conjecture.

(4) The space of smoothly embedded 2-spheres in $\mathbf{R}^3$ is contractible.

This space is the orbit space $\mathrm{Emb}(S^2, \mathbf{R}^3)/\mathrm{Diff}(S^2)$. Since $\mathrm{Diff}(S^2) \simeq O(3)$, contractibility of this orbit space is equivalent to the map ρ above being a homotopy equivalence.

From the fibration

$$\begin{array}{ccccc} \mathrm{Diff}(D^3_+ \operatorname{rel} \partial D^3_+) & \to & \mathrm{Emb}(D^3_+, D^3 \operatorname{rel} S^2_+) & \to & \mathrm{Emb}(D^2, D^3 \operatorname{rel} \partial D^2) \\ \wr\! \mid & & \wr\! \mid & & \\ \mathrm{Diff}(D^3 \operatorname{rel} \partial D^3) & & * & & \end{array}$$

where $D^n = D^n_+ \cup D^n_-$, $D^n_+ \cap D^n_- = D^{n-1}$, and $S^{n-1}_\pm = D^n_\pm \cap S^{n-1}$, we see that $\mathrm{Diff}(D^3 \operatorname{rel} \partial D^3) \simeq \Omega\, \mathrm{Emb}(D^2, D^3 \operatorname{rel} \partial D^2)$. So (1) is equivalent to:

(5) $\mathrm{Emb}(D^2, D^3 \operatorname{rel} \partial D^2)$ is contractible.

Next consider the fibration

$$\begin{array}{ccccc} \mathrm{Emb}(D^2_+, D^3 \operatorname{rel} \partial D^2_+) & \to & \mathrm{Emb}(D^2_+, D^3 \operatorname{rel} S^1_+) & \to & \mathrm{Emb}(D^1, D^3 \operatorname{rel} \partial D^1) \\ & & \wr\! \mid & & \\ & & * & & \end{array}$$

According to [5], (adapted to the smooth category as in [7]), the inclusion $\mathrm{Emb}(D^2_+, D^3 - D^2_- \operatorname{rel} \partial D^2_+) \hookrightarrow \mathrm{Emb}(D^2_+, D^3 \operatorname{rel} \partial D^2_+)$ is a homotopy equivalence. The smaller of these two spaces can be identified with $\mathrm{Emb}(D^2, D^3 \operatorname{rel} \partial D^2)$. Hence

$$\mathrm{Emb}(D^2, D^3 \operatorname{rel} \partial D^2) \simeq \Omega\, \mathrm{Emb}(D^1, D^3 \operatorname{rel} \partial D^1),$$

and (5) is equivalent to:

(6) The identity component of $\mathrm{Emb}(D^1, D^3 \operatorname{rel} \partial D^1)$ (consisting of the unknotted arcs) is contractible.

(It follows from the Smale Conjecture and [5] that the other components of $\text{Emb}(D^1, D^3 \text{rel } \partial D^1)$ are aspherical, though their π_1's can be quite non-trivial.)

There is an easy homotopy equivalence

$$\text{Emb}(S^1, S^3) \simeq \text{Emb}(D^1, D^3 \text{rel } \partial D^1) \times O(4)/O(2),$$

quite similar to $\text{Diff}(S^3) \simeq \text{Diff}(D^3 \text{rel } \partial D^3) \times O(4)$. Since $\text{Diff}(S^1) \simeq O(2)$, this yields

$$\text{Emb}(S^1, S^3)/\text{Diff}(S^1) \simeq \text{Emb}(D^1, D^3 \text{rel } \partial D^1) \times O(4)/(O(2) \times O(2));$$

hence (6) is equivalent to:

(7) The space of smoothly embedded unknotted circles in S^3 deformation retracts onto the space of great circles in S^3 (i.e., $O(4)/O(2) \times O(2)$).

We leave to the reader proofs that the following statements are all equivalent to the Smale Conjecture:

(8) $\text{Diff}(S^2 \times D^1 \text{rel } \partial) \simeq \Omega O(3)$.

(9) $\text{Diff}(S^1 \times D^2 \text{rel } \partial) \simeq *$.

(10) $\text{Emb}(S^2, S^2 \times \mathbf{R}) \simeq O(3)$.

(11) $\text{Diff}(S^2 \times \mathbf{R}) \simeq O(3) \times O(1)$.

(12) $\text{Emb}(S^1, S^1 \times \mathbf{R}^2) \simeq O(2)$.

(13) $\text{Diff}(S^1 \times \mathbf{R}^2) \simeq O(2) \times O(2) \times \Omega O(2)$.

(14) The space of smooth 2-spheres in S^3 deformation retracts onto the subspace of great 2-spheres.

(15) The space of smooth unknotted circles in $\mathbf{R}^3$ deformation retracts onto the subspace of great circles in S^2.

(16) $\text{PL}(S^3) \simeq O(4)$ (or $\text{TOP}(S^3) \simeq O(4)$).

(17) The PL (or TOP) versions of (4), (7), (8), (10), (11), (14), (15).

University of California, Los Angeles

References

[1] J. W. Alexander, On the subdivision of 3-space by a polyhedron, P.N.A.S. **10** (1924), 6–8.

[2] J. Cerf, *Sur les Difféomorphismes de la Sphère de Dimension Trois* ($\Gamma_4 = 0$), Springer Lecture Notes, **53** (1968).

[3] ———, Groupes d'automorphismes et groupes de difféomorphismes des variétés compactes de dimension 3, Bull. Soc. Math. France **87** (1959), 319–329.

[4] A. Hatcher, Linearization in 3-dimensional topology, Proc. I.C.M. Helsinki (1978), 463–468.

[5] ———, Homeomorphisms of sufficiently large P^2-irreducible 3-manifolds, Topology **15**, (1976), 343–347.

[6] R. Kirby and L. Siebenmann, *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations*, Annals of Math. Studies **88** (1977), Princeton Univ. Press.
[7] F. Laudenbach and O. Zambrano, Disjonction à paramètres dans une variété de dimension 3, P^2-irréductible, C.R. Acad. Sci. Paris **288** A (1979), 331–334.
[8] S. Smale, Diffeomorphisms of the 2-sphere, Proc. A.M.S. **10** (1959), 621–626.
[9] ———, review of the paper [3], Math Reviews **22** (1961), 1218.

(Received June 29, 1981)
(Revised September 23, 1982)