PUBLICATIONS MATHÉMATIQUES DE L'I.H.É.S.

MATTHEW GRAYSON
CHARLES PUGH

Critical sets in 3-space

Publications mathématiques de l'I.H.É.S., tome 77 (1993), p. 5-61
<http://www.numdam.org/item?id=PMIHES_1993__77__5_0>

CRITICAL SETS IN 3-SPACE

by MATTHEW GRAYSON *and* CHARLES PUGH

Abstract. — Given a non-empty compact set $C \subset \mathbf{R}^3$, is C the set of critical points for some smooth proper function $f: \mathbf{R}^3 \to \mathbf{R}_+$? In this paper we prove that the answer is "yes" for Antoine's Necklace and most but not all tame links.

1. Introduction

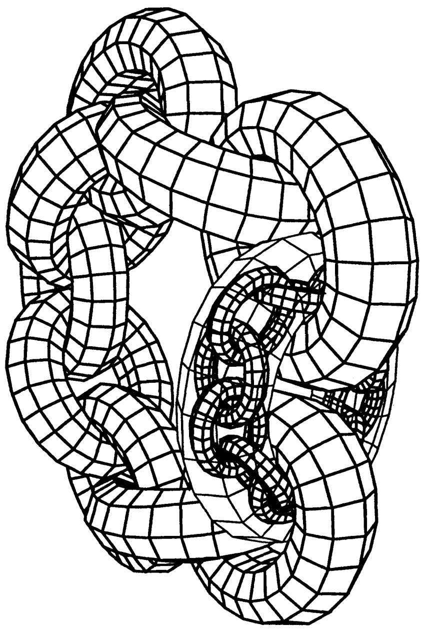
Given a non-empty compact set $C \subset \mathbf{R}^m$, is C the set of critical points for some smooth (C^∞) proper function $f: \mathbf{R}^m \to \mathbf{R}_+$? If $m = 1$, the answer is easily seen to be "yes", always. If $m = 2$, Norton and Pugh (1991) show that the answer is "yes" if and only if no component of $\mathbf{R}^2 \backslash C$ is simply connected. In this paper some results are given in dimension $m = 3$, the most remarkable being that for Antoine's Necklace the answer is "yes it is a critical set". In dimension $m \geqslant 4$ there are doubtlessly generalizations of what we do below but stronger hypotheses will be needed and we feel that $m = 3$ provides enough problems already.

To be more precise, recall that $f: \mathbf{R}^m \to \mathbf{R}$ is *critical* at a point $p \in \mathbf{R}^m$ if and only if its derivative at p is zero, $(Df)_p = 0$. The set of all critical points is denoted by $\mathrm{cp}(f)$. The f-image of $\mathrm{cp}(f)$ is the set of *critical values*, $\mathrm{cv}(f) := f(\mathrm{cp}(f))$. The Morse-Sard Theorem concerns $\mathrm{cv}(f)$, not $\mathrm{cp}(f)$, and asserts that $\mathrm{cv}(f)$ has zero measure if f is at least of class C^{m+1}. Note that this implies that $\mathrm{cv}(f)$ is compact and totally disconnected when f is smooth and $\mathrm{cp}(f)$ is compact.

A *proper* function has the property that $f^{-1}(K)$ is compact for all compact sets K in its target space. Equivalently, when $f: \mathbf{R}^m \to \mathbf{R}$, $|f(x)| \to \infty$ as $|x| \to \infty$ and vice versa. We say that $C \subset \mathbf{R}^m$ is *critical* if and only if $C = \mathrm{cp}(f)$ for some smooth function $f: \mathbf{R}^m \to \mathbf{R}$ and *properly critical* if f can be chosen to be proper.

Recall from Rolfsen (1976) that *Antoine's Necklace* is a wild Cantor set in $\mathbf{R}^3$ constructed as follows. Starting with the solid torus R, draw a cyclic chain of small linked solid tori $R_0, \ldots, R_{n-1}$ around the longitudinal core of R; call $R^{(1)} = R_0 \cup \ldots \cup R_{n-1}$. In the figure $n = 10$. Inside each R_i, repeat the picture, scaled down to the size of R_i. Call the resulting union of n^2 very small solid tori $R^{(2)}$. Continue. Antoine's Necklace is by definition

$$A = \bigcap R^{(k)}.$$



A has some bizarre properties. Although it is totally disconnected, no embedded 2-sphere in $\mathbf{R}^3$ is able to separate it—it is " indivisible " by 2-spheres. (It *is* divisible by 2-tori.) Besides, any loop γ in $\mathbf{R}^3$ that links one of the tori in the construction of A, also links A in the sense that γ cannot be shrunk to a point in $\mathbf{R}^3\backslash A$; the fundamental group of $S^3\backslash A$ is non-trivial. Even though A has topological dimension zero, it acts as if it were a curve.

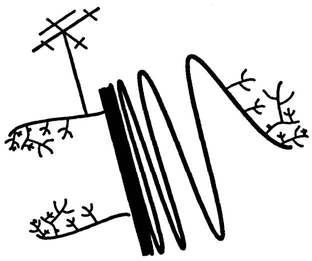
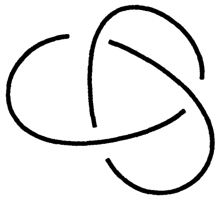
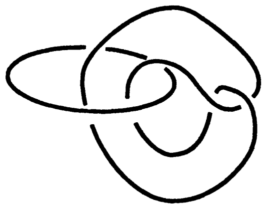
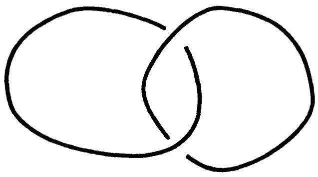
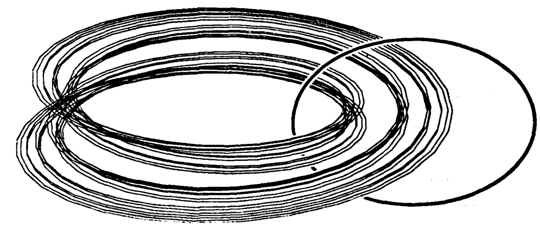
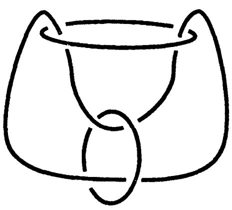
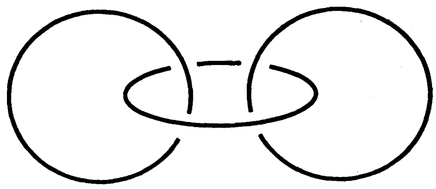
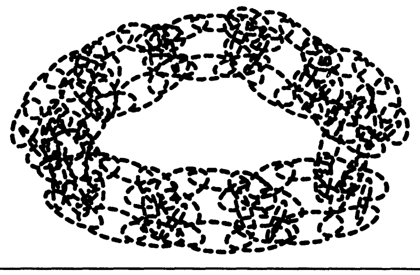
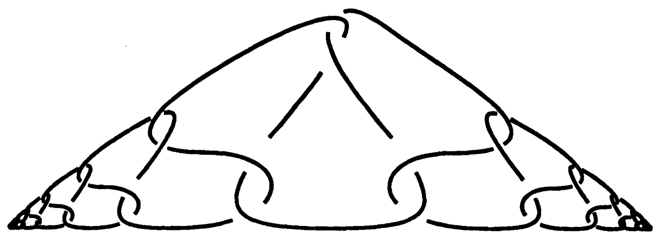
Theorem A. — Antoine's Necklace is properly critical. (See § 5 for the proof.)

Theorem B. — All tame links in $\mathbf{R}^3$ *are critical and some but not all are properly critical.* (Below are tables summarizing what we know. The proofs appear in § 2, 4.)

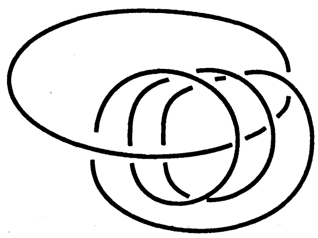
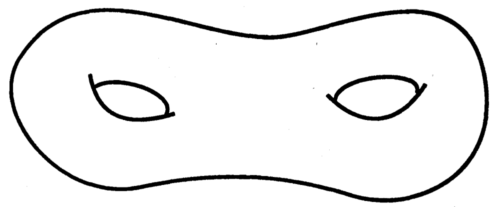
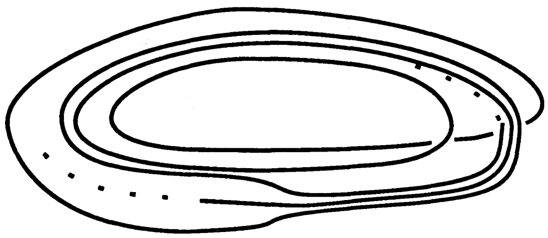
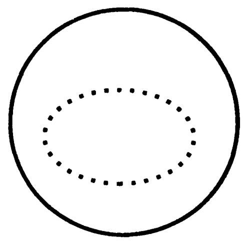
If we relax the smoothnes hypothesis from C^∞ to C^r where $r < \dim(\mathbf{R}^m)$ then an example of Whitney (1932) suggests that critical sets can be more general than those we consider here. In particular, Harrison and Pugh (1990) show that a fractal circle can be the critical set for a C^1 function on $\mathbf{R}^2$. This is impossible in the C^∞ case as is shown in Theorem B.

One may view the question of classifying critical sets as part of " Morse Theory with degenerate singularities ". It also has an interpretation in dynamical systems in terms of chain recurrence. A point p is *chain recurrent* under a flow φ if for any $\varepsilon > 0$ there is an ε-chain from p to itself, an *ε-chain* being a sequence of trajectory arcs $\gamma_i = \{ \varphi_t(p_i) : 0 \leqslant t \leqslant t_i \}$, $i = 0, \ldots, n$, where $t_i > 1$ and the distance from the end

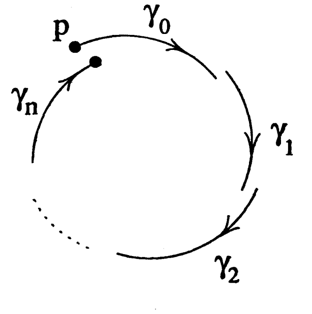
Properly critical

Yes	No
1. Any cellular set. 	1. A circle or knot. 
2. The unlink. 	2. Any 2-component link except the Hopf link and the unlink. 
3. The Hopf link. 	3. The *p*-adic solenoid link. 
4. All tame n-component links, n $\geqslant$ 3. 	4. A chain of three circles with only two critical values. 
5. Antoine's Necklace. 	5. Some wild arcs. 

Improperly critical

Yes	No
1. Any tame link. 	1. Any compact surface. 
2. The Denjoy solenoid. 	2. S^2 with a circle enclosed. 

$\varphi_{t_i}(p_i)$ of γ_i to the beginning p_{i+1} of γ_{i+1} is $< \varepsilon$ for all i. To say that the ε-chain $\gamma_0, \gamma_1, \ldots, \gamma_n$ goes from p to itself means that $p_0 = p$ and the distance from the last end $\varphi_{t_n}(p_n)$ to p is $< \varepsilon$. Chain recurrence is the most general form of recurrence occurring in dynamics. The set of chain recurrent points, $CR(\varphi)$, contains all the fixed points, periodic orbits, ω-limit sets, α-limit sets, homoclinic orbits, and non-wandering orbits. According to Conley (1976) and Wilson and Yorke (1973), given a smooth flow φ on a compact manifold, there is a *global Lyapunov function* f for φ. The function f is real valued, smooth, increases strictly along all non-chain-recurrent orbits of φ, and is critical exactly on $CR(\varphi)$. Thus, every chain recurrent set is a critical set. The converse is clear: for if φ is the flow generated by the vector field $\mathrm{grad}(f)$ then its only chain recurrent points are its fixed points, and they are exactly the zeros of $\mathrm{grad}(f)$; i.e.,



CR(grad(f)-flow) $=$ cp(f). If f is proper then ∞ is a sink (an attractor) for the gradient flow considered on $S^3 = \mathbf{R}^3 \cup \infty$. The upshot is

Classifying properly critical sets in $\mathbf{R}^3$ amounts to classifying chain recurrent sets for flows in $\mathbf{R}^3$ with a sink at ∞.

Corollary. — If a Morse-Smale flow on S^3 has exactly two closed orbits and a point sink then its closed orbits form a Hopf link or the unlink, never other links such as the Whitehead link.

1.1. Section 2 starts with some basic results and tools. Most of the results are positive; they say what sets are properly critical. It concludes with the proof that all $\geqslant 3$ component links are properly critical, and with some curious examples. Section 3 is a technical section with machinery necessary for the 2-component link case, proved in section 4. Section 3A is an appendix to section 3 which demonstrates by an example how difficult it can be to make Antoine's Necklace properly critical. As previously noted, Theorem A is proved in section 5.

We thank Elise Cawley, Peter Jones, Bill Massey, Mike Shub, and Bill Thurston for several valuable comments.

2. Initial answers

In this section we dispose of some of the assertions in Theorem B. The first lemma lets us eliminate certain critical points of a smooth function $f: M \to \mathbf{R}$. If in some coordinate system $(x^1, \ldots, x^m)$, $f = f(x^1, \ldots, x^m)$ is a strictly increasing function of one of the variables then its critical points there are " clearly irrelevant ". More precisely, we say that f has only *superfluous* critical points in $N \subset M$ if there is a smooth coordinate system $(x^1, \ldots, x^m)$ in which N appears to be a product,

$$N = \{(x^1, \ldots, x^{m-1}) \in N_0\} \times \{x^m \in [a, b]\},$$

N_0 being a compact subset of $\mathbf{R}^{m-1}$, and

$$\frac{\partial f(x)}{\partial x^m} \geqslant 0 \text{ if } x \in N \quad \text{while } \frac{\partial f(x)}{\partial x^m} > 0 \text{ if } x \in N \text{ and } x^m = b. \tag{1}$$

2.1. *Erasing Lemma. — If $f: M \to \mathbf{R}$ is smooth and obeys* (1) *then f can be C^∞-approximated by $F: M \to \mathbf{R}$ such that $F = f$ off N and $\frac{\partial F}{\partial x^m} > 0$ on int*(N). (All critical points interior to N get *erased.*)

Proof. — By compactness, there exist $\mu, \nu > 0$ such that $\frac{\partial f(x)}{\partial x^m} \geqslant \mu$ on $N_0 \times [b - \nu, b]$. Let $\alpha : \mathbf{R}^{m-1} \to [0, 1]$ and $\beta : \mathbf{R} \to [0, 1]$ be smooth bump functions such that α is positive on $\mathrm{int}(N_0)$, and zero off N_0; β is positive on (a, b), zero off $[a, b]$, and $\beta' > 0$ on the interval $(a, b - \nu]$. Set

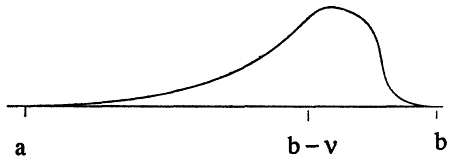
$$F(x) = f(x) + \varepsilon\alpha(x^1, \ldots, x^{m-1})\,\beta(x^m),$$

where $x = (x^1, \ldots, x^m)$ and $\varepsilon > 0$ is small. Clearly, $F = f$ off N and $F \to f$ in the C^∞ sense as $\varepsilon \to 0$. Also, for ε small, $|\varepsilon\beta'| < \mu$ and so

$$\frac{\partial F}{\partial x^m} = \frac{\partial f}{\partial x^m} + \varepsilon\alpha\beta' > 0 \quad \text{on } N_0 \times [b - \nu, b],$$

while on $\mathrm{int}(N_0) \times (a, b - \nu]$, $\beta' > 0$ and $\dfrac{\partial f}{\partial x^m} \geqslant 0$ imply that $\dfrac{\partial F}{\partial x^m} > 0$. Thus $\dfrac{\partial F}{\partial x^m} > 0$ on all of $\mathrm{int}(N)$.

QED



2.2. *Theorem (Enlargement of critical sets).* — *If* $f : M \to \mathbf{R}$ *is smooth and* C *is a compact subset of some level set* $f^{-1}(c)$, *then there is a smooth function* $F : M \to \mathbf{R}$ *such that*

$$\mathrm{cp}(F) = \mathrm{cp}(f) \cup C.$$

If f *is proper, so is* F.

Proof. — Let $g : \mathbf{R} \to \mathbf{R}$ be a smooth homeomorphism such that $g' > 0$ except at c, where $g'(c) = 0$. Set $f_0 = g \circ f$. Clearly f_0 is smooth and

$$\mathrm{cp}(f_0) = \mathrm{cp}(f) \cup f^{-1}(c).$$

We are going to use 2.1 to erase superfluous critical points in $P = f^{-1}(c)\setminus(C \cup \mathrm{cp}(f))$. For each $p \in P$ we can find a coordinate system $(x^1, \ldots, x^m)$ on a neighborhood N_p of p which is a *flowbox* for the grad(f)-flow; i.e., for some interval $[a, b]$ with $a < c < b$ and some $\delta > 0$,

$$f(x^1, \ldots, x^m) = x^m \text{ if } a \leqslant x^m \leqslant b \text{ and } |(x^1, \ldots, x^{m-1})| \leqslant \delta.$$

Then $\dfrac{\partial f_0(x)}{\partial x^m} \geqslant 0$ if $x \in N_p$, while $\dfrac{\partial f_0(x)}{\partial x^m} > 0$ if $x \in N_p$ and $x^m \neq c$. In the flowbox coordinate system, the neighborhood N_p corresponds to the product of the δ-disc in $\mathbf{R}^{m-1}$ and the interval $[a, b]$. We take a locally finite cover of P by the interiors of such neighborhoods N_p, say $N_1, N_2, \ldots$, where $N_n = N_{p_n}$, making sure that each N_n misses $C \cup \mathrm{cp}(f)$. Working in N_1 we use 2.1 to replace f_0 by F_1. Since flowboxes overlap naturally, $\dfrac{\partial F_1}{\partial x^m}$ remains non-negative when judged in any of the other flowbox coordinate systems N_n that meet N_1. This lets us continue, replacing F_1 by F_2 in N_2, and so on, never introducing new critical points where old ones were erased.

Since $\{N_n\}$ is locally finite, $F_n(x)$ is independent of n for all $n \geqslant$ some $n_0(x)$. Thus, $F = \lim_{n\to\infty} F_n$ is well defined. For safety's sake we can also require that the C^r-size of $F_r - F_{r-1}$ is $\leqslant 2^{-r}$ respecting some fixed C^r norm on functions defined on M. Then F is C^∞ and has no critical points in $U = \bigcup \operatorname{int}(N_n)$. Critical points of f_0 off U are unaffected by this construction and so we see that F has all the old critical points of f_0 in U^c, but has no critical points in U. Since $\{N_n\}$ covers $P = f^{-1}(c)\backslash(C \cup \operatorname{cp}(f))$ and $N_n \cap (C \cup \operatorname{cp}(f)) = \emptyset$, it follows that $\operatorname{cp}(F) = C \cup \operatorname{cp}(f)$ as claimed.

If f is proper then so is $f_0 = g \circ f$, and since F approximates f_0, so is F.

QED

2.3. *Theorem. — The Hopf link is properly critical.*

Proof. — It suffices to show that some ambiently diffeomorphic copy of the Hopf link is properly critical, for the property of being critical or properly critical is clearly invariant under ambient diffeomorphism. Consider the function $f: \mathbf{R}^3 \to \mathbf{R}$ defined as

$$f(x, y, z) = z^4 - 2(x^2 + y^2) + (x^2 + y^2)^2 + 1.$$

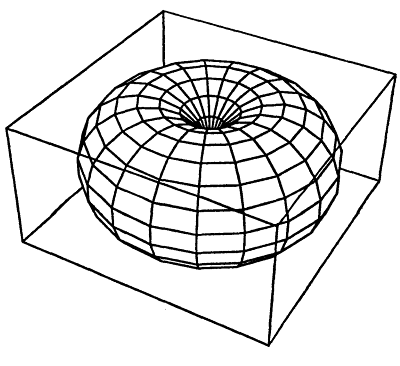
It is easy to see that $|f| \to \infty$ as $|(x, y, z)| \to \infty$. That is, f is proper. The critical points of f are found from

$$-4x + 2(x^2 + y^2)(2x) = 0, \quad -4y + 2(x^2 + y^2)(2y) = 0, \quad 4z^3 = 0,$$

and these equations hold if and only if $x^2 + y^2 = 1$ and $z = 0$, or $(x, y, z) = (0, 0, 0)$. Calling A the unit circle in the $z = 0$ plane, we see that $f(A) = 0$, $f(0, 0, 0) = 1$, $\operatorname{cp}(f) = A \cup \{(0, 0, 0)\}$. The level set $f^{-1}(1)$ is the *bagel pinched* torus shown. The doughnut hole has been shrunk to a point to prevent butter leaking out. Except at the origin, $f^{-1}(1)$ is a regular surface.

On $f^{-1}(1)$, we consider $B = \{(x, 0, z) : z^4 + x^4 - 2x^2 + 1 = 1\}$. It is a smooth unknotted Jordan curve that links A once. By Theorem 2.2 we can modify f to a function $F: \mathbf{R}^3 \to \mathbf{R}_+$ so that $\operatorname{cp}(F) = \operatorname{cp}(f) \cup B$; i.e., $\operatorname{cp}(F) = A \cup B$, a Hopf link.

QED



Remark. — We believe that an analytic expression for a properly critical Hopf link could be found, but we leave it for the interested reader to do so. The general question of which critical sets occur for analytic functions is probably much harder than the C^∞ case treated in this paper.

2.4. *Theorem.* — *The unlink is properly critical.*

Proof. — Again, it suffices to show that a diffeomorphic copy of the unlink is properly critical. Consider the function $f : \mathbf{R}^3 \to \mathbf{R}$ defined as

$$f(x, y, z) = z^4 - 4(x^2 + y^2)\, z + 2(x^2 + y^2)^2 + 1.$$

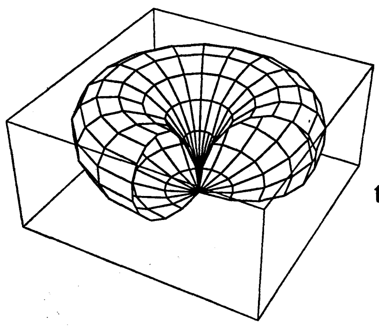
It is not hard to check that f is proper. (Break things down into the cases that $x^2 + y^2 \leqslant 4\,|\,z\,|$ and $x^2 + y^2 \geqslant 4|\,z\,|$.) The critical points of f are found from

$$-8xz + 4(x^2 + y^2)\,(2x) = 0, \qquad -8yz + 4(x^2 + y^2)\,(2y) = 0,$$
$$4z^3 - 4(x^2 + y^2) = 0.$$

These equations hold if and only if $x^2 + y^2 = 1$ and $z = 1$ or $(x, y, z) = (0, 0, 0)$. Thus, $f(\mathrm{A}) = 0$ where A is the unit circle in the $z = 1$ plane; and $f(0, 0, 0) = 1$. The level set $f^{-1}(1)$ is a *bialy pinched torus*—it has a smooth bottom face. The bagel pinch point lies midway up the hole while for the bialy it lies at the bottom. (See the remark that follows for more details on its shape.)

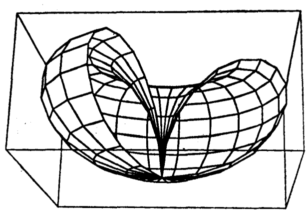
Let B be any smooth Jordan curve on the level set $f^{-1}(1)$ passing through $(0, 0, 0)$. (For instance, we could take for B the intersection of the lower branch of $f^{-1}(1)$ with a vertical cylinder $(x - \rho)^2 + y^2 = \rho^2$ having small radius ρ.) By Theorem 2.2 there is a proper smooth function $\mathrm{F} : \mathbf{R}^3 \to \mathbf{R}_+$ with $\mathrm{cp}(\mathrm{F}) = \mathrm{cp}(f) \cup \mathrm{B}$.

QED



top view

bottom view



Remark. — Let us say a little more about the bialy, especially its bottom face. Introducing cylindrical coordinates, the equation for the bialy becomes

$$z^4 - 4r^2 z + 2r^4 = 0.$$

To check the bialy's shape at the origin, one would like to solve this equation there for $z = z(r)$, but this is difficult since its z-partial derivative vanishes at $(0, 0)$. However, one can write $r = r(z)$ as

$$r^2 = z\sqrt{\left\{1 \pm \sqrt{\left(1 - \frac{z^2}{2}\right)}\right\}}, \qquad z \geqslant 0.$$

The choice of $+\sqrt{\ }$ corresponds to the bottom branch of the bialy while $-\sqrt{\ }$ corresponds to the pinched upper branch. Thus, for (r, z) on the bottom face, we have $zg(z) - r^2 = 0$ where $g = \sqrt{\left\{1 + \sqrt{\left(1 - \frac{z^2}{2}\right)}\right\}}$ is analytic at $z = 0$. The z-partial of this equation at $(0, 0)$ is $g(0) = \sqrt{2} \neq 0$ and so one has an analytic solution $z = z(r^2)$. Since $z(0) = 0$, we see that to second order the bottom of the bialy is a paraboloid through $(0, 0, 0)$.

In the preceding proofs we observed that proper criticality is invariant under ambient diffeomorphism. Under non-ambient diffeomorphism, criticality and proper criticality may be lost. For example, a sphere plus its center point is properly critical, but a sphere plus a point outside it is not critical at all. In a different direction the next result drops from ambient diffeomorphism to ambient homeomorphism.

2.5. *Theorem.* — *If* C, C′ *are compact subsets of* $\mathbf{R}^3$ *and* $h : \mathbf{R}^3 \circlearrowleft$ *is a homeomorphism sending* C′ *to* C, *then* C *is properly critical if and only if* C′ *is.*

Proof. — Since we are working in $\mathbf{R}^3$, not $\mathbf{R}^4$, we may replace h with another ambient homeomorphism $\mathrm{H} : \mathbf{R}^3 \circlearrowleft$ which sends C′ to C and is a diffeomorphism of C'^c to C^c. See Munkres (1972). We assume that $\mathrm{C} = \mathrm{cp}(f)$ for some smooth proper $f : \mathbf{R}^3 \to \mathbf{R}_+$ and produce another smooth proper function on $\mathbf{R}^3$ whose critical point set is C′. Let $\mathrm{V} = \mathrm{cv}(f) = f(\mathrm{C})$.

We pull f back by H, getting $f \circ \mathrm{H} : \mathbf{R}^3 \to \mathbf{R}_+$. We know that $f \circ \mathrm{H}$ is continuous. Restricted to C'^c it is smooth and regular. By the Morse-Sard Theorem, $\mathrm{V} \subset \mathbf{R}$ is compact, totally disconnected and so we can find a smooth orientation-preserving homeomorphism $g : \mathbf{R} \to \mathbf{R}$ which is very flat at V and otherwise is a diffeomorphism. By Lemma 3 of Norton and Pugh it follows that $\mathrm{F} = g \circ f \circ \mathrm{H}$ is smooth on all of $\mathbf{R}^3$. Its set of critical values is $\mathrm{cv}(\mathrm{F}) = g(\mathrm{V})$ and its set of critical points is the entire F-inverse image of cv(F), $\mathrm{cp}(\mathrm{F}) = \mathrm{F}^{-1}(g(\mathrm{V}))$. In particular, cp(F) includes C′, but it also includes superfluous critical points in C'^c. These we erase using 2.1.

For any $p \in \mathrm{C}'^c$, we can find a flowbox neighborhood $\mathrm{N}_p = \mathrm{N}_0 \times [a, b]$ of p respecting the grad$(f \circ \mathrm{H})$-flow; on N_p, $\mathrm{F}(x^1, \ldots, x^m) = g(x^m)$. We always choose b

to be a regular value of F, i.e., $b \in \mathbf{R}\backslash g(\mathrm{V})$. Then $\mathrm{F}(x)$ is strictly increasing respecting x^m and $\dfrac{\partial \mathrm{F}(x)}{\partial x^m} > 0$ at $x^m = b$. The critical points of F in N_p are superfluous. As in the proof of 2.2 we form a locally finite cover of $\mathrm{C}'^c \cap \mathrm{cp}(\mathrm{F})$ by such neighborhoods N_p, and proceed to erase the superfluous critical points via 2.1. The result is a smooth proper function f' with $\mathrm{cp}(f') = \mathrm{C}'$.

QED

Even so, we have not made full use of the hypotheses of 2.5. We needed to know that H diffeomorphs C'^c to C^c and that $\mathrm{cp}(f) = \mathrm{C}$, but we did not use the fact that H homeomorphs C′ onto C. It suffices that H is globally continuous and sends C′ into C. Thus.

2.5′. *Theorem. — If* C, C′ *are compact subsets of* $\mathbf{R}^m$ *for which there is a continuous endomorphism of* $\mathbf{R}^m$ *sending* C′ *into* C, *diffeomorphing* C'^c *onto* C^c, *and if* C *is properly critical, then so is* C′.

Let us draw some conclusions from 2.5, 2.5′.

2.6. *Corollary. — Any cellular set* $\mathrm{C} \subset \mathbf{R}^m$ *is properly critical,* $m \neq 4$.

Proof. — See also Norton and Pugh. Recall that a set $\mathrm{C} \subset \mathbf{R}^m$ is said to be *cellular*, a term invented by Morton Brown, if it is the monotone intersection of compact topological m-balls, $\mathrm{C} = \bigcap_n \mathrm{B}_n$. Any cellular set $\mathrm{C} \subset \mathbf{R}^m$ is a compact non-empty connected set—known as a *continuum*—and if $m \neq 4$ then $\mathrm{S}^m\backslash\mathrm{C}$ is diffeomorphic to $\mathbf{R}^m$, ∞ being sent to 0. See Brown (1962). (If $m = 4$ one must assume that C is the monotone intersection of *smooth* 4-balls. See Norton and Pugh.) The diffeomorphism extends to a continuous map $\mathrm{S}^m \circlearrowleft$ sending C to ∞. Reflecting in the equator gives a map $\varphi : \mathbf{R}^m \circlearrowleft$ which diffeomorphs C^c to $\mathbf{R}^m\backslash\{0\}$ and sends C to 0. According to 2.5′, since $\{0\}$ is properly critical, so is C.

QED

2.7. *Corollary. — Any finite disjoint union of cellular sets in* $\mathbf{R}^3$ *is properly critical.*

Proof. — This is easy and left as an excercise. So is its generalization to a " tame union of cellular sets ".

QED

2.8. *Theorem. — If a properly critical set in* $\mathbf{R}^m$ *is a compact non-empty connected set then it is cellular.*

Proof. — The proof is easy and appears in Norton and Pugh.

QED

Combining 2.6 and 2.8 we get

2.9. *Theorem. — A compact non-empty connected set in* $\mathbf{R}^3$ *is properly critical if and only if it is cellular.*

2.10. *Corollary.* — *No circle in* $\mathbf{R}^3$ *is properly critical.*

Proof. — A circle—unknotted, knotted, wild, or whatever—is compact and connected but is not cellular. Its complement is never simply connected and is therefore not diffeomorphic to $\mathbf{R}^3\backslash\{0\}$.

QED

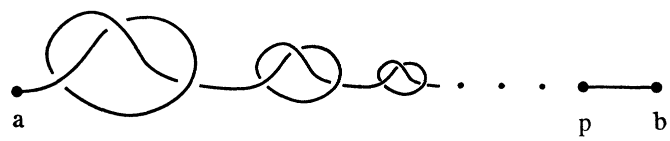
2.11. *Corollary.* — *The Alexander horned ball is properly critical if and only if it horns are internal.*

Proof. — The horned ball is a continuum that is cellular if and only if the horns curl inward.

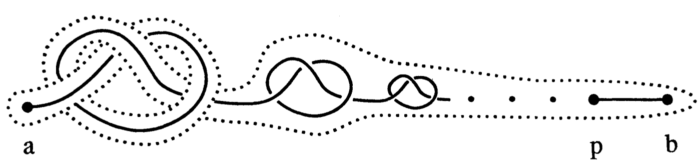
QED

2.12. *Corollary.* — *Some wild arcs in* $\mathbf{R}^3$ *are properly critical and others are not.*

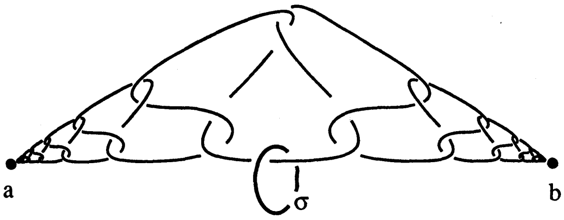
Proof. — An arc $A \subset \mathbf{R}^3$ is *tame* if there is an ambient homeomorphism of $\mathbf{R}^3$ carrying A onto a straight segment. If it is not tame it is *wild.* One example of a wild arc is gotten by tying a sequence of smaller and smaller disjoint overhand knots that limit down to an interior point p of the arc A from a to b. Between p and b, A is a segment.



According to Wilder (1930), A is wild. (Interestingly, without the segment $[p, b]$, A is tame.) It is not hard to prove that A is cellular. For we can enclose it in a decreasing sequence of spheres as shown. (In particular, this proves that $\mathbf{R}^3\backslash A$ is simply connected. Note too that making such disjoint knots accumulate at other points interior to A still leaves A cellular, even if the knots accumulate at a Cantor subset of A.) Being cellular, A is properly critical according to 2.6.



To exhibit a wild arc that is not properly critical, we consider B as shown. It is taken from Fox and Artin (1948), p. 981. The clasps converge to the endpoints a, b of B. (This time it is uneccessary to add segments beyond the knot accumulation points. The complement of B is not simply connected as can be seen by trying to unlink the loop σ shown. Therefore B is not cellular and by 2.8 it can not be properly critical.



QED

Remark. — There is probably a similar result for other non-trivial properly critical sets—they can be re-embedded so that in their new incarnation they are no longer properly critical.

2.13. *Corollary.* — *The p-adic solenoid and the Denjoy solenoid are not properly critical.*

Proof. — The p-adic solenoid is the nested intersection of longer and longer, thinner and thinner solid tori T_i that wrap more and more times around the core of a fixed solid torus T. See Shub (1987), p. 27. The Denjoy solenoid is the suspension over a circle of a Cantor subset of the circle, see Nemytskii and Stepanov (1960), p. 381-383 and p. 391-392. Both are continua but neither has simply connected complement.

QED

Remark. — Simply connectedness of the complement of C is not enough to conclude that it is cellular. For example, the components of the complement of the 2-sphere are simply connected. A more enlightening example is the Whitehead continuum C. See Rolfsen, p. 82. Its complement W in S^3 is a contractible open set, so $\mathbf{R}^3 \setminus C$ is extremely simply connected, but W is not homeomorphic to $\mathbf{R}^3$, and so C is not cellular. We have just indicated a proof of

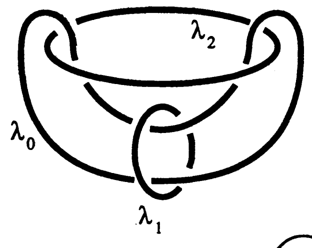
2.14. *Corollary.* — *The Whitehead continuum is not properly critical.*

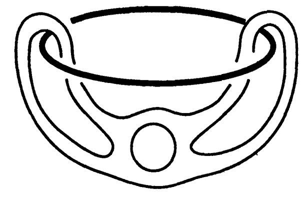
Next we turn to links with more than two components.

2.15. *Theorem.* — *Any tame link in* $\mathbf{R}^3$ *with at least three components is properly critical.*

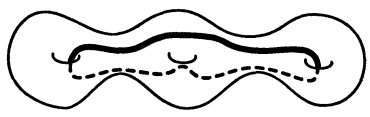
Example. — Consider the Borromean rings $L = \lambda_0 \cup \lambda_1 \cup \lambda_2$. We construct a smooth proper function f on $\mathbf{R}^3$ such that $\mathrm{cp}(f) = L$ and $f(\lambda_i) = i$, $i = 0, 1, 2$. The level surfaces of f are shown in the figures on next page.

The level surface $f = 0.1$ is a torus enclosing λ_0. At level $f = 1$, the torus has grown larger and has hit λ_1 in two points. The level surface $f = 1.1$ has genus 3 and encloses $\lambda_0 \cup \lambda_1 =$ the set of critical points in $f \leqslant 1$. The last component λ_2 snakes through two holes of this level surface. As $t \uparrow 2$, the level surface $f = t$ grows so that the two holes through which λ_2 passes shrink to critical points p, q in the bagel fashion of 2.3, while the empty hole shrinks to a critical point r in the bialy fashion of 2.4. We arrange λ_2 (by an isotopy) so it lies on the pinched level surface $f = 2$ and contains the three critical points p, q, r. The level surfaces $f = t > 2$ are spheres. By Theorem 2.2 we enlarge the critical set on $f = 2$ to λ_2, introducing no new critical points. This gives a function with the Borromean rings as critical set.

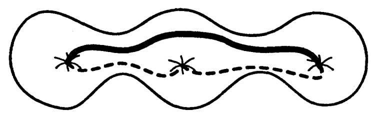




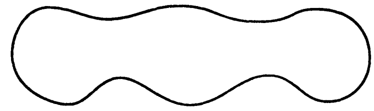
$f = 1.1$



$f = 1.5$



$f = 2.0$



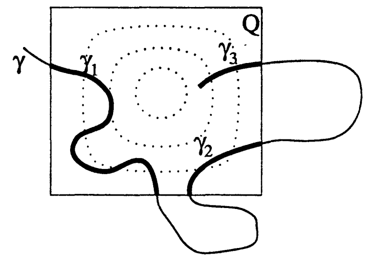
$f = 2.1$

The general case shares the same last step. As $t \uparrow 2$, the largest critical value, the level surface $f = t$ is an unknotted handlebody whose holes are shrinking to points. The shrinking takes place in disjoint compressing discs ([1]) into which $f = t$ grows. It is vital that the last component λ_n of the link meet each compressing disc once, either transversally as in the bagel pinch or tangentially as in the bialy pinch. The trick is to arrange the components of the link in space so that the connections at the critical level surfaces are easily determined. It will be convenient to first work piecewise linearly and then smooth things off.

The integer lattice $\mathbf{Z}^3$ determines the division of $\mathbf{R}^3$ into integer cubes Q with integer faces, integer edges, and integer vertices—the last being the points of $\mathbf{Z}^3$ themselves. On each Q choose a smooth function $d = d_Q$ whose level surfaces form a smoothed-off square bullseye. That is, $d = 0$ at the center of Q, $d = 1$ on ∂Q, and the only critical points of d occur at the center of Q and on its edges. We have in mind a smoothed-off version of the max-norm distance ([2]) to the center of Q. A *strand* (in Q) of a smooth curve $\gamma \subset \mathbf{R}^3$ is a compact arc that is a connected component of $\gamma \cap Q$. A strand is a *d-radial* if it is topologically transverse to the d-level surfaces, while it is a *d-hook* if it is non-transverse at exactly one point p, a d-minimum on the strand, and p is interior to both Q and the strand. We also require d-radials and d-hooks to miss the edges of Q. A curve γ is *d-hooked* if for each Q, $\gamma \cap Q$ is either empty or a d-hook. The next three lemmas show that d-hookedness is a mild and flexible condition.

([1]) A compressing disc is actually a solid cylinder $D^2 \times [0, 1]$ glued to the boundary of a handlebody $H \subset \mathbf{R}^3$ along the belt $S^1 \times [0, 1]$. Except for the belt, it is exterior to H. A handlebody is unknotted in $\mathbf{R}^3$ if and only if it becomes a ball under the addition of a finite number of compressing discs.

([2]) The max-norm in $\mathbf{R}^3$ is $|(x, y, z)| = \max\{|x|, |y|, |z|\}$. Its spheres are not smooth.



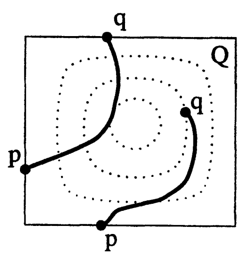
$\gamma_1 =$ a strand but not a d-hook

$\gamma_2 =$ a d-hook

$\gamma_3 =$ a d-radial.

2.16. *Lemma.* — *Given $p, q \in Q$, not on its edges, there is a d-hook λ from p to q in Q. If $d(p) \neq d(q)$ then λ can be chosen to be a d-radial.*

Proof. — Draw it.



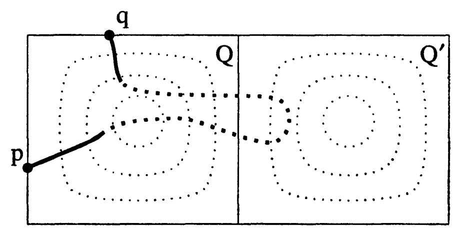
QED

2.17. *Lemma.* — *Let Q, Q' be abutting integer cubes with bullseye functions d, d'. If λ is a d-hook in Q from p to q and Λ is a 1-complex in $Q \cup Q'$ disjoint from λ then there is an isotopic arc λ' in $Q \cup Q'$ from p to q such that*

(2) *the isotopy takes place in $(Q \cup Q')\backslash\Lambda$ and leaves λ fixed near p, q;*

(3) *the connected components of $\lambda' \cap Q$ are d-hooks and those of $\lambda' \cap Q'$ are d'-hooks.*

Proof. — Pluck λ across $Q \cap Q'$ as shown. Since Λ has codimension two it can't obstruct the isotopy.



QED

2.18. *Lemma.* — *A d-hook joining $p, q \in \partial Q$ is unknotted.*

Proof. — A d-hook is ambiently diffeomorphic rel ∂Q to the union of two straight segments.

QED

Proof of **2.15.** — Let $L = \lambda_0 \cup \ldots \cup \lambda_n$ be the given link. By a PL isotopy we may assume that each loop λ_i consists of integer edges. As above we refer to integer cubes as Q. By subdividing $\mathbf{Z}^3$, or equivalently by dilating L, it is also fair to assume:

(4) If λ_i meets Q then it does so in a strand that is either an integer edge or a pair of integer edges forming a right angle. At most one λ_i meets any Q.

After making a small parallel translation of $\lambda_0, \ldots, \lambda_{n-1}$ it follows from (4) that, if $\lambda_i \cap Q \neq \emptyset$, then $\lambda_i \cap Q$ is a strand that is either an axis parallel straight line segment or is a right angle pair of axis parallel segments. In both cases, the strands are interior to Q except at their endpoints and these endpoints do not lie on integer edges. By Lemma 2.16, each such strand can be replaced with an isotopic d-hook, the isotopy fixing a neighborhood of ∂Q. Since proper criticality is invariant under isotopy, we may assume that $\lambda_0, \ldots, \lambda_{n-1}$ are already d-hooked and miss the centers of the integer cubes Q.

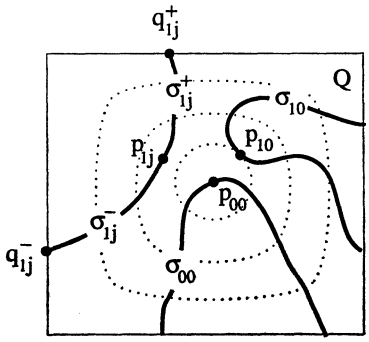
Choose a large cube C with integer vertices containing L in its interior. Consider abutting integer cubes Q, Q' in C, and a strand λ of λ_0 in Q. Let $\Lambda = L \backslash \lambda$. By Lemma 2.17, we replace λ with λ' where λ' is isotopic to λ in $Q \cup Q'$ and obeys (2), (3). Applying this construction repeatedly to a string of abutting integer cubes lets us replace λ_0 with a d-hooked loop λ'_0 in C that meets each C-interior face of every integer cube in C. (Repetition in this string of abutting integer cubes is permitted.) All this is done without disturbing $\lambda_1, \ldots, \lambda_n$. We then do the same thing with λ_1, getting an isotopic d-hooked loop λ'_1 that meets each C-interior face of Q. Again, since proper criticality is invariant under isotopy, it is fair to assume that L already has these properties and, for all integer cubes $Q \subset C$,

(5) $L = \lambda_0 \cup \ldots \cup \lambda_n \subset \operatorname{int}(C)$, $\lambda_0, \ldots, \lambda_{n-1}$ are d-hooked, λ_0 and λ_1 meet each C-interior face of Q, and λ_n consists of integer edges.

Consider the λ_i-strands σ_{ij} in Q where $0 \leqslant i \leqslant n-1$ and $0 \leqslant j \leqslant J(i)$. Each d-hook σ_{ij} splits into two d-radials $\sigma_{ij}^{\pm}$ from the d-minimum point p_{ij} to the points $q_{ij}^{\pm}$ where σ_{ij} meets ∂Q. By a further isotopy in Q we may also assume that

(6) $$0 = d(p_{00}) < d(p_{10}) < d(p_{ij})$$

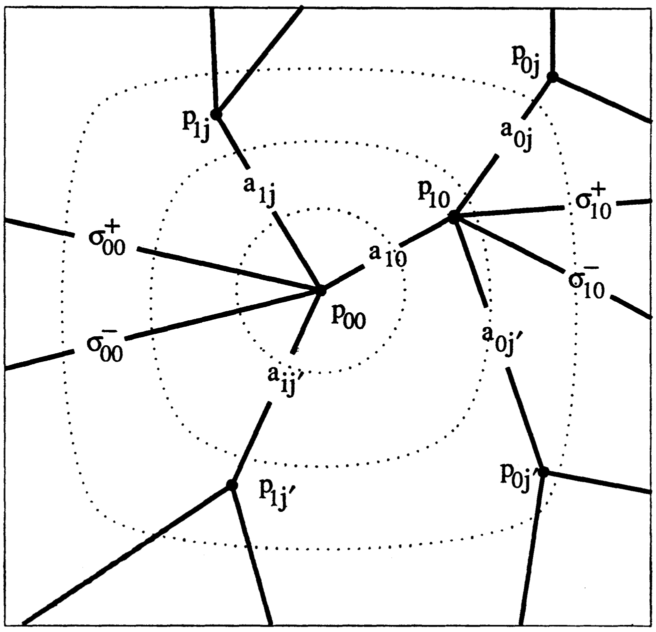
for all other i, j. By (5), p_{00}, p_{10} exist. Let a_{ij} be a d-radial arc from p_{ij} to p_{00} or p_{10} according as $i \geqslant 1$ or $i = 0$. (If $i = j = 0$, let $a_{ij} = p_{00}$.)



By 2.16, these a_{ij} exist and we choose them disjoint from each other and from L, except at their endpoints. The set

$$T = \bigcup a_{ij} \cup \sigma_{ij}$$

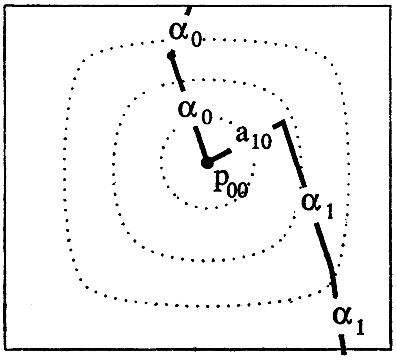
is a graph with nodes p_{ij}, q_{ij}^-, q_{ij}^+ and edges a_{ij}, σ_{ij}^-, σ_{ij}^+. From the following diagram, it is clear that T is *tree*—i.e., it is connected and contains no cycles.



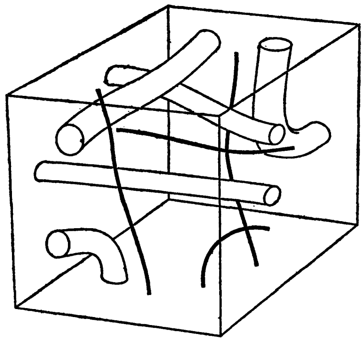
Let α be an arc in T whose endpoints lie in ∂Q. We claim that α is a d-hook. If α contains both nodes p_{00}, p_{10}, then

$$\alpha = \alpha_0 \cup a_{10} \cup \alpha_1$$

where α_0, α_1 are d-radials. (Here we use (6), $d(p_{10}) < d(p_{ij})$ for all $ij \neq 00, 10$.) On the other hand, if α contains only one node or neither node then it does not contain a_{10} and is even more clearly a d-hook. By Lemma 2.18, T is unknotted relative to ∂Q.

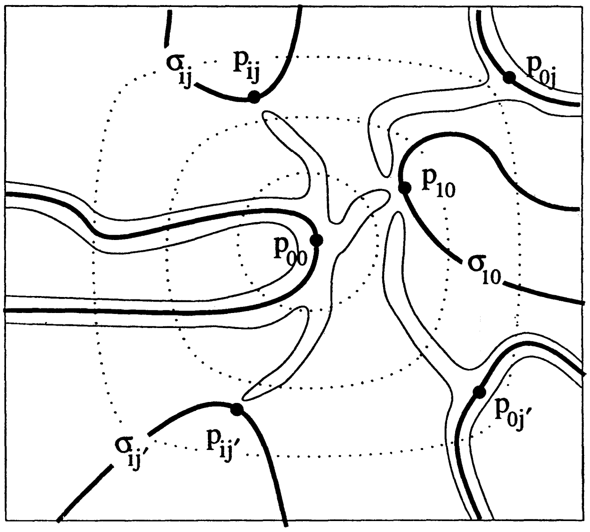


Now we define a function $f: \mathbf{R}^3 \to \mathbf{R}$ so that it takes a minimum on λ_0, say $f(\lambda_0) = 0$. The nearby level surface $f = 0.1$ is a thin 2-torus around λ_0. It meets every $Q \subset C$ in a finite set of tubes cutting ∂Q transversally, one of which contains the center of Q. A typical cube is shown.



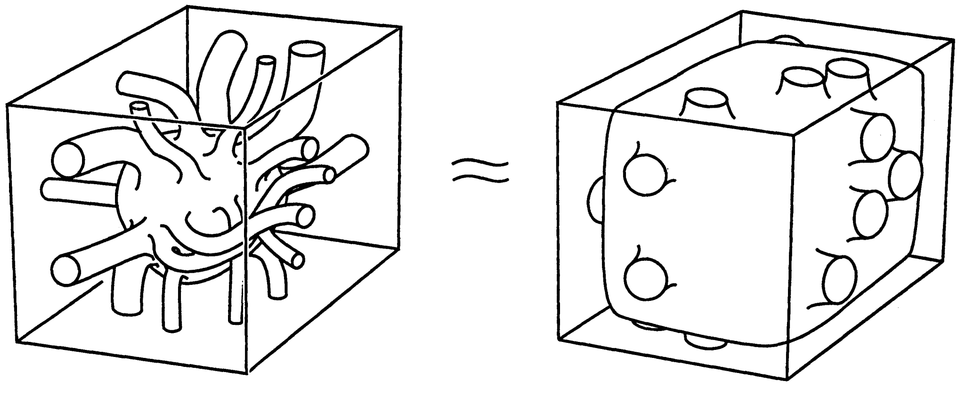
As $t \uparrow 1$ the level surface $f = t$ exudes thin, outward d-radial feelers from p_{00} along the edges a_{ij} of T to the points p_{ij}, $i \geqslant 1$. Simultaneously, a single feeler extends along each a_{0j} from p_{0j}, $j \geqslant 1$, to p_{10}. When $t = 1$, the feelers touch their intended target points and we define $f^{-1}(1)$ to be the strands σ_{ij}, $1 \leqslant i \leqslant n-1$, together with the limits of the feelers. (Thus, $f(\lambda_i) = 1$, $1 \leqslant i \leqslant n-1$, although it would be equally easy to make the values distinct.) We define $f^{-1}(1.1)$ to be a surface closely approximating these limit feelers and strands. Since T is an unknotted tree in Q, $f^{-1}(1.1) \cap Q$ is a multiply punctured sphere; its punctures are circles on the face of Q corresponding to $q_{ij}^{\pm}$, $0 \leqslant i \leqslant n-1$, and its embedding is unknotted. By (5), such puncture circles occur on each C-interior face of Q.

After the touch, the level surface $f^{-1}(1.1)$ encloses $\lambda_0 \cup \ldots \cup \lambda_{n-1}$ and in Q it is isotopic, rel ∂Q, to a multiply punctured sphere. The isotopy is realized by expanding through d-level surfaces. As $s \uparrow 1$, the circles of intersection between $f^{-1}(1.1)$ and the level surface $d = s$ may split, but they never appear, disappear, or join. For example, the circle of intersection corresponding to the outward radial feeler to a strand σ splits

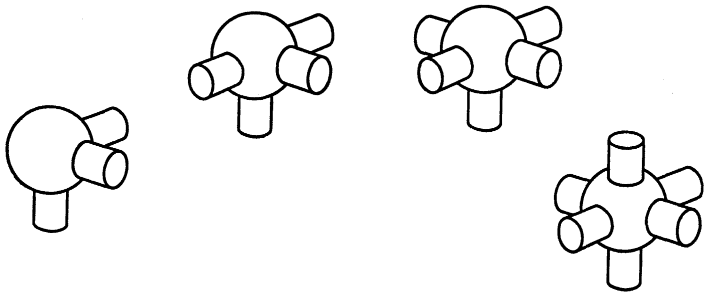


into two circles as s passes through the d-minimum value on σ. If $\sigma \neq \sigma_1$ then after this one bifurcation nothing else happens to the two circles. However, if $\sigma = \sigma_1$ then the circles re-split as s passes through the values $d(q)$ where the inward feelers touch σ_1. This picture is the basic reason the construction works. Tubes branch outward toward ∂Q, never inward.

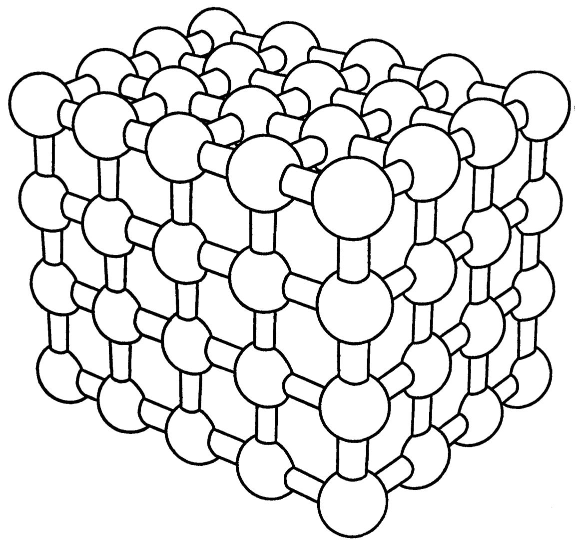
Thus, the handlebody $H = \{f \leqslant 1.1\}$ is built out of cubical units H_Q, where H_Q is (ambiently diffeomorphic rel ∂Q to) a round ball in Q joined to the C-interior faces of ∂Q by d-radial tubes.



With compressing discs Δ_k disjoint from the integer edges, we first amalgamate all the tubes that pass through a given face of Q. This forms a surface $M = f^{-1}(1.1) \cup \bigcup \Delta_k$. Since $f^{-1}(1.1)$ encloses $\lambda_0 \cup \ldots \cup \lambda_{n-1}$ and λ_n consists of integer edges, M misses L and is built from cubical units M_Q, each of which is a 3-, 4-, 5-, or 6-pronged sphere, according to the number of C-interior faces that Q has.



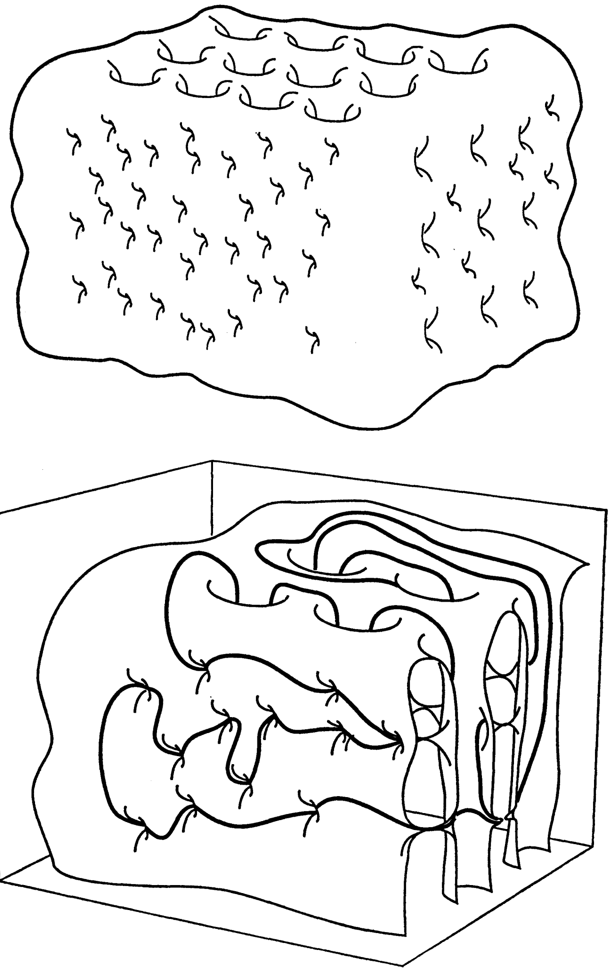
If we attach more compressing discs Δ'_k in all the vertical axis-planes and in the bottom horizontal plane, then the resulting surface $M' = M \cup \bigcup \Delta'_k$ is a sphere. The



surface M' is built like a thickened, open, upright wine carton, complete with cardboard dividers. Its fat vertices lie at the centers of the integer cubes, not on the integer lattice itself. Thus, each of the vertical and horizontal compressing discs is crossed once by an integer edge. We define f so that its level surfaces $f = t$ grow smoothly into all of the compressing discs as $t \uparrow 2$.

To recapitulate: we have shown that for the handlebody $H = \{f \leqslant 1.1\}$ there exist disjoint compressing discs $S_k \approx D^2 \times [0, 1]$ such that the 2-disc $D_k \subset S_k$ corresponding to $D^2 \times 0$ meets λ_n at most once, and $H \cup \bigcup S_k$ is a 3-ball. Finally, we arrange λ_n (by an isotopy in $\mathbf{R}^3 \backslash H$) so that it is smooth, lies on the pinched sphere $\lim_{t \to 2} \{f = t\}$, and touches every D_k exactly once, say at p_k. Thus, $f(\lambda_n) = 2$. For those D_k meeting λ_n transversally, we apply the bagel pinch at p_k, and for the others we apply the bialy pinch. The pinched sphere $f^{-1}(2)$ looks like an upholstered version of the wine carton. Its upholstery has bialy dimples on its external vertical faces and on the underside of its horizontal base face since L is interior to C. The level surfaces $f = t > 2$ are spheres.

This gives a continuous proper function on $\mathbf{R}^3$ that is smooth and non-critical off L. By Lemma 3 of Norton and Pugh there is a smooth homeomorphism $g : \mathbf{R} \to \mathbf{R}$ fixing the set $\{0, 1, 2\} = f(\mathrm{L})$ such that $\mathrm{F} = g \circ f$ is smooth. The critical set of F is $f^{-1}\{0, 1, 2\}$, but the critical points of F off L are superfluous since f is non-critical there. By Lemma 2.1 we erase these superfluous critical points and are left with a smooth proper function on $\mathbf{R}^3$ having critical set L.



$\lambda_n =$ ———. QED

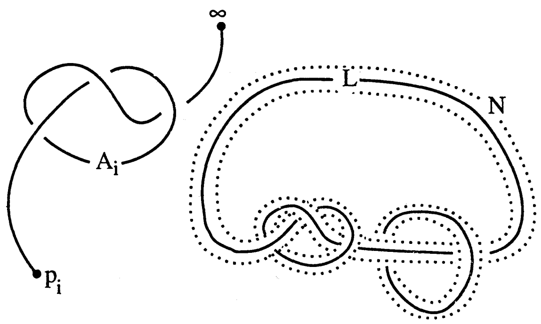
Thanks to Bill Thurston for the suggestion of a " universal jungle gym ". The same methods lead to a proof of

2.19. *Theorem. — If* L *is any tame* 1*-complex in* $\mathbf{R}^3$ *then* $\mathrm{L} \sqcup \mathrm{A} \sqcup \{p\}$ *is properly critical where* A *is a segment,* p *is a point, and* " $\sqcup$ " *denotes disjoint union.*

Finally, we come to the topic of improper criticality—the function f need not be proper.

2.20. *Theorem.* — *Any tame link in* $\mathbf{R}^3$ *is improperly critical, but no compact boundaryless surface is. Except for the 2-torus, neither is a compact surface together with a circle that it encloses.*

Proof. — Let L be a tame link in $\mathbf{R}^3$ and choose a smooth function $f_0 : \mathbf{R}^3 \to \mathrm{R}_+$ that has L as a global minimum and has no critical points in $\mathrm{N}\backslash\mathrm{L}$ where N is a tubular neighborhood of L. Modify f_0 as follows. Leave f_0 alone on N and approximate it by a function f_1 on $\mathbf{R}^3\backslash\mathrm{N}$ which has isolated critical points $p_1, p_2, \ldots$. Draw disjoint smooth arcs $\mathrm{A}_1, \mathrm{A}_2, \ldots$ in $\mathbf{R}^3\backslash\mathrm{N}$ from these critical points to ∞. Excising the arcs from $\mathbf{R}^3$



(which amounts to dragging the critical points p_i off to ∞) does not change the topology of $\mathbf{R}^3$. In fact there is a diffeomorphism $\varphi : \mathbf{R}^3 \to \mathbf{R}^3\backslash\bigcup \mathrm{A}_i$ leaving N fixed. The critical points of the pulled-back function $\varphi^*(f_1) = f_1 \circ \varphi$ are exactly L.

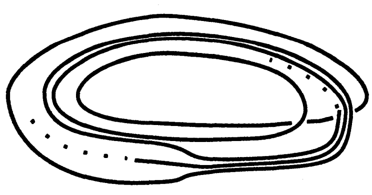
If a compact boundaryless surface Σ is improperly critical for f then f is constant on each connected component Σ_0 of Σ. Let R be the region in $\mathbf{R}^3$ bounded by Σ_0. Since f is constant on $\partial\mathrm{R}$, either $f|_{\mathrm{R}}$ has an interior maximum, an interior minimum, or it is constant. In any case we have a critical point in R contrary to the assumption that $\Sigma = \mathrm{cp}(f)$.

Suppose that $\Sigma \cup \mathrm{C}$ is improperly critical where Σ is a compact boundaryless surface and C is a circle enclosed by Σ. Then C is a local extremum and by 4.1 below, its tubular neighborhood sweeps out an open solid torus N under the gradient flow. This torus N fills out the inside of Σ and therefore Σ is the 2-torus.

QED

2.21. *Theorem.* — *The Denjoy solenoid* D *is improperly critical.*

Proof. — All we need to use is that D is a compact proper subset of the 2-torus, or for that matter, of any smooth compact connected surface Σ in $\mathbf{R}^3$. For it is easy to find a smooth function f_0 on $\mathbf{R}^3$ that is negative inside Σ, zero on Σ, positive outside Σ,



and has only finitely many critical points $p_1, \ldots, p_k$, none of which lie on Σ. By Theorem 2.2 we modify f_0 to a function f_1 whose critical set is $\mathrm{cp}(f_0) \cup \mathrm{D}$. Since D is not all of Σ and Σ is connected, there exist smooth arcs A_i from p_i to ∞ that miss D, and we use them to drag the critical points p_i off to ∞. This leaves the set D as the only critical points.

QED

Remark. — In fact it is not hard to see that any finite disjoint union of sets which are improperly critical is also improperly critical. More interesting is a case of uncountably many such sets. Consider the Denjoy solenoid D. It is improperly critical and it lies on the 2-torus T. Consider a tubular neighborhood $\mathrm{T} \times [-1, 1]$ of T in $\mathbf{R}^3$ and let C be a Cantor subset of $[-1, 1]$. Then $\mathrm{D} \times \mathrm{C}$ is improperly critical. The proof is left as an exercise.

Conjecture. — The p-adic solenoid is not improperly critical.

If $\mathrm{P} \mid \mathrm{Q}$ separates a critical set $\mathrm{cp}(f)$ (i.e., $\mathrm{cp}(f) = \mathrm{P} \sqcup \mathrm{Q}$, where P, Q are proper closed subsets of $\mathrm{cp}(f)$) then P, Q are called *critical pieces*. They are the most general type of smooth critical set, for as in 2.20, 2.21, it is easy to show that any critical piece is improperly critical. However,

2.22. *Theorem.* — *Some compact sets in* $\mathbf{R}^3$ *are neither critical pieces themselves nor homeomorphic to critical pieces.*

Proof. — Consider the Hawaiian earring

$$\left\{x \in \mathbf{R}^2 : x = 0 \text{ or } \left|x - \left(\frac{1}{n}, 0\right)\right| = \frac{1}{n} \text{ for } n \in \mathbf{N}\right\}.$$

As a subset of the plane it is not improperly critical, although as a subset of the 3-space, it is. (We think of $\mathbf{R}^2$ as the (x, y)-plane in $\mathbf{R}^3$.) If we knot the filaments of the earring, it is no longer a critical piece, but it is homeomorphic to one. The Hawaiian lightbulb,

$$\left\{x \in \mathbf{R}^3 : x = 0 \text{ or } \left|x - \left(\frac{1}{n}, 0, 0\right)\right| = \frac{1}{n} \text{ for } n \in \mathbf{N}\right\}$$

is not a critical piece nor is it homeomorphic to a critical piece.

QED

A complementary question is: which compact subsets of $\mathbf{R}^3$ can be re-embedded so they are not critical. For instance, some wild re-embeddings of the arc and ball are not critical (Corollaries 2.11, 2.12) while it is easy to see that all re-embeddings of a finite set or of the set $\{0\} \cup \left\{\frac{1}{n} : n \in \mathbf{N}\right\}$ are properly critical. More interestingly, one can generalize the proof in § 5 to show that not only is Antoine's Necklace properly critical, but so is *every* re-embedding of the Cantor set.

3. Gradients

In this section we review some general concepts from dynamical systems needed to analyze critical sets, and introduce topological criteria for homotopy near the critical set. Let $f\colon M^m \to \mathbf{R}$ be a fixed smooth proper function defined on a manifold. By properness, we may assume that the values of f lie in $\mathbf{R}_+ = [0, \infty)$. Fix a Riemann structure on M and let $x \mapsto \mathrm{grad}_x(f)$ be the resulting gradient vector field (1).

Being a smooth vector field on M, $\mathrm{grad}(f)$ generates a local flow φ. It would be convenient if φ were a (global) flow—i.e., if its trajectories $\varphi_t(x)$ were defined for all time t. However, this may not be so. For instance, the smooth proper function $x \mapsto x^4$ defined on $M = \mathbf{R}$ has gradient field $4x^3$ and its trajectory through $x = 1$ is $t \mapsto (1 - 8t)^{-1/2}$, a curve defined only on $-\infty < t < \frac{1}{8}$. The trouble is that $\mathrm{grad}(f)$ grows too fast as $f(x) \to \infty$. If we post-compose f with a diffeomorphism $g : \mathbf{R}_+ \circlearrowleft$ then we get a new function $f_1 = g \circ f\colon M \to \mathbf{R}_+$ with $\mathrm{grad}_x(f_1) = g'(fx)\,\mathrm{grad}_x(f)$. The critical points of f and f_1 are the same; the trajectories under $\mathrm{grad}(f_1)$ are merely reparameterizations of those under $\mathrm{grad}(f)$. Let us observe that g can be chosen so that $\mathrm{grad}(f_1)$ does generate a flow—that is, the trajectories under $\mathrm{grad}(f_1)$ exist for all time (2).

Let $K_n = f^{-1}[0, n]$. We know that K_n is compact, $K_n \subset \mathrm{int}(K_{n+1})$, and $M = \bigcup K_n$. By compactness, there is a constant $c_n > 0$ such that if $c : \mathbf{R}_+ \to \mathbf{R}_+$ is any function with $c(v) \leqslant c_n$ for all $v \in [n, n+1]$, then any trajectory of the field $c(fx)\,\mathrm{grad}_x(f)$ takes more than unit time to go from K_n to $K_{n+2} \backslash K_{n+1}$. We then choose a smooth function $g : \mathbf{R}_+ \circlearrowleft$ such that

(i) $g(0) = 0$ and $g' > 0$ everywhere,
(ii) $g' = 1$ on $[n, n+1]$ if $n = 2, 6, 10, 14, \ldots$,
(iii) $g' \leqslant c_n$ on $[n, n+1]$ if $n = 0, 4, 8, 12, \ldots$.

By (i), g is injective and g^{-1} is smooth. By (ii), g is onto. Thus g is a diffeomorphism $\mathbf{R}_+ \circlearrowleft$, and clearly, g is proper. By (iii), any $\mathrm{grad}(g \circ f)$-trajectory $\varphi_t(x)$ stays inside a compact set on any bounded time interval. Thus, maximal trajectories are defined for all time and $\mathrm{grad}(g \circ f)$ generates a flow. The upshot is:

It is no loss of generality to assume that the local flow $\varphi = \varphi_t(x)$ generated by $\mathrm{grad}(f)$ *is defined on all of* $\mathbf{R} \times M$.

Now we discuss some simple dynamics of the gradient flow φ. The *reverse orbit* through $x \in M$ and the *forward orbit* through $x \in M$ are the point sets

$$\mathcal{O}_-(x) := \{ \varphi_t(x) : t \leqslant 0 \} \quad \text{and} \quad \mathcal{O}_+(x) := \{ \varphi_t(x) : t \geqslant 0 \}$$

---

(1) We assume that M is complete respecting the metric d determined by the Riemann structure. In $\mathbf{R}^m$ we may as well use the Euclidean metric. Then $\mathrm{grad}_x(f) = (\partial f/\partial x^1, \ldots, \partial f/\partial x^m)$.

(2) In the example $f(x) = x^4$ on $\mathbf{R}$, a suitable choice of g is $g(x) = \log(1 + x)$. In general more trickiness is needed, but the idea is the same—g' should kill off $|\mathrm{grad}(f)|$.

respectively. The α-*limit set* and the ω-*limit set* are where they accumulate as $t \to -\infty$ and $t \to +\infty$.

$$\alpha(x) := \{y \in M : \varphi_{t_n}(x) \to y \text{ for some sequence } t_n \to -\infty\}$$
$$\omega(x) := \{y \in M : \varphi_{t_n}(x) \to y \text{ for some sequence } t_n \to +\infty\}.$$

Since f is proper, $\alpha(x)$ and $\omega(x)$ are compact connected subsets of $\mathrm{cp}(f)$. Thus, f is constant on each $\alpha(x)$ and $\omega(x)$. (Here we use the assumption that f is smooth; it suffices by the Morse-Sard Theorem that f be C^{m+1}.) Since $\mathcal{O}_-(x) \subset f^{-1}[0, fx]$, which is a compact set, $\alpha(x) \neq \emptyset$. It is possible that $\omega(x) = \emptyset$ since $\varphi_t(x)$ may run off to ∞ as $|x| \to \infty$.

3.1. *Lemma. — If* $\omega(x) \neq \emptyset$ *then* $\mathrm{dist}(\varphi_t(x), \omega(x)) \to 0$ *as* $t \to \infty$. *The same is true for* $\alpha(x)$ *as* $t \to -\infty$.

Proof. — By $\mathrm{dist}(y, \omega(x))$ we mean $\inf\{d(y, w) : w \in \omega(x)\}$, d being the Riemann metric on M. Suppose the assertion is false. Then there is a sequence $\tau_n \to \infty$ such that $\mathrm{dist}(\varphi_{\tau_n}(x), \omega(x)) > \nu > 0$ for all n. Since $\omega(x) \neq \emptyset$, there is a sequence $t_n \to \infty$ with $\mathrm{dist}(\varphi_{t_n}(x), \omega(x)) \to 0$. By the intermediate value theorem, we find a sequence $T_n \to \infty$ such that

$$\mathrm{dist}(\varphi_{T_n}(x), \omega(x)) = \nu.$$

By local compactness of M we may assume that $\varphi_{T_n}(x)$ converges to some point $y \in M$. Clearly, $y \in \omega(x)$ and $\mathrm{dist}(y, \omega(x)) = \nu$, a contradiction. The case of α-limit sets is checked similarly.

QED

Consider two level sets U, V of f such that $f(U) < f(V)$. Some points of U may flow to V under the gradient flow. Let

$$U_0 = \{u \in U : \varphi_t(u) \in V \text{ for some } t = t(u) > 0\}.$$

Since $f(\varphi_t(u))$ is either constant or strictly monotone increasing in t, the time $t(u)$ is unique, and

$$h : U_0 \to V \qquad u \mapsto \varphi_{t(u)}(u)$$

is well-defined. It is called the *Poincaré map* from U to V. Its image in V is the set $V_0 = \{v \in V : \varphi_s(v) \in U \text{ for some } s = s(v) < 0\}$. Clearly, $s(v) = -t(u)$ for $v = h(u)$, and $\varphi_{s(v)}(v) = h^{-1}(v)$. It is standard by transversality and the implicit function theorem that

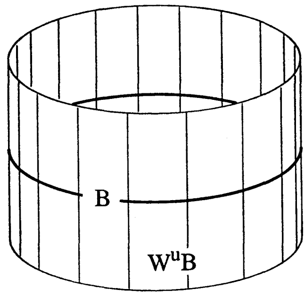
The Poincaré map is a diffeomorphism from U_0 *to* V_0 *and* U_0, V_0 *are open subsets of* U, V.

A set $P \subset M$ is φ-*invariant* if $\varphi_t(P) = P$ for all $t \in \mathbf{R}$. For example any set of critical points is φ-invariant. The *unstable set* of a φ-invariant set P and the *stable set* of P are

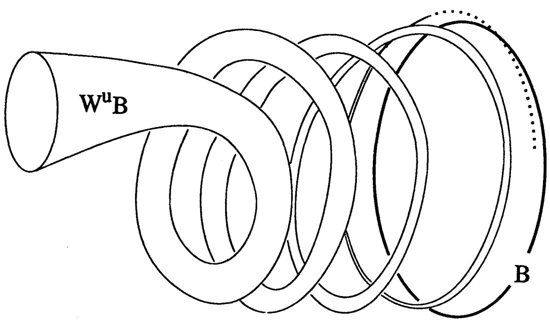
$$W^u(P) := \{x \in M : \mathrm{dist}(\varphi_t(x), P) \to 0 \text{ as } t \to -\infty\}$$
$$W^s(P) := \{x \in M : \mathrm{dist}(\varphi_t(x), P) \to 0 \text{ as } t \to +\infty\}$$

respectively. By Lemma 3.1, $W^u(P) = \{ x : \alpha(x) \subset P \}$ and $W^s(P) = \{ x : \emptyset \neq \omega(x) \subset P \}$. We are going to describe how the unstable set attaches itself to its invariant set.

We assume that $f\colon M \to \mathbf{R}$ is smooth, proper, and has no critical values in the interval $(b, c]$. We set $V = f^{-1}(b)$, $B = \mathrm{cp}(f) \cap V$, and $S = f^{-1}(c)$. There are two ways that $W^u(B)$ attaches itself to B. The first is exemplified by B being a circle and $W^u B$ being an annulus (or several annuli) whose boundary is B. They are glued to B in the



same way that the unstable manifold is glued to a hyperbolic periodic orbit. The second way is like the Fuller (1952) whisker where B is a circle and $W^u B$ is a funnel coiling away from B in its exterior. Since all points of B are fixed under the gradient flow, the coiling is extremely slow. We could also imagine the funnel not coiling at all or coiling first one way and then the other. It is this second type of attachment that is more troublesome to analyze.



To be more precise we say that $W^u B$ *Fuller attaches* to B if for every neighborhood N of B in M, we can find a small neighborhood S_0 of $W^u B \cap S$ in S and a time $t_0 < 0$ such that $\varphi_{t_0}(S_0) \subset N$ and

(i) For every Jordan curve $\gamma \subset S_0$, $\varphi_{t_0}(\gamma)$ is null homotopic in N.

Since $\pi_1(S_0)$ is generated by loops that are Jordan curves, we can rephrase (i) as

(ii) $(\varphi_{t_0})_* : \pi_1(S_0) \to \pi_1(N)$ has null image.

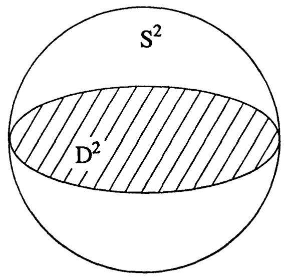
That is, $\varphi_{t_0}(S_0)$ is homotopically trivial in N. It is possible to avoid mention of S_0 altogether and rephrase (i) in terms of Čech homotopy (see Christie (1944) for its definition) by requiring that for all $t \ll 0$,

(iii) $$(\varphi_t)_* : \check{\pi}_1(W^u B \cap S) \to \check{\pi}_1(N) \text{ has null image.}$$

We call the opposite of Fuller attachment *essential* attachment. If B is a Jordan curve, we can take N to be a solid torus with core B. Then $W^u B$ attaches essentially to B if and only if there exist essential loops $\varphi_{t_0}(\gamma)$ in N, N being a thin solid torus at B, t_0 being very negative, and γ being a Jordan curve in S near $W^u B$. (Recall that an essential loop in N winds along the core of N with non-zero total winding—its total signed intersection with a meridian disc is non-zero.)

3.2. *Lemma.* — *Suppose that* $M = \mathbf{R}^3$, $f : M \to \mathbf{R}$ *is smooth, has no critical values in* $[a, b) \cup (b, c]$, $f^{-1}(a)$ *is connected and non-empty,* $B = f^{-1}(b) \cap \mathrm{cp}(f)$, $f^{-1}(c) = S \approx S^2$, *and* $W^u B$ *Fuller attaches to* B. *Then* $S \backslash W^u B$ *is connected.*

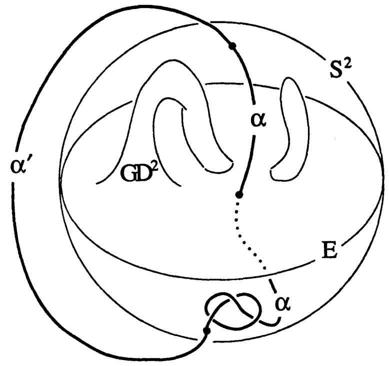
It is convenient to first prove a separation lemma concerning the space $\Theta = S^2 \cup D^2$ formed by rotating the letter " theta " around its vertical bisector. Here, S^2 is the unit sphere in $\mathbf{R}^3$ and D^2 is the open unit disc in the (x, y)-plane. They are disjoint.



The space Θ

3.3. *Lemma.* — *Suppose that* $g : \Theta \to S^3$ *is continuous,* g *smoothly embeds* S^2, *and* $g(S^2)$ *misses* $g(D^2)$. *Then* $S^3 \backslash g(\Theta)$ *includes at least three components* R_0, R_1, R_2 *where* R_0 *is one of the two components of* $S^3 \backslash g(S^2)$ *and* R_1, R_2 *abut the g-images of the Northern and Southern hemispheres.*

Proof. — Since g smoothly embeds S^2, we can follow g by an ambient diffeomorphism $g' : S^3 \circlearrowleft$ such that $G = g' \circ g$ is the identity on S^2. Since $g(S^2)$ misses $g(D^2)$, $G(D^2)$ misses S^2 and lies entirely inside it or entirely outside it; say it lies in the inside of S^2, D^3. We claim that no arc α can be drawn in D^3 from the Northern hemisphere to the Southern hemisphere which misses $G(\Theta)$. For such an arc α could be continued



in S^3 to form a loop α' which links the equator E of S^2, while $G(\bar{D}^2)$ would provide a null homotopy of E in $D^3\backslash\alpha$. Since α cannot exist, $D^3\backslash G(\Theta)$ includes distinct components that abut the Northern and Southern hemispheres of S^2. They are carried by the ambient diffeomorphism g' to the regions R_1, R_2, while the outside of S^2 is carried to R_0.

QED

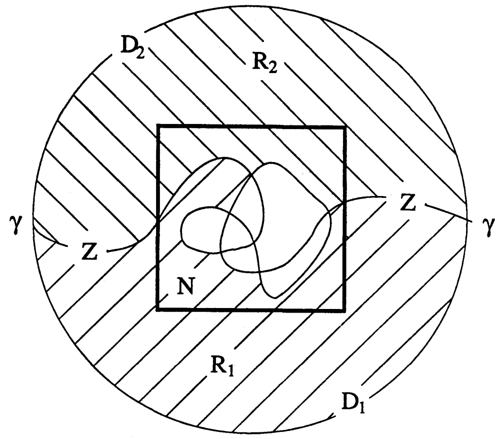
Remarks. — Possible knottedness of α prevents us from using α to retract $g(\bar{D}^2)$ to E. A similar result of Feign (1988) states that any smoothly immersed compact surface in $\mathbf{R}^3$ separates. Examples like Bing's house with two rooms show that such separation results are not entirely trivial. Smoothness of g in 3.3 can be relaxed to continuity. The analysis must be done without the ambient diffeomorphism g'.

Proof of **3.2.** — Given that $W^u B$ Fuller attaches to B, $f^{-1}(a)$ is connected and non-empty, and $S = f^{-1}(c) \approx S^2$, we want to show that $S\backslash W^u B$ is connected. Suppose it is disconnected and S_1, S_2 are two of its components. Choose $s_1 \in S_1$, $s_2 \in S_2$. Let N be any neighborhood of B in $\mathbf{R}^3$ and let S_0, t_0 satisfy (i). Since $S \approx S^2$ and S_0 is a neighborhood of $\partial S_1 \cup \partial S_2$, we can choose a smooth Jordan curve $\gamma \subset S_0$ separating s_1 from s_2. The gradient points outward across S and so $\{\varphi_t \gamma : t_0 \leqslant t \leqslant 0\}$ is an embedded cylinder which we denote by Z. By (i) its inner boundary $\varphi_{t_0}(\gamma)$ is null homotopic in N. Let D_1, D_2 be the open discs whose union is $S\backslash\gamma$, labelled so that $s_1 \in D_1$, $s_2 \in D_2$.

Define a continuous map $g : \Theta \to \mathbf{R}^3$ that smoothly embeds the Southern hemisphere of Θ onto D_1, the Northern hemisphere onto D_2, the cylinder

$$\{(x, y, 0) \in D^2 : x^2 + y^2 \geqslant 1/2\}$$

onto the cylinder Z, and on the disc $D' \subset D^2$ of radius 1/2 define g to express the null homotopy of $\varphi_{t_0}(\gamma)$ in N. As above, $\Theta = S^2 \cup D^2$. Then g is continuous and on S^2 it is a smooth embedding. Besides, $g(S^2)$ misses $g(D^2)$. From 3.3 it follows that $g(\Theta)$ separates the inside of S and two of the complementary components, R_1, R_2, abut D_1, D_2.



Since $f^{-1}(a)$ is connected, it misses R_1 or R_2. Say it misses R_1. Consider the reverse orbit of s_1, $\varphi_t(s_1)$ for $t \leqslant 0$. It enters R_1 and never crosses D_1 or the cylinder Z by uniqueness of flow trajectories. Thus, it either meets $g(D')$ or it stays forever in R_1. In the latter case, $f(\varphi_t(s_1)) \neq c$ for all $t < 0$ and we see that $f(\varphi_t(s_1)) > b$ for all $t < 0$. Then $\alpha(s_1) \subset B$ and $s_1 \in W^u B$. In the former case, $\varphi_t(s_1)$ passes through N. But N is an arbitrarily small neighborhood of B and s_1 is fixed, independent of N. Thus, in either case, $s_1 \in W^u B$. This contradicts the assumption that S_1 is complementary to $W^u B$ in S.

QED

Lemma 3.2. Can be summarized as saying that, under the right circumstances, Fuller attachment implies connectedness of $S \backslash W^u B$. The next lemma says that connectedness of $S \backslash W^u B$ implies a key property of loops at B—they can be " slipped below " V.

3.4. *Slip Lemma.* — *Suppose that* $f: M \to \mathbf{R}$ *is smooth, has no critical values in* $(b, c]$, $B = f^{-1}(b) \cap \mathrm{cp}(f)$, $f^{-1}(c) = S \approx S^2$, *and* $S \backslash W^u B$ *is connected, non-empty. Then any loop* $\sigma_0 \subset B^c$ *near* B *is locally homotopic in* B^c *to a loop* σ_1 *lying* below V *in the sense that* $f(\sigma_1) < b$.

3.5. *Lemma.* — *If* $W \subset S^2$ *is compact and does not separate* S^2 *then any path* γ *whose endpoints* p, q *lie in* W^c *is homotopic (rel.* p, q*) to a path in* W^c. *The homotopy occurs near* W.

Proof. — Let $g: S^2 \to [0, 1]$ be a smooth function with $g^{-1}(0) = W$. For small regular values $\varepsilon > 0$, let G_ε be the connected component of $g \geqslant \varepsilon$ containing p. Then W^c is the monotone union of G_ε as $\varepsilon \to 0$, and G_ε^c is a small neighborhood of W consisting of a finite number of disjoint smooth discs D_i. Each piece of γ in D_i can be replaced by a homotopic arc on ∂D_i, the homotopy occuring in D_i. The result is a homotopic path from p to q avoiding W.

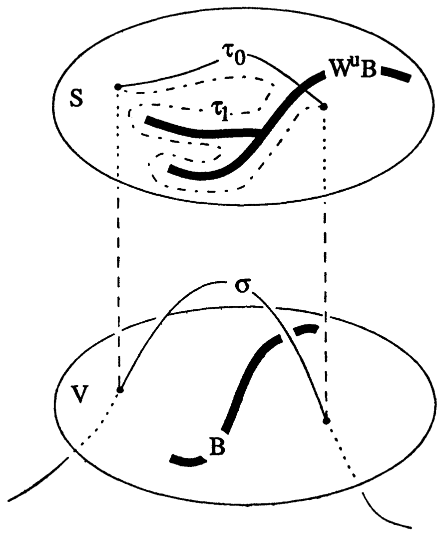
QED

Proof. of **3.4.** — Outside B, $V = f^{-1}(b)$ is a smooth surface normal to the gradient of f. Let N be a small neighborhood of B in M, and let σ_0 be a loop in $N \backslash B$. By a small homotopy, we may assume that σ_0 is a smooth Jordan curve transverse to V. Then σ_0 is the union of arcs alternately above and below V, or it is entirely above or below V.

If σ_0 is already below V there is nothing to prove. If it is entirely above V then it smoothly projects up to S along the flow trajectories, and we get a loop $\tau_0 \subset S$. Since N is small, τ_0 lies near $W^u B \cap S$. After a homotopy near $W^u B$, we may assume that $\tau_0 \not\subset W^u B$ since $S \backslash W^u B \neq \emptyset$. Because $W^u B$ does not disconnect S and $S \approx S^2$, 3.5 says that there is a homotopy τ_t in S near $W^u B$ so that τ_1 misses $W^u B$. When we flow backward by φ, τ_1 is pushed below $f^{-1}(b) = V$. That is, σ_0 is homotopic to σ_1 by a homotopy σ_t in B^c and σ_1 lies below V. However, the homotopy does not occur near B. To rectify this is easy. We just push σ_t downward by φ. More precisely, we may replace S with a regular level surface $S' = f^{-1}(b + \varepsilon)$. For the pair $(S, W^u B \cap S)$ is diffeomorphic to the pair $(S', W^u B \cap S')$. When ε is small, the neighborhood N of B is also a neighborhood of $W^u B \cap S'$, and by 3.5 we know that the homotopy τ_t occurs in such a neighborhood.

If an arc of σ_0 lies above V we treat it similarly. We homotop it to an arc consisting of two " legs " from V to S and the projected arc in S. We homotop the arc in S to miss $W^u B$, keeping the endpoints (where the legs meet S) fixed, and then we flow back down below V. To make the homotopy occur in N, we replace S by S′ as above.

QED



3A. *Another appearance of Antoine's Necklace.* In this appendix, we say something about the relation between chain recurrence and the concept of gradient-like flow invented by Charles Conley. Our remarks are due to Clark Robinson (1976) and (1990).

A flow φ is *gradient-like* if there is a globally defined continuous function g that is strictly decreasing on all non-constant orbits of φ; g is a *generalized Lyapunov function* for φ. If φ is smooth then Wilson and Yorke showed that g can be made smooth too. Generalized Lyapunov functions should not be confused with global Lyapunov functions, which are constant on periodic orbits, and, indeed, the entire chain recurrent set for the flow.

As one would expect, many properties of gradient flows hold also for gradient-like flows. However, if one wants to use the generalized Lyapunov function g to control the chain recurrence of φ then it is necessary to make an additional assumption—the critical values of g must be totally disconnected. Conley calls such a flow *strongly gradient-like*. The following example of Robinson shows what can happen without this totally disconnected condition.

Example. — There is a smooth gradient-like flow φ on $\mathbf{R}^3$ whose chain recurrent set is connected but whose fixed point set is totally disconnected. In fact, φ is generated by a vector field $\mathrm{X} = -h(x, y, z)\, \partial/\partial z$ where h is smooth, $h \geqslant 0$, and $h^{-1}(0)$ is Antoine's Necklace. The generalized Lyapunov function g for the flow is the z-height.

The point is that one would expect the level surfaces of g (the horizontal planes in Robinson's example) to prohibit non-trivial chain recurrence. This *is* the case for gradient flows. Their only chain recurrent points are their fixed points. The upshot is that the fixed point set of the gradient-like flow φ is strictly smaller than its chain recurrent set.

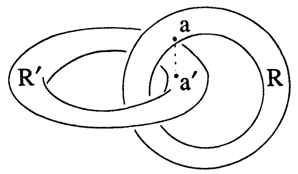
The relevance of this example to the present paper is that at first one might think it gives a way to realize Antoine's Necklace as the critical set of some smooth function f by solving the equation $\mathrm{X} = \mathrm{grad}(f)$. However, the requirements that $\partial f/\partial x = \partial f/\partial y = 0$ and $\partial f/\partial z = h$, everywhere, imply that h is constant. (Differentiate and commute the mixed partials.)

Here is a brief description of Robinson's example. It relies on the following fact about the self-linking of Antoine's Necklace A.

(*) Given two linked rings (solid tori) R, R′ in the construction of A, there is no embedded 2-sphere in $\mathbf{R}^3$ that separates $\mathrm{A} \cap \mathrm{R}$ from $\mathrm{A} \cap \mathrm{R}'$.

From (*) it follows that for such linked rings R, R′ there exist points $a \in \mathrm{A} \cap \mathrm{R}$, $a' \in \mathrm{A} \cap \mathrm{R}'$ with $a > a'$ in the sense that

$$a = (x, y, z) \qquad a' = (x, y, z') \quad \text{and} \quad z > z'.$$



For if no such a, a' exist then $A \cap R < A \cap R'$ and we can vertically translate $A \cap R'$ far above R, missing $A \cap R$ during the translation. Then $A \cap R$ and $A \cap R'$ can be separated by a 2-sphere, contradicting $(*)$.

Let R be the collection of all the rings R used in the construction of A. Define φ to be the flow generated by $X = -h\partial/\partial z$ where $h: \mathbf{R}^3 \to \mathbf{R}_+$ is any smooth function that vanishes only on A. We say that φ is *ε-transitive* on $A \cap R$, for $R \in \mathsf{R}$, if for any $p, q \in A \cap R$, there is an ε-trajectory from p to q in R. Given $\varepsilon > 0$, φ is ε-transitive on all sufficiently small rings $R \in \mathsf{R}$. For h is small on small rings, and this permits ε-trajectories to move around freely.

We claim that φ is ε-transitive on all of A. It suffices to show that if φ is ε-transitive on each $A \cap R_i$, where $R_0, \ldots, R_{n-1}$ are the primary subrings of a ring $R \in \mathsf{R}$, then it is also ε-transitive on $A \cap R$. (The rings R_i need not be small.) We choose a sequence of pairs a_i, a_i' such that

$$a_i \in A \cap R_i, \qquad a_i' \in A \cap R_{i+1}, \qquad \text{and} \qquad a_i > a_i'.$$

(We understand, $R_n = R_0$.) Since $h \geqslant 0$, there is an ε-trajectory from a_i downward to a_i'. It lies in R. Since φ is ε-transitive on $A \cap R_{i+1}$, there is an ε-trajectory from a_i' to a_{i+1} in R_{i+1}. The ε-trajectory moves upstream along the rings $R_0, \ldots, R_{n-1}$ like a salmon up a weir. Thus, φ is ε-transitive on A. Since $\varepsilon > 0$ is arbitrary, one says that φ is *chain transitive* on A.

Given $\varepsilon > 0$, let $\mathsf{T}(\varepsilon)$ be the collection of all ε-trajectories T joining points of A, and let $T(\varepsilon) \subset \mathbf{R}^3$ be the union of all the $T \in \mathsf{T}(\varepsilon)$. Since φ is ε-transitive on A, $T(\varepsilon)$ is compact, contains A, and is *ε-connected* in the sense that any pair of points in $T(\varepsilon)$ can be joined by an ε-chain in $T(\varepsilon)$. Let $T = \bigcap T(\varepsilon)$. Clearly $T(\varepsilon) \downarrow T$ as $\varepsilon \downarrow 0$. Thus, T is compact, connected, and contains A. Being connected, $T \neq A$. We claim that

$$T = CR(\varphi).$$

Since φ is chain transitive on A, it is clear that $T \subset CR(\varphi)$. To prove that $CR(\varphi) \subset T$ it suffices to show that $CR(\varphi) \subset T(\varepsilon)$ for all $\varepsilon > 0$.

Clearly, $T(\varepsilon)$ contains the ε-neighborhood of A and so, for some $\nu > 0$, $h \geqslant \nu$ off $T(\varepsilon)$. It follows that no $\nu/2$-trajectory which lies entirely off $T(\varepsilon)$ can be periodic. For each unit time trajectory-segment off $T(\varepsilon)$ falls by at least ν and recovers by at most $\nu/2$. Now take $p \in CR(\varphi)$. Through p there is a periodic $\nu/2$-trajectory, P. Somewhere P meets an ε-trajectory $T \subset T(\varepsilon)$. Say T joins a to a'. Amalgamating P and T gives an ε-trajectory from a to a' that passes through p. This shows that $p \in T(\varepsilon)$, $CR(\varphi) \subset T$, and thus $CR(\varphi) = T$.

4. Further obstructions to criticality

In this section we complete the proof of Theorem B. We begin with a simple lemma about the type of neighborhood permitted at a minimum. Keep in mind that the critical points may be totally degenerate. Analytic Morse Theory is inadequate. See M. Morse (1934), Theorem 4.1 on p. 154, and M. Saito (1990), Lemma 2.

4.1. *Lemma.* — *Suppose that* $f: \mathbf{R}^3 \to \mathbf{R}_+$ *is a smooth proper function,* $f^{-1}(0)$ *is a Jordan curve* A, *and* f *has no other critical values near* 0. *Then for each small* $\varepsilon > 0$ *the set* $f^{-1}[0, \varepsilon]$ *is a solid torus of which* A *is the core circle.*

The simplest proof of 4.1 is: "how else *could* $f^{-1}[0, \varepsilon]$ look?" For an alternate proof we use some piecewise linear (PL) 3-manifold topology, especially a result from Hempel (1976) asserting that certain 3-manifolds are of the form $V \times [0, 1]$. Recall that a *fake* 3-*cell* is a compact 3-manifold with boundary which is contractible but is not homeomorphic to the standard 3-ball B^3. Equivalent to the Poincaré Conjecture is the assertion that no fake PL 3-cell exists.

4.2. *Proposition.* — *No fake* PL 3-*cell exists in* S^3.

Proof. — Suppose that $M \subset S^3$ is such a fake 3-cell. Being contractible, $H^1(M) = 0$. Also, $S^3 \backslash \partial M = \mathrm{int}(M) \sqcup (S^3 \backslash M)$ and so by Alexander Duality,

$$\check{H}^1(\partial M) = H_1(S^3 \backslash \partial M) = H_1(\mathrm{int}(M)) \oplus H_1(S^3 \backslash M) = 0 \oplus \check{H}^1(M) = 0.$$

Therefore $\partial M \cong S^2$. By the PL Schönflies Theorem in S^3 of Alexander (1928), a PL 2-sphere separates S^3 into two PL balls, so M is not fake after all.

QED

To avoid the Poincaré Conjecture one introduces the *Poincaré associate* of a 3-manifold M as follows; see Hempel, p. 88. It is a 3-manifold M^* defined up to PL equivalence by the requirements

(i) M^* contains no fake 3-cell;

(ii) $M' = M^* \# S$ where M' is M with all 2-sphere components of ∂M capped off by gluing on 3-cells, "$\#$" denotes connected sum, and S is a homotopy 3-sphere.

The idea is that M^* is M with fake 3-cells replaced by true ones. In terms of homotopy, this is no change at all.

4.3. *Corollary.* — *If* $M \subset \mathbf{R}^3$ *is a compact* 3-*manifold and no component of* ∂M *is a* 2-*sphere then the Poincaré associate of* M *is* M.

Proof. — By 4.2, M contains no fake 3-cell. By assumption $M' = M$. Since $M \# S^3 = M$ we see that M satisfies (i), (ii) and therefore by uniqueness $M = M^*$.

QED

4.4. *Proposition.* — *Any smooth h-cobordism in* $\mathbf{R}^3$ *is diffeomorphic to a product.*

Proof. — A *cobordism* between disjoint compact boundaryless surfaces V_0, V_1 is a compact 3-manifold M such that $\partial M = V_0 \cup V_1$. It is an *h-cobordism* if there are deformation retractions $r_0 : M \to V_0$ and $r_1 : M \to V_1$. We show that if $M \subset \mathbf{R}^3$ is a smooth h-cobordism between V_0 and V_1 then $M \approx V \times [0, 1]$ with V_0, V_1 corresponding to $V \times 0$, $V \times 1$.

Since r_0, r_1 induce isomorphisms on homotopy, $\pi_1(V_0) = \pi_1(V_1)$, and because they are compact boundaryless surfaces, we conclude that $V_0 \approx V_1$. Being in $\mathbf{R}^3$, V_0 and V_1 bound solid compact regions R_0, R_1 by the Jordan Separation Theorem, and either $M = R_0 \backslash \operatorname{int}(R_1)$ or $M = R_1 \backslash \operatorname{int}(R_0)$. We assume that $M = R_1 \backslash \operatorname{int}(R_0)$, so M is the compact region outside R_0 and inside R_1. Note that V_0, V_1 are orientable since they are compact boundaryless surfaces in $\mathbf{R}^3$.

Case 1. — The surfaces V_0, V_1 are 2-spheres. Then $M \approx S^2 \times [0, 1]$ by the Annulus Theorem in $\mathbf{R}^3$.

Case 2. — The surfaces V_0, V_1 are k-handled tori with $k \geqslant 1$. According to Theorem 10.2 of Hempel, p. 89, in the boundaryless case, the Poincaré associate of M is PL equivalent to $V \times [0, 1]$ (with V corresponding to $V \times 0$) provided that V is a compact connected boundaryless surface in ∂M such that

a) $V \neq S^2, \mathbf{P}^2$;

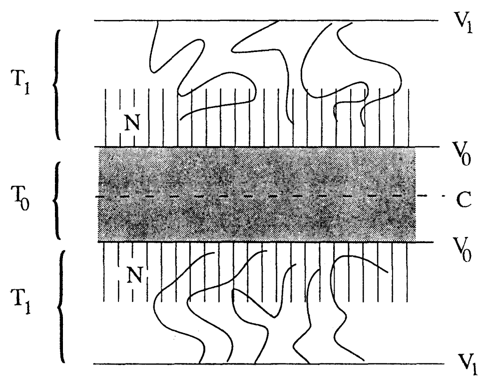
b) the inclusion $i : V \hookrightarrow M$ induces an isomorphism $i_* : \pi_1(V) \to \pi_1(M)$.

Take $V = V_0$. Then *a)* holds because we are assuming V_0, V_1 are k-handled tori; *b)* is true since M deforms to V_0. By 4.3 the Poincaré associate of M is M and thus M is PL equivalent to $V \times [0, 1]$ with V_0, V_1 corresponding to $V \times 0$, $V \times 1$. By Munkres we may replace the PL equivalence with a diffeomorphism in dimension three.

QED

Proof of **4.1**. — We make take $\varepsilon = 1$ and define $N = f^{-1}[0, 1]$, $V_1 = \partial N = f^{-1}(1)$. Post-composing f with a function on $\mathbf{R}$ which is very flat at 0, we may assume that f is flat at A. Then we blow A up to a solid torus N_0 with boundary $V_0 \approx T^2$. Naturally, f becomes a smooth function defined on the solid region N_1 between V_0 and V_1 which is critical only at V_0 and takes on the value 0 there. We extend the function to the interior of N_0 so it vanishes identically there. Call the new function $F : \mathbf{R}^3 \to \mathbf{R}_+$. It is smooth and agrees with f off the solid $N = N_0 \cup N_1$.

We claim that N_1 provides an h-cobordism between V_0 and V_1. That is, we claim that there are deformation retractions r_0, r_1 of N_1 onto V_0, V_1. Since there are no



critical values of F in $(0, 1]$, the gradient flows provides a deformation isotopy of N_1 to $N_1' \subset T$ where T is a small tubular neighborhood of V_0 in N_1. Then we use the tubular neighborhood structure of T to deform N_1' to V_0. This gives r_0; the retraction r_1 is constructed similarly. By 4.4,

$$N_1 \approx T^2 \times [0, 1] \quad \text{with } T^2 \times 0,\ T^2 \times 1$$

corresponding to V_0, V_1. Blowing N_0 back down to A shows that $N \approx S^1 \times D^2$ with A corresponding to $S^1 \times 0$.

QED

Next we prove a theorem which implies that many links $A \cup B$ are not properly critical. We first recall some terminology from Rolfsen, p. 88-89, 132-135, 297-298.

The *p-fold cyclic branched cover* of $\mathbf{R}^3$ around a tame knot $A \subset \mathbf{R}^3$ is a map $\pi : M \to \mathbf{R}^3$, such that $\pi : \pi^{-1}(A) \to A$ is a diffeomorphism, such that π is an unbranched cover of $M \backslash \pi^{-1}(A)$ to $\mathbf{R}^3 \backslash A$, and in some pair of tubular neighborhood coordinate systems at $\pi^{-1}(A)$ in M and A in $\mathbf{R}^3$, π takes the form $(x, r, \theta) \mapsto (x, r, p\theta)$. The coordinate x refers to A and its preimage, and the polar coordinates (r, θ) refer to the normal directions. (Tameness is equivalent to the existence of these coordinate systems.)

A link $A \cup B$ is *splittable* if there exist disjoint topological 3-balls containing A, B. In other words A, B are *trivially linked*, they are not truly linked at all. The *linking number* $Lk(A, B)$ is the total number of times that A, B wind around each other, taking into account cancellation due to winding backwards and forwards.

The linking number of the Hopf link is 1 and that of the Whitehead link is 0. If a link is splittable then clearly its linking number is zero, but the Whitehead link shows the linking number can be zero for unsplittable links. Splittability is invariant under isotopy while linking number is invariant under homotopy. During the homotopy, A_t and B_t stay disjoint, although each is allowed to self intersect. Thus, the Whitehead link is homotopic but not isotopic to the unlink.

4.5. *Theorem. — Suppose that* $A \cup B$ *is an unsplittable tame link in* $\mathbf{R}^3$ *such that for some cyclic branched cover* $\pi : M \to \mathbf{R}^3$ *around* A, $\pi^{-1}(B)$ *is disconnected. Then there exists no smooth proper function* $f : \mathbf{R}^3 \to R_+$ *such that* $\mathrm{cp}(f) = A \cup B$ *and* $f(A) \leqslant f(B)$.

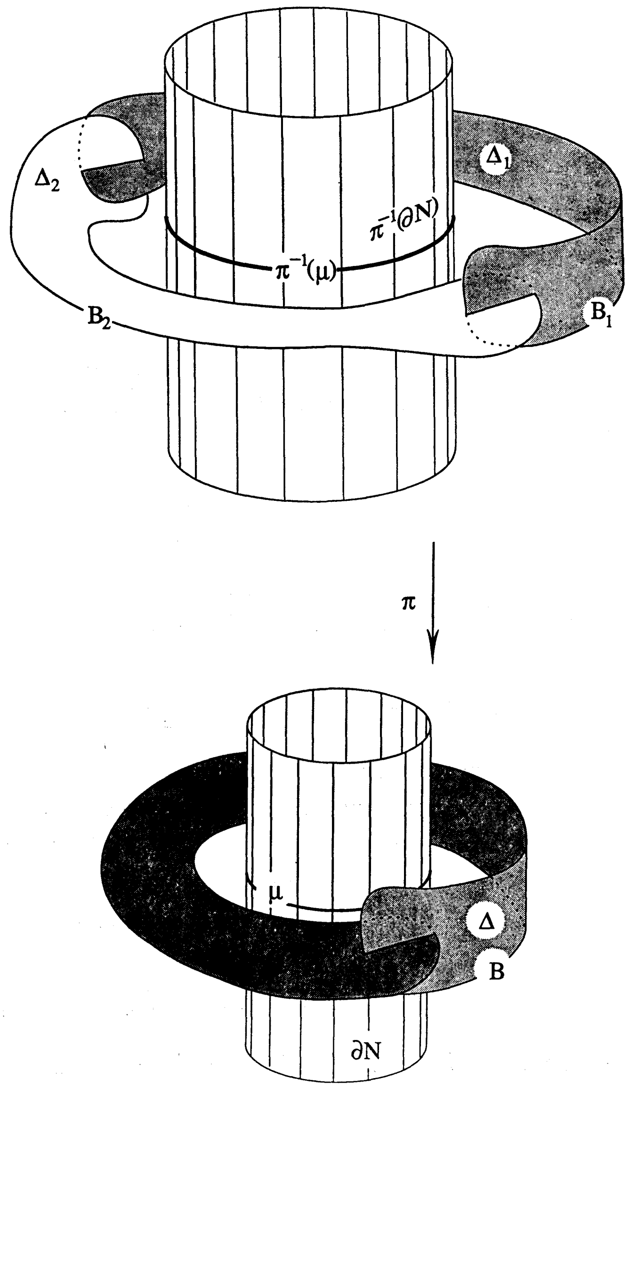
Note the asymmetry in A, B.

4.6. *Corollary. — All unsplittable, properly critical 2-component links in* $\mathbf{R}^3$ *have linking number* ± 1. *In particular, the Whitehead link and the double wrap link are not properly critical.*

Proof. — If the linking number of $A \cup B$ is 0, take the q-fold branched cover of $\mathbf{R}^3$ around A, $q \geqslant 2$. (To do this, take q copies of $\mathbf{R}^3 \backslash A$, split each along the Seifert surface for A, and glue cyclically.) Then $\pi^{-1}(B)$ is q copies of B. If the linking number of $A \cup B$ is $\pm p$, $p \geqslant 2$, take the p-fold branched cover around A. Then $\pi^{-1}(B)$ is p copies of B. In either case the hypothesis of 4.5 is satisfied.

QED

Remark. — It is instructive to see that the 2-fold branched cover of the Whitehead link $A \cup B$ actually is disconnected. The link is symmetric in the sense that there is an ambient isotopy of $\mathbf{R}^3$ which exchanges A, B. Thus, it is fair to draw A as a round flat circle in $\mathbf{R}^3$, with B clasping a solid torus neighborhood N of A. Consider the 2-fold cyclic branched cover around A, $\pi : M \to \mathbf{R}^3$. In the picture below we indicate $\pi^{-1}(B) = B_1 \sqcup B_2$. The loops B_1, B_2 are once linked to each other and are exchanged by τ, $\tau(B_1) = B_2$, $\tau(B_2) = B_1$. The tops and bottoms of the cylinders are identified. The immersed discs Δ, Δ_1, Δ_2 bounded by B, B_1, B_2 are drawn to indicate horizontal or vertical inclinations. Note that τ is rotation by 180°. The disc Δ twists by 90° as an observer goes once around the meridian μ of ∂N. The same is true of Δ_1, Δ_2 as one goes



halfway around the meridian $\pi^{-1}(\mu)$ of $\pi^{-1}(\partial N)$. For half of $\pi^{-1}(\mu)$ corresponds to the whole of μ under the projection π.

4.7. *Lemma. — If Σ is a connected non-spherical surface and $D_1, \ldots, D_k$ are disjoint compact discs interior to Σ then $E = \Sigma \backslash (D_1 \cup \ldots \cup D_k)$ is path connected and contains a Jordan curve that is* essential *in the sense that it does not bound a disc in Σ.*

Proof. — If D is a compact disc in Σ then $\Sigma \backslash D$ is homeomorphic to the punctured surface $\Sigma \backslash p$ where p is a point in Σ. Removing p from Σ does not disconnect it and any homology class of Σ has a representative 1-chain that misses p. The lemma follows by induction on k.

QED

4.8. *Lemma. — If W is a compact subset of the 2-torus T^2 which meets every essential Jordan curve in T^2, then there is a compact connected subset $G \subset W$* carrying the one-dimensional Čech homology of T^2 *in the sense that the inclusion $G \hookrightarrow T^2$ induces a surjection $\check{H}_1(G) \to \check{H}_1(T^2)$.*

Remark. — In our application of 4.8, W is part of the stable set of a Jordan curve and could be quite nasty in terms of its local topology. This is why we must use Čech theory.

Proof. — Let $g : T^2 \to [0, 1]$ be a smooth function such that $g^{-1}(0) = W$. If $\varepsilon > 0$ is a small regular value of g, then $g^{-1}[0, \varepsilon)$ is a small neighborhood of W whose boundary $g^{-1}(\varepsilon)$ is a finite union of smooth Jordan curves γ. If any such γ is essential, we contradict the assumption on W. Thus, each $\gamma \subset g^{-1}(\varepsilon)$ bounds an open disc D. Since we are working in T^2, not S^2, the disc D is unique. Inside D, g may be greater than ε, less than ε, or both.

Let D be the collection of all open discs $D \subset T^2$ such that $\partial D \subset g^{-1}(\varepsilon)$ for some regular value ε of g. Two discs $D, D' \in \mathsf{D}$ are either disjoint, nested, or equal since their boundaries are disjoint or equal. That is, D forms a lattice. To avoid unnecessary complication in dealing with *all* regular values of g, we fix some sequence of regular values $\varepsilon_n \to 0$ and set

$$\mathsf{D}_n = \{ D \in \mathsf{D} : \partial D \subset g^{-1}(\varepsilon_k) \text{ for some } \varepsilon_k \geqslant \varepsilon_n \}.$$

The collection of discs D_n increases with n.

We claim that $\bigcup_n \mathsf{D}_n$ does not cover W. For if it does then we can find finitely many discs $D_1, \ldots, D_k \in \mathsf{D}_n$ for some n, such that $\bigcup_i D_i \supset W$ and $\bar{D}_1, \ldots, \bar{D}_k$ are disjoint. By 4.7 there is a non-trivial Jordan curve in $T^2 \backslash (D_1 \cup \ldots \cup D_k)$, which contradicts the assumption on W. It follows that G is compact and non-empty, where by definition

$$G = W \cap E, \qquad E = \bigcap_n E_n \quad \text{and} \quad E_n = T^2 \backslash \bigcup_{D \in \mathsf{D}_n} D.$$

We must show that G is connected and carries the one-dimensional Čech homology of T^2.

Clearly, the sets E_n decrease as n increases and $G \subset E_n$ for all n. We claim that $g \leqslant \varepsilon_n$ on E_n. We know that E_n is the complement in T^2 of finitely many smooth discs whose closures are disjoint. (This is the virtue of working with a sequence of regular values ε_n instead of with all regular values ε.) Suppose that $g(y) > \varepsilon_n$ for some $y \in E_n$. We perturb y so that it belongs to $\mathrm{int}(E_n)$, retaining the inequality $g(y) > \varepsilon_n$. Since $G \subset E_n$ and $\emptyset \neq G \subset W = g^{-1}(0)$, there exists a point $x \in E_n$ with $g(x) = 0$. Clearly, $x \in \mathrm{int}(E_n)$. By 4.7 $\mathrm{int}(E_n)$ is connected. The Intermediate Value Theorem produces a point $z \in \mathrm{int}(E_n)$ with $g(z) = \varepsilon_n$, and this contradicts $\mathrm{int}(E_n)$ being disjoint from $g^{-1}(\varepsilon_n)$. Since z can not exist, neither can y, and $g \leqslant \varepsilon_n$ on E_n.

Since $g \leqslant \varepsilon_n$ on E_n, we see that $g = 0$ on $E = \bigcap E_n$. Since $g^{-1}(0) = W$ and $G = W \cap E$ it follows that $E \subset W$ and $E = G$. Under nested intersection, path connectedness can deteriorate as arcs become topologist's sine curves, but Čech homology persists. Thus, $E = G$ is compact, connected and carries the one-dimensional Čech homology of T^2.

QED

Proof of **4.5.** — We are given an unsplittable link $A \cup B$ such that for some cyclic branched cover around A, $\pi^{-1}(B)$ is disconnected. Seeking a contradiction, we suppose that there is a smooth proper function $f: A \to \mathbf{R}_+$ with $\mathrm{cp}(f) = A \cup B$ and $f(A) \leqslant f(B)$. The case $f(A) = f(B)$ is trivial because $A \cup B$ is disconnected while it is also the nested decreasing intersection of the balls $\{ x \in \mathbf{R}^3 : f(x) \leqslant f(A) + 1/n \}$. Thus, we may assume that $f(A) = 0$ and $f(B) = 2$. According to Lemma 4.1, the neighborhood $N = f^{-1}[0, 1]$ of A is a solid torus with core circle A.

Consider any essential Jordan curve $\gamma \subset \partial N$. We can express γ as (m, ℓ) respecting the meridian and longitude of ∂N. Since γ is essential, $(m, \ell) \neq (0, 0)$. If $m \neq 0$ then γ links A. In particular, $\gamma \cup A$ is unsplittable. If $m = 0$ then $\ell = \pm 1$ since γ is a Jordan curve, and then $\pm$ A is isotopic to γ in N. Since $A \cup B$ is unsplittable, so is $\gamma \cup B$. The net result is that $\gamma \cup A$ or $\gamma \cup B$ is unsplittable.

Let $W = W^s(B) \cap \partial N$. It is compact and its points flow to B under the gradient flow. Any essential Jordan curve γ must meet W, for otherwise under the gradient flow γ passes beyond $f^{-1}(2)$ and flows out to ∞, splitting it from both A and B. By 4.8, we conclude that W contains a compact connected subset G carrying the one-dimensional Čech homology of T^2.

Let $\pi : M \to \mathbf{R}^3$ be the hypothesized p-fold cyclic branched cover around A such that $\pi^{-1}(B)$ is disconnected, and let $B_1, \ldots, B_k$ be the components of $\pi^{-1}(B)$. Under the basic deck transformation $\tau : M \to M$, the B_i are permuted transitively because τ is transitive on any π-fiber. Since $k \geqslant 2$, the permutation is non-trivial.

Since we may take the function f to be very flat at A, it lifts to a smooth function $F : M \to \mathbf{R}_+$. The lift is τ-invariant, $F \circ \tau = F$. The gradient flow φ of f lifts to the gradient flow Φ of F and it too is τ-invariant, $\pi \circ \Phi_t(x) = \varphi_t(\pi(x))$, $\tau \circ \Phi_t(x) = \Phi_t(\tau(x))$. In particular, the stable sets of the B_i, $W^s(B_i, F)$ are permuted by τ. Set

$$W_i = W^s(B_i, F) \cap \pi^{-1}(\partial N).$$

The surface $\pi^{-1}(\partial N)$ is a 2-torus in M which encloses A. Under π it covers ∂N by wrapping its meridian p-times around the meridian of ∂N. By 3.3, $W_1, \ldots, W_k$ are compact and disjoint. They are non-trivially permuted by τ.

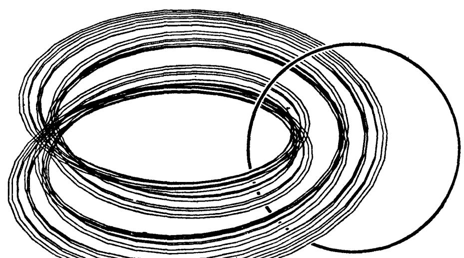
Since G is connected and carries the homology class of the meridian, its lift $\pi^{-1}(G)$ is connected. (This is the trick!) Under τ, it is sent to itself. Since $G \subset W$, $\pi^{-1}(G) \subset \pi^{-1}(W)$. But since $\pi^{-1}(G)$ is connected, it lies entirely in some W_i, and this is incompatible with the facts that $\pi^{-1}(G)$ is τ-invariant while τ non-trivially permutes $W_1, \ldots, W_k$.

QED

Remark. — If $A \cup B$ is the unlink, then the proof of 4.5 falters at only one place—although in the 2-fold branched cover around A, $\pi^{-1}(B)$ is disconnected, there is no set G to lift up and give a contradiction. For in the proof that G exists, 4.8, we needed to prohibit a longitude curve γ on the boundary of the tubular neighborhood of A from running off to ∞ under the gradient flow. If $A \cup B$ is splittable then there is no such prohibition. Considering 2.4, this is a lucky thing.

Using similar ideas, we dispose of another assertion in Theorem B.

4.9. *Corollary.* — *The link of a circle* A *and the p-adic solenoid* B *is not properly critical.*



Proof. — Suppose that $f: \mathbf{R}^3 \to \mathbf{R}_+$ is smooth, proper, and $\mathrm{cp}(f) = A \cup B$. Since the complement of $A \cup B$ in $\mathbf{R}^3$ is not diffeomorphic to $\mathbf{R}^3 \setminus \{0\}$, f has at least two critical values, and since A, B are connected it has at most two; say they are 0, 1. According to Churchill (1972), p. 349, the p-adic solenoid is never a source for a flow in $\mathbf{R}^3$, so $f(A) = 0$.

As in the proof of Theorem 4.5, consider a tubular neighborhood N of A and the set $W = W^s(B) \cap \partial N$. Since A links B, W meets every essential loop on ∂N and 4.8 yields the connected set G that carries the one-dimensional Čech homology of ∂N. Let N' be a solid torus closely approximating B. The core of N' links A p-times for some large p. In the p-fold branched cover around A, $\pi^{-1}(B)$ has p components and this leads to the same contradiction as in the proof of 4.5 above.

QED

Next we use the subtler result 3.4 to rule out knots in properly critical 2-component links. This goes most of the way to proving that a tame 2-component link in $\mathbf{R}^3$ is properly critical if and only if it is the unlink or the Hopf link. We are going to use two basic facts from knot theory. Here, " essential " means " not null-homotopic ".

4.10. *Incompressibility Lemma.* — *If* K *is a non-trivial knot in* S^3 *and* γ *is an essential loop on the boundary of its tubular neighborhood then* γ *is essential in* K^c *as well.*

4.11. *z-axis Lemma.* — *If a tame 2-sphere* Σ *in* S^3 *meets an unknot transversally and in exactly two points then there is a homeomorphism of* S^3 *to itself which carries* Σ *to the standard 2-sphere and carries the unknot to the z-axis through* ∞.

We will also use the Generalized Loop Theorem of Shapiro and Whitehead (1958). See Hempel, p. 55. The proofs of 4.10, 4.11 appear on pages 103, 345 of Rolfsen.

First we draw a topological conclusion about a tame 2-component link $A \cup B \subset S^3$ satisfying the following *slip condition*:

$$\operatorname{range}(i_A) \supset \operatorname{range}(i_B)$$

where N_A, N_B are disjoint tubular neighborhoods of A, B in S^3, and

$$i_A : \pi_1(\partial N_A) \to \pi_1(M) \qquad i_B : \pi_1(\partial N_B) \to \pi_1(M)$$

are induced by the inclusions of ∂N_A, ∂N_B into $M = S^3 \backslash (N_A \cup N_B)$. We think of γ " slipping down " from B to A as can a meridian of the Hopf link. The slip condition is the conclusion of the Slip Lemma, 3.4 where $f : \mathbf{R}^3 \to \mathbf{R}_+$ is smooth, proper, $A \cup B = \operatorname{cp}(f)$, $A = f^{-1}(0)$, $B \subset f^{-1}(2)$, and $N_A = f^{-1}[0, 1]$.

4.12. *Proposition.* — *If* $\operatorname{Lk}(A, B) = \pm 1$ *and the slip condition holds, then* $A \cup B$ *is the Hopf link.* (The converse is immediate.)

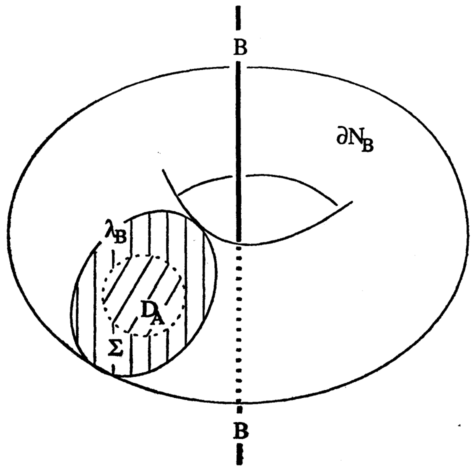
Proof. — First we show that A, B are unknotted. Consider a meridian μ_B on ∂N_B. It is homotopic in M to a loop μ'_B in ∂N_A. Since μ_B is essential in N_B^c, μ'_B is essential in ∂N_A. But $\mu'_B \simeq \mu_B \simeq$ a point in N_A^c. Therefore, ∂N_A contains an essential loop that is not essential in A^c and, by 4.10, A is unknotted.

Consider a longitude λ_B on ∂N_B. It is essential on ∂N_B and satisfies $\operatorname{Lk}(\lambda_B, B) = 0$. Let $\lambda'_B \subset \partial N_A$ be homotopic in M to λ_B. Express λ'_B as (m, ℓ) respecting the meridian and longitude of ∂N_A, μ_A, λ_A. We assume $\operatorname{Lk}(A, B) = 1$. Since λ_B is homotopic in N_B to B,

$$\operatorname{Lk}(A, \lambda'_B) = \operatorname{Lk}(A, \lambda_B) = \operatorname{Lk}(A, B) = 1.$$

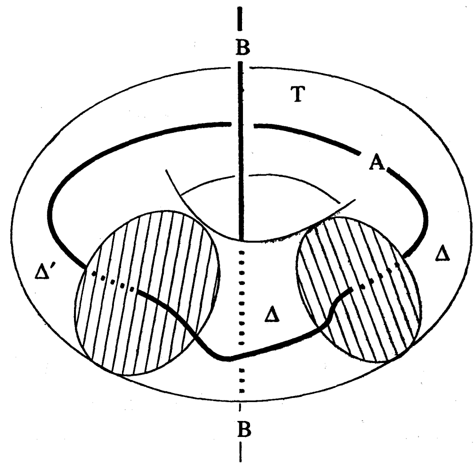
Thus, $m = 1$. On the other hand, $\operatorname{Lk}(\lambda'_B, B) = \operatorname{Lk}(\lambda_B, B) = 0$ and $\operatorname{Lk}(A, B) = 1$ imply that $\ell = 0$. Thus, $\lambda_B \simeq \mu_A$ in M. Since μ_A is null-homotopic in N_A, we see that $\lambda_B \simeq$ a point in B^c, and it follows from 4.10 that B is unknotted.

Now we know that A, B are unknotted and $\mu_A \simeq \lambda_B$ in M. That is, there is a continuous map of the cylinder $S^1 \times [0, 1]$ into M such that $S^1 \times 0$ is carried to μ_A and $S^1 \times 1$ is carried to λ_B. By the Generalized Loop Theorem, there is an embedded



cylinder $\Sigma \subset M$ with $\partial\Sigma = \mu_A \cup \lambda_B$. Cap off this cylinder with a disc $D_A \subset N_A$ such that $\partial D_A = \mu_A$ and D_A meets the core A just once, transversally. Then $D = \Sigma \cup D_A$ is an embedded disc in N_B^c with $\partial D = \lambda_B$. It meets A transversally and exactly once.

Since B is unknotted, N_B^c is a solid torus T, of which λ_B is a meridian and D is a meridian disc. We may assume that T is the standard solid torus and D is a standard meridian disc. Using its tubular neighborhood, we thicken D to a solid cylinder



$D \times [-\varepsilon, \varepsilon] = \Delta$ in T. Then $T \backslash \Delta$ is another solid cylinder Δ' whose boundary is a tame 2-sphere meeting the unknot A transversally and exactly twice. By 4.11, after a homeomorphism, A crosses the solid cylinders in the standard way, as their axis. Thus, $A \cup B$ is ambiently homeomorphic to the Hopf link.

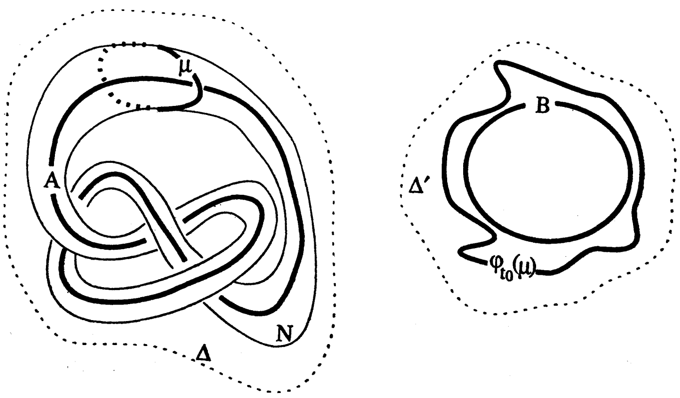
QED

4.13. *Theorem. — A tame 2-component link in* $\mathbf{R}^3$ *is properly critical if and only if it is the unlink or the Hopf link.*

Proof. — Let $L = A \cup B$ be the given link. If L is the Hopf link or the unlink it is properly critical according to 2.3, 2.4. Assume that L is properly critical for the function $f: \mathbf{R}^3 \to [0, \infty)$. We may take $f(A) = 0$, $f(B) = 2$. By 4.6 we know that $\mathrm{Lk}(A, B) = \pm 1$ or 0, and in the latter case, the link is splittable.

Case 1. $\mathrm{Lk}(A, B) = \pm 1$. — Assume that $W^u B$ attaches essentially to B. Given a tubular neighborhood N_B of B, there exists a Jordan curve $J \subset f^{-1}(3)$ near $W^u B$ such that for some $t > 0$, $J' = \varphi_{-t}(J)$ is an essential loop in N_B. An essential loop in N_B winds ℓ times along B, $\ell \neq 0$, and so $\mathrm{Lk}(A, J') = \pm \ell \neq 0$. But under φ, J' flows to J and then off to ∞, de-linking itself from A, a contradiction. Thus, we may assume that $W^u B$ Fuller attaches to B. Then by 3.4 the slip condition holds and by 4.12, L is the Hopf link.

Case 2. $\mathrm{Lk}(A, B) = 0$ *and* A, B *are splittable.* — It suffices to show that A, B are unknotted, since a splittable link of unknots is clearly the unlink. Let Δ, Δ' be disjoint 3-balls that contain A, B. *Assume that* A *is knotted.* By 4.10, all essential γ in ∂N_A are essential in A^c, and therefore must meet $W^s B$ to avoid flowing off to ∞ under φ. By 4.8 there is a compact connected set $G \subset (W^s B \cap \partial N_A)$ carrying the one-dimensional Čech homology of ∂N_A. Under φ_t, G flows to B. Choose a time $t_0 > 0$ such that $\pi_{t_0}(G) \subset \mathrm{int}(\Delta')$. Choose a Jordan curve $\mu \subset \partial N_A$ which is a homotopy merdian and lies so close

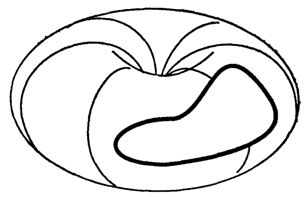


to G that $\varphi_{t_0}(\mu) \subset \mathrm{int}(\Delta')$ also. Since G carries the one-dimensional Čech homology of ∂N_A, such a μ exists. The loop $\varphi_{t_0}(\mu)$ is null-homotopic in Δ' because Δ' is a ball. Since Δ' is disjoint from A, we get a null-homotopy of μ in A^c and this is impossible, whether or not A is knotted. Thus, A is unknotted.

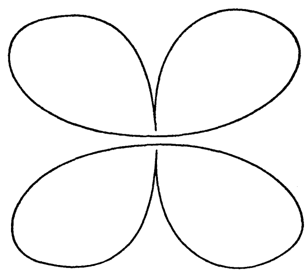
Assume that B *is knotted.* By 4.10, no essential loop in N_B can flow off to ∞ under φ. Thus, $W^u B$ Fuller attaches to B. Let μ_B be a meridian loop at B. By 3.4 it slips down to a loop μ'_B near A. Since A lies in the ball Δ, we can homotop μ'_B to a point in B^c, and thereby homotop μ_B to a point in B^c. This is impossible, whether or not B is knotted. We conclude that neither A nor B is knotted and L is the unlink.

QED

Next, we discuss links L as critical sets and ask: " to what extent can the values of f on the components of the link be specified in advance? " For example, it is impossible to specify that all the values on the components of L are equal when $L = \mathrm{cp}(f)$ and $f: \mathbf{R}^3 \to \mathbf{R}_+$ is proper. (The proof is easy and left to the reader.) On the other hand, if L is the unlink of three circles then we may require that f has only two critical values. This we do by looking at the bialy construction of the unlink of two circles and enlarging the critical set to include a new irrelevant circle on the bialy $f^{-1}(1)$.

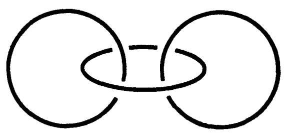


More interestingly, it is also possible to do the construction when two of the components of L are required to be minima. For this we make a " bialy/bialy join ".



The only result we prove in this vein is meant to be suggestive. Its linking hypothesis can surely be weakened to some sort of unsplittability.

4.14. *Theorem. — Suppose that* $f: \mathbf{R}^3 \to \mathbf{R}_+$ *is smooth, proper, and* $\mathrm{cp}(f) = L$ *is a link consisting of a simple chain of three unknots. Then* f *has three distinct critical values.*



Proof. — Suppose instead that f has only two critical values. Let J, J′, J″ denote the components of L, labelled so that

$$f(J) \leqslant f(J') \leqslant f(J'').$$

Choose thin tubular neighborhoods N, N′, N″ of J, J′, J″. Let γ be a loop on ∂N expressed as (m, ℓ) respecting the meridian and longitude of ∂N. If $m \neq 0$ then γ links J and can not flow off to ∞ under the gradient flow. If $\ell \neq 0$ then γ links J′ or J″ and also can not flow off to ∞. But if γ misses $W^s(J' \cup J'')$ then it does flow off to ∞ and necessarily $(m, \ell) = (0, 0)$ so γ is null-homotopic on ∂N. In other words, every essential loop on ∂N meets $W^s(J' \cup J'')$ and, by Lemma 4.8, there exists a compact connected set $G \subset (\partial N \cap W^s(J' \cup J''))$ carrying $\check{H}^1(\partial N)$. For large t,

$$\varphi_t(G) \subset\subset N' \cup N''.$$

Since G is connected, $\varphi_t(G)$ is contained in N′ or in N″. We may assume the components of L are labelled so that $\varphi_t(G) \subset N''$. For if $f(J) = f(J')$ then $\varphi_t(G)$ has no chance of being contained in N′, while if $f(J') = f(J'')$ and $\varphi_t(G) \subset N'$ then we merely exchange the names of J′, J″. Choose a meridian and longitude μ, λ of ∂N which so nearly lie in G that $\varphi_t(G) \subset\subset N''$ implies

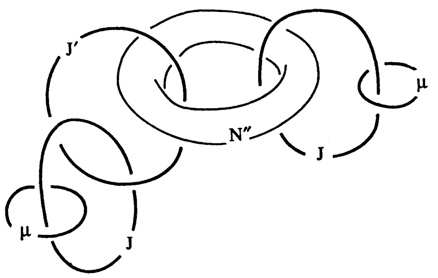
$$\varphi_t(\mu) \cup \varphi_t(\lambda) \subset N''.$$

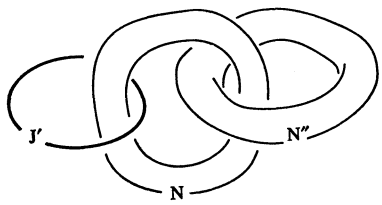
Any loop in N″ is homotopic to a multiple of J″, say $\varphi_t(\mu) \simeq kJ''$ in N″. Since the gradient flows carries μ to $\varphi_t(\mu)$ in L^c, and $N'' \subset J^c$, we see that if $k = 0$ then we have a null-homotopy of μ in J^c. But it never happens in $\mathbf{R}^3$ that a meridian of a loop is null-homotopic in the loop's complement. Thus, $k \neq 0$.

First suppose that $\mathrm{Lk}(J', J'') = 1$. Then

$$0 = \mathrm{Lk}(\mu, J') = \mathrm{Lk}(\varphi_t(\mu), J') = \mathrm{Lk}(kJ'', J') = k,$$

contradicting $k \neq 0$.





On the other hand, if $\mathrm{Lk}(J', J'') = 0$ then

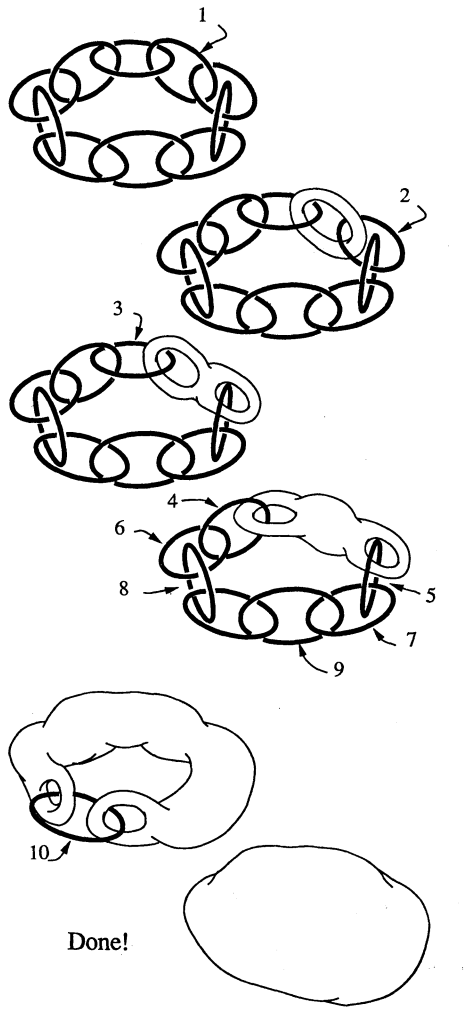
$$1 = \mathrm{Lk}(J, J') = \mathrm{Lk}(\lambda, J') = \mathrm{Lk}(\varphi_t(\lambda), J') = 0$$

QED

5. Antoine is properly critical

There are many different Antoine's Necklaces, depending on the number of solid tori forming each ring cycle. Although the following construction can be modified for as few as two tori per cycle (Bing's necklace) it is easier to visualize with at least four per cycle. Ten is convenient for computer graphics and we use it for the rest of the discussion.

By describing its level surfaces, we will define a smooth proper function f on $\mathbf{R}^3$ with the Antoine Necklace A as its critical set and a Cantor set C as its critical values, $\mathrm{cp}(f) = \mathrm{A}$, $\mathrm{cv}(f) = \mathrm{C}$. In fact, f will send A homoemorphically onto C. Just as a Cantor set in $\mathbf{R}$ is the nested intersection of interval families, so our Antoine level sur-

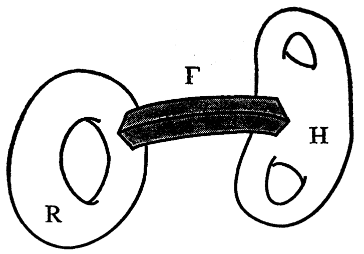


faces will be the nested intersection of (smooth) handlebody filtrations ([1]) of $\mathbf{R}^3$. The construction of these handlebody filtrations, and thus of f, is inductive. First we describe coarse handlebody filtrations, then we refine them, and eventually we take limits.

Before we get too technical, consider how we might engulf A. On the coarsest level, A is made up of ten rings. Our first move will be to swallow one of these rings; we go from the preimage of zero, a point, to the premiage of 0.1, a torus. The next step engulfs one more ring, yielding a surface of genus two. We have to be careful about creating too many holes, so when we add the next ring, we keep the level surface's genus at two. We continue eating rings in this way until there is only one ring left. In eating this final ring, we must close up the big hole in the center of the necklace, and this yields a sphere.

This required four types of moves. One to eat the first ring, one, very similar, to eat the second, another type to eat the third through ninth, and a final move to enclose all of A. We can now define these moves in a more technical language.

To be more precise, consider disjoint handlebodies H, $\mathrm{R} \subset \mathbf{R}^3$. Let Γ be a handlebody with piecewise smooth boundary such that $\mathrm{int}(\Gamma)$ is disjoint from both H and R. We call this a *bridge* from H to R if $\Gamma \cap \mathrm{H}$ and $\Gamma \cap \mathrm{R}$ are bounded by piecewise smooth curves in $\partial\mathrm{H}$ and $\partial\mathrm{R}$, and relative to some tubular neighborhood structures at $\partial\mathrm{H}$ and $\partial\mathrm{R}$, Γ locally appears to be $(\Gamma \cap \mathrm{H}) \times [0, \varepsilon]$ and $(\Gamma \cap \mathrm{R}) \times [0, \varepsilon]$. For example, Γ could be a smooth 3-cube with two opposite faces glued to $\partial\mathrm{H}$, $\partial\mathrm{R}$.



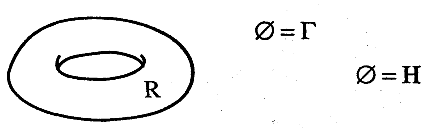
After smoothing the corners this gives the boundary connected sum of H and R.

We are going to assume that R is a solid torus (anchor ring) and

$$\mathrm{H}' \supset\supset \mathrm{G} = \mathrm{H} \cup \Gamma \cup \mathrm{R}$$

is handlebody that smooths G, Γ being a bridge from H to R. The four types of *Antoine moves* $\mathrm{H} \to \mathrm{H}'$ mentioned above are defined below. An *E̲-move* glues R to the empty

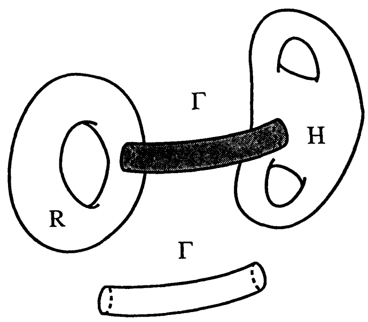
E̲-move



([1]) A smooth handlebody is a solid that is diffeomorphic to the boundary connected sum of compact solid tori. See Rolfsen, p. 46. It is a " solid torus of genus g ". All our handlebodies will be contained in $\mathbf{R}^3$ and unknotted. Generally, a *filtration* of a space X is an ordered collection of subsets $X_\alpha \subset X$ such that $\bigcap X_\alpha = \emptyset$, $\bigcup X_\alpha = X$, and if $\alpha < \beta$ then $X_\alpha \subset\subset X_\beta$. In our case the filtrations will be finite collections of handlebodies $\{H_i\}$. We refer to the transition from H_i to H_{i+1} as a " move ".

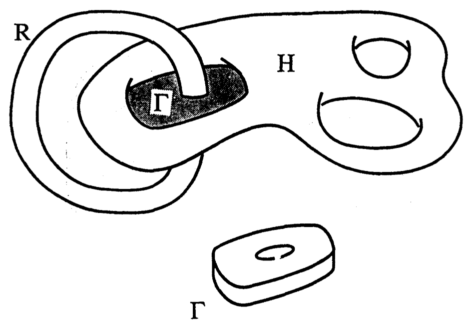
set using the empty bridge. Then H′ is R plus a collar. This was our first move.

An *E-move* glues R to H using a solid cylinder $\Gamma = D^2 \times [0, 1]$, one of whose end discs is contained in ∂H and the other in ∂R. Then H′ is the boundary connected sum of H and R plus a smoothing collar. Our second move was of this type. (If one admits the empty set as a handlebody, an $\underline{\mathrm{E}}$-move is a special type of E-move.)



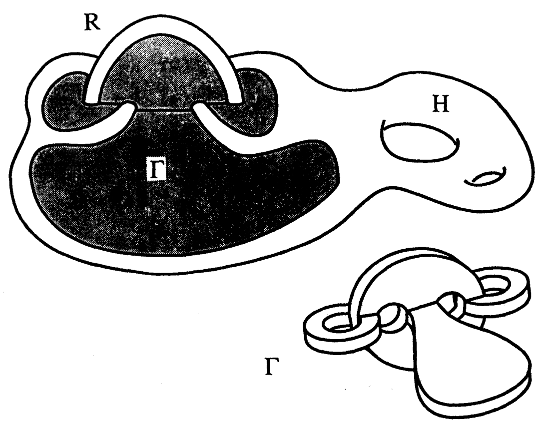
E-move

A *B-move* glues R to H using a solid torus Γ whose boundary contains two longitude belts, one glued to a longitude belt of ∂H and the other glued to a meridian belt of ∂R. This accomplishes the third move above.



B-move

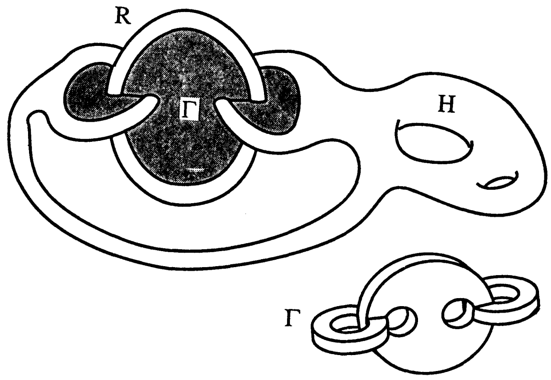
A $\bar{J}$*-move* glues R to H using a solid 4-holed torus Γ whose boundary contains two compound belts and one strip. One belt lies on ∂H and the other on ∂R. The former



$\bar{\mathrm{J}}$-move

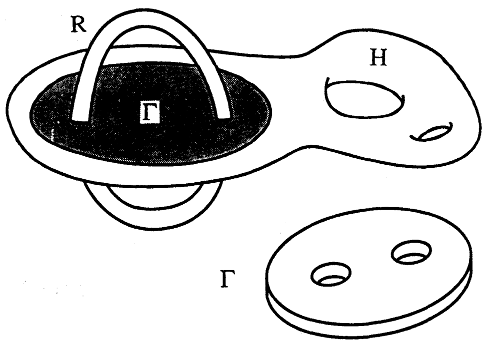
consists of two disjoint pairs of longitude and meridian belts lying on two handles of H. The latter consits of a longitude belt and two meridian belts. The torus Γ is also glued to ∂H along a strip joining the two pairs of longitude and meridian belts on the two handles of H. This move closes off some holes.

A *J-move* glues R to H using a solid 4-holed torus Γ in the same manner as the *$\bar{J}$-move*. The final strip gluing is omitted, leaving the hole in the center of the torus.



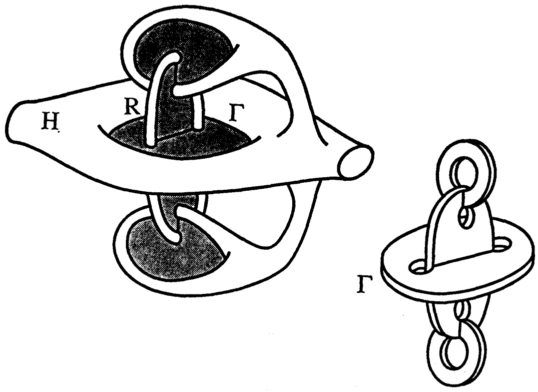
J-move

A *C-move* glues R to H using a solid 2-holed torus Γ whose boundary contains an outer longitudinal belt and two inner ones. The former is glued to a longitude belt on ∂H and the latter are glued to meridian belts of ∂R.



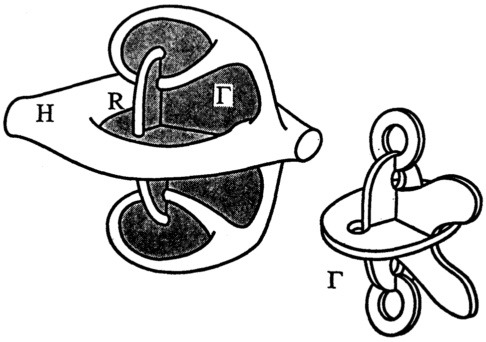
C-move

A *D-move* glues R to H using a solid 6-holed torus Γ whose boundary contains two compound belts, one on ∂H and the other on ∂R. The former consists of three disjoint longitude belts lying on three handles of H, together with two meridian belts lying on two of these handles. The latter consists of a longitude belt and four meridian belts.



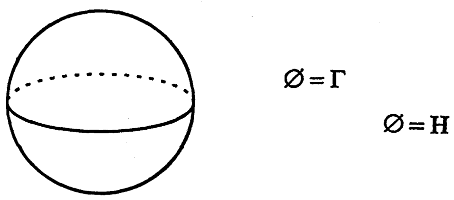
D-move

A $\bar{D}$*-move* glues R to H using a solid 6-holed torus Γ in the same way as the D-move, except that Γ is also glued to ∂H along two strips joining the free longitude belt to the meridian belt. This closes off two more holes of H than does a D-move.



$\bar{\mathbf{D}}$-move

The ninth type of Antoine move is slightly different. An *O-move* glues the ball $\mathbf{B}^3$ to the empty set using the empty bridge.



O-move

Except for the O-move, each Antoine move is a *smooth self-engulfing isotopy modulo* R, the isotopy motion taking place in H′. That is, for a family $t \mapsto \Omega_t$ of handlebodies in H′, $0 \leqslant t \leqslant 1$, we have

(i) $\Omega_0 = \mathrm{H}$, $\Omega_1 = \mathrm{H}'$, and if $s < t$ then $\Omega_s \subset\subset \Omega_t$;

(ii) the surfaces $\partial\Omega_t \backslash \mathrm{R}$ smoothly foliate the half open set $\mathrm{H}' \backslash \{ \mathrm{R} \cup \mathrm{H} \}$;

(iii) the Hausdorff distance between the boundaries of H and H′ is a fixed multiple of the diameter of R.

For example the handlebodies Ω_t for the E-move are, for small t, slightly enlarged copies of H from which a tongue extends toward R in Γ. As t increases, the tongue pierces ∂R, thickens, and eventually Ω_t becomes a slightly compressed copy of H′ which expands out to H′ as $t \to 1$. Discontinuity of $t \mapsto \Omega_t$ is necessary but the discontinuity occurs inside R.

The B-, C-, J-, $\bar{\mathrm{J}}$-, D- and $\bar{\mathrm{D}}$-moves are similar. Self-engulfing tongues extend across Γ as Ω_t grows from H to H′ modulo R. See the factorization figures below. The $\underline{\mathrm{E}}$-move is somewhat special. We take $\Omega_0 = \emptyset$ and $\Omega_t =$ expanding copies of R, $0 < t \leqslant 1$, with $\lim_{t \to 0} \Omega_t = \mathrm{R}$. Even though $t \mapsto \Omega_t$ is discontinuous at $t = 0$, properties (i)-(iii) are valid.

We want to define the handlebody filtrations leading to the Antoine level surfaces. Let R denote the collection of all the solid tori involved in the construction of A. Each $\mathrm{R} \in \mathsf{R}$ contains a canonical copy $\mathrm{A_R}$ of A. Let A denote the collection of all these small copies $\mathrm{A_R}$ of A. The handlebody filtration $\mathsf{H} = \{ \emptyset = \mathrm{H}_0, \mathrm{H}_1, \ldots, \mathrm{H}_k \}$ is an *Antoine filtration* if

a) any ring $\mathrm{R} \in \mathsf{R}$ whose boundary meets H_i is contained completely in H_i;
b) each transition $\mathrm{H}_i \to \mathrm{H}_{i+1}$ is an Antoine move as described above.

Property *a)* implies that no $\mathrm{A_R} \in \mathsf{A}$ meets the boundary of an H_i, and for fixed H_i, each small enough $\mathrm{A_R}$ is either entirely inside H_i or entirely outside it. If $\mathrm{H}_k = \mathbf{B}^3$ we say that H *eats Antoine*, while the move $\mathrm{H}_i \to \mathrm{H}_{i+1}$ *eats* $\mathrm{R} \in \mathsf{R}$ if $\mathrm{R} \subset (\mathrm{H}_{i+1} \backslash \mathrm{H}_i)$. The coarsest handlebody filtration eating Antoine is $\mathsf{H}_0 = \{ \emptyset, \mathbf{B}^3 \}$. It is the result of an O-move. We will refine H_0, getting Antoine filtrations H_n with $\mathsf{H}_0 \subset \mathsf{H}_1 \subset \mathsf{H}_2 \subset \ldots$.

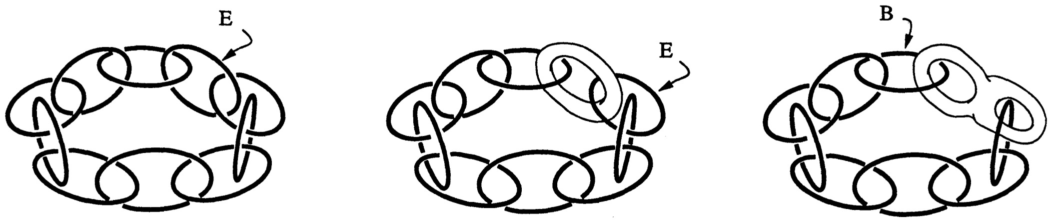
The general situation is this. We have a handlebody H, a disjoint solid torus R, and a second handlebody H′ produced by an Antoine move eating R all at once from H. We factor the Antoine move $\mathrm{H} \to \mathrm{H}'$ as the product of ten smaller scale Antoine moves eating the subrings $\mathrm{R}_0, \ldots, \mathrm{R}_9$ of R one at a time, starting from H and ending at H′. Instead of eating R all at once, this eats it bit by bit. For example, we factor the most basic move, the O-move, as

$$\mathrm{O} = \bar{\mathrm{J}}\ \mathrm{B}\ \mathrm{B}\ \mathrm{B}\ \mathrm{B}\ \mathrm{B}\ \mathrm{B}\ \mathrm{B}\ \mathrm{E}\ \underline{\mathrm{E}}.$$

The product is read from right to left, the rightmost factor operates first and the leftmost last. This particular product indicates that we start with the empty set $\mathrm{H} = \mathrm{H}_0$, make an $\underline{\mathrm{E}}$-move from H_0 eating the first subring R_0 of the main ring R and producing a solid torus H_1, then an E-move from H_1 eating the second subring R_1 of R and producing a solid 2-holed torus H_2, then a B-move eating R_2 and producing a solid 2-holed torus H_3, then another B-move eating R_3 and producing a solid 2-holed torus H_4, etc., then a seventh B-move eating R_8 and producing a solid 2-holed torus H_9, and finally a $\bar{\mathrm{J}}$-move eating R_9, producing $\mathrm{H}' = \mathbf{B}^3$, and thereby eating Antoine.

Here are the factorizations of the Antoine moves and the corresponding handlebody filtrations.

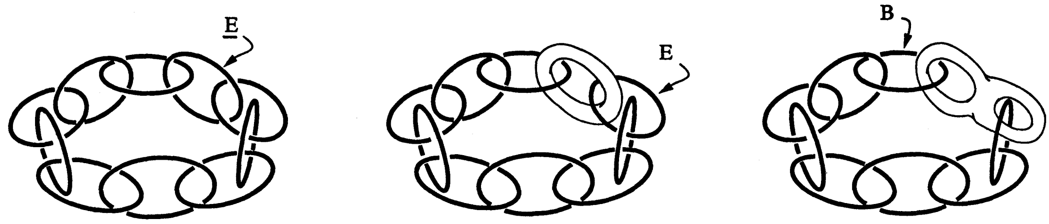
$$O = \bar{J}\ B\ B\ B\ B\ B\ B\ B\ E\ \underline{E}.$$



(a) $f = 0$ (b) $f = 0.1$ (c) $f = 0.2$

(d) $f = 0.3$ (e) $f = 0.9$ (f) $f = 1.1$

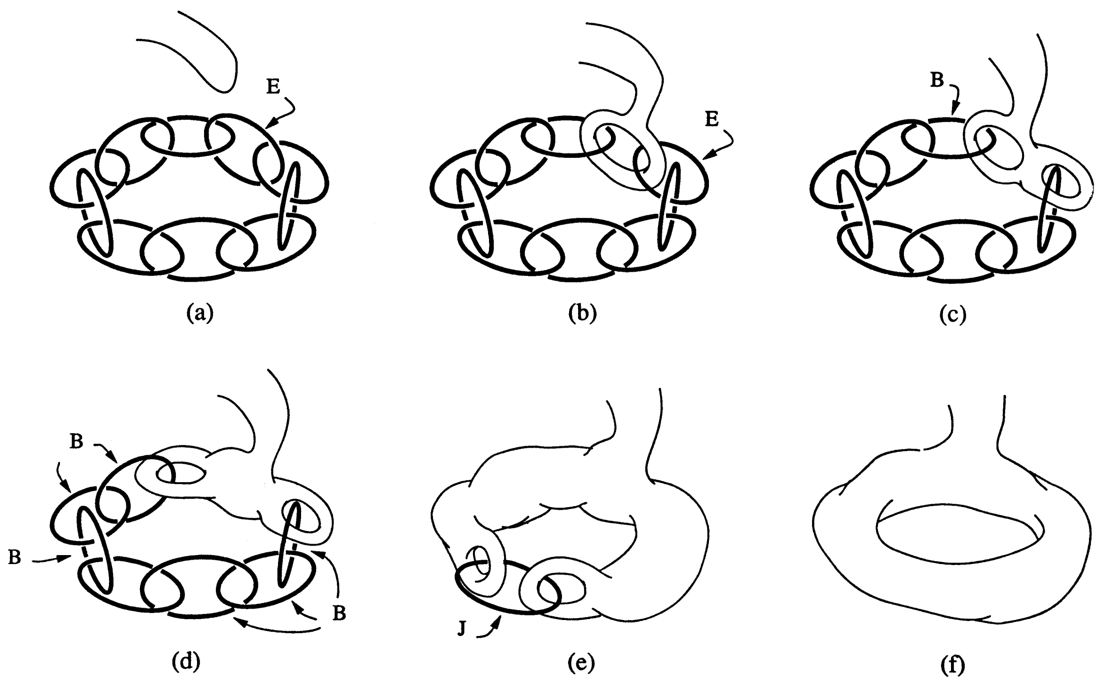
$$\underline{E} = J\ B\ B\ \ldots\ B\ E\ \underline{E}.$$



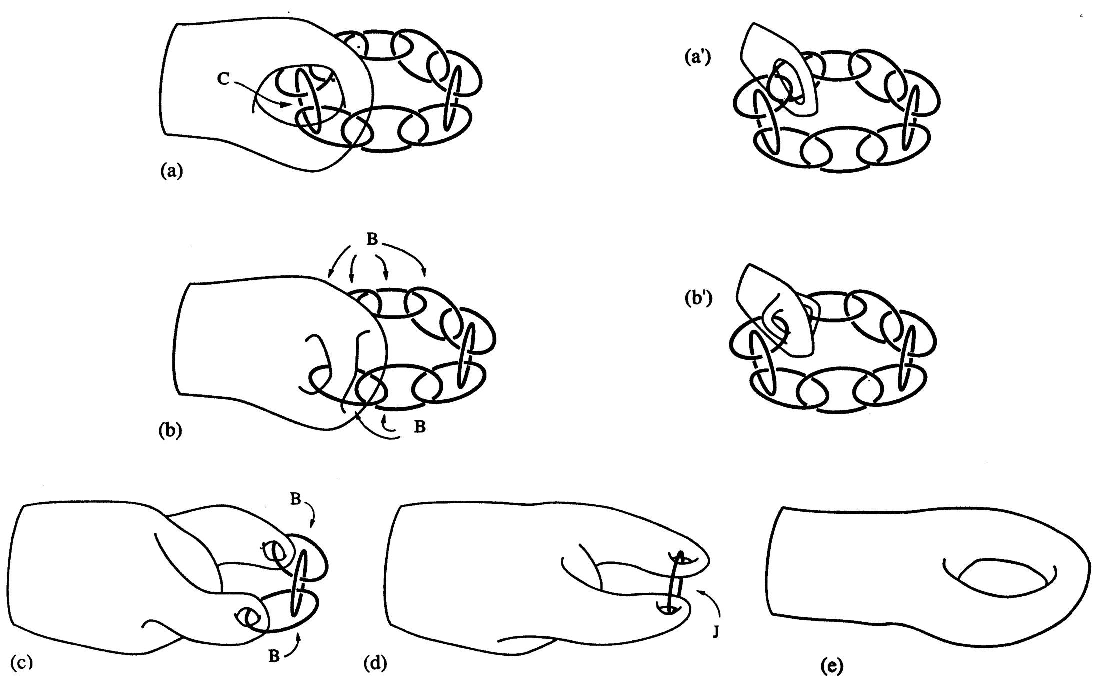
(a) (b) (c)

(d) (e) (f)

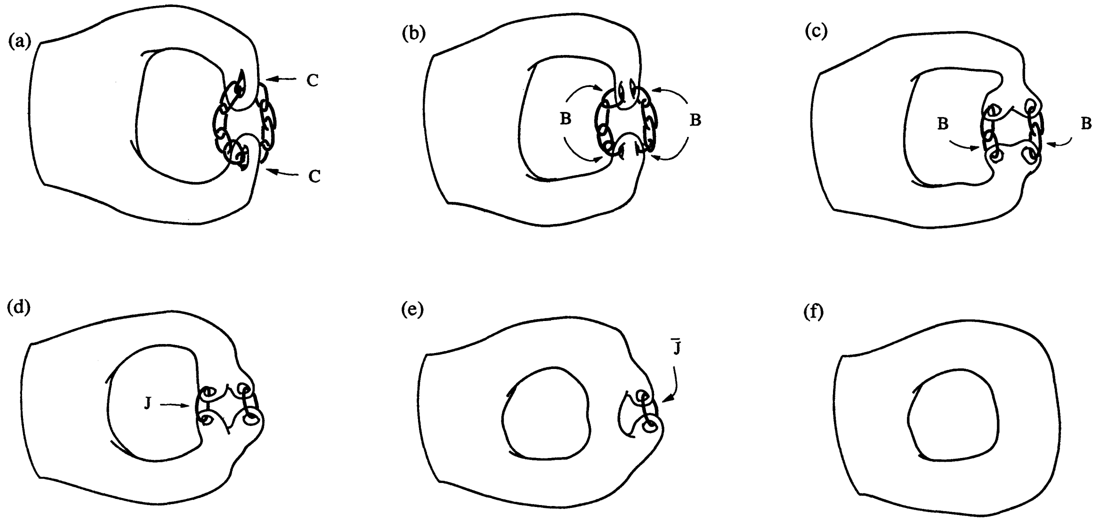
$E = J\ B\ B\ \ldots\ B\ E\ E.$



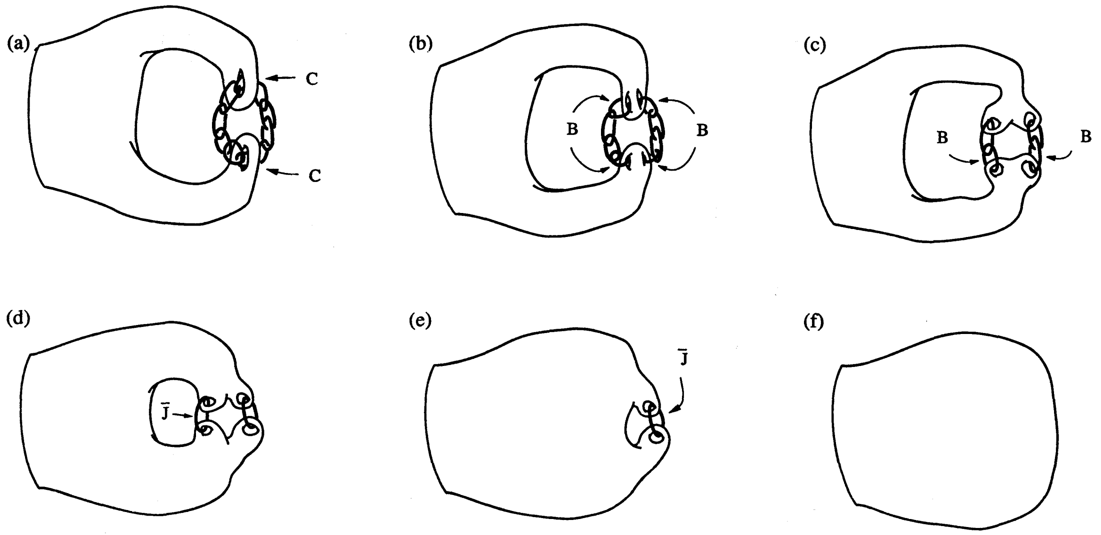
$B = J\ B\ B\ \ldots\ B\ B\ C.$



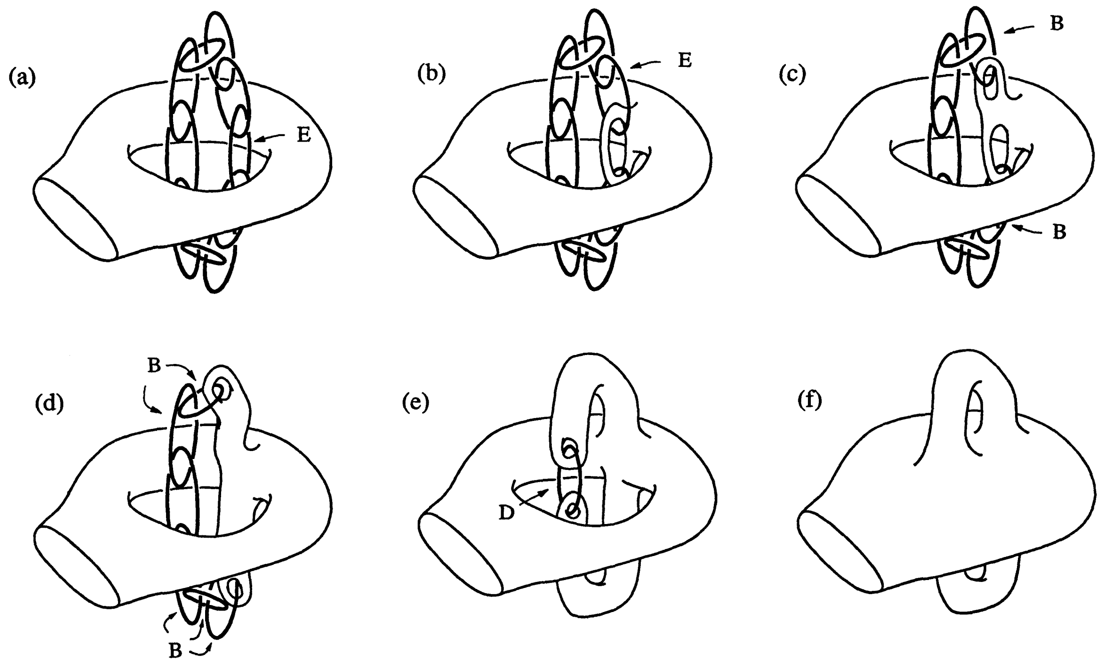
$$J = \bar{J}\, J\, B\, B \ldots B\, C\, C.$$



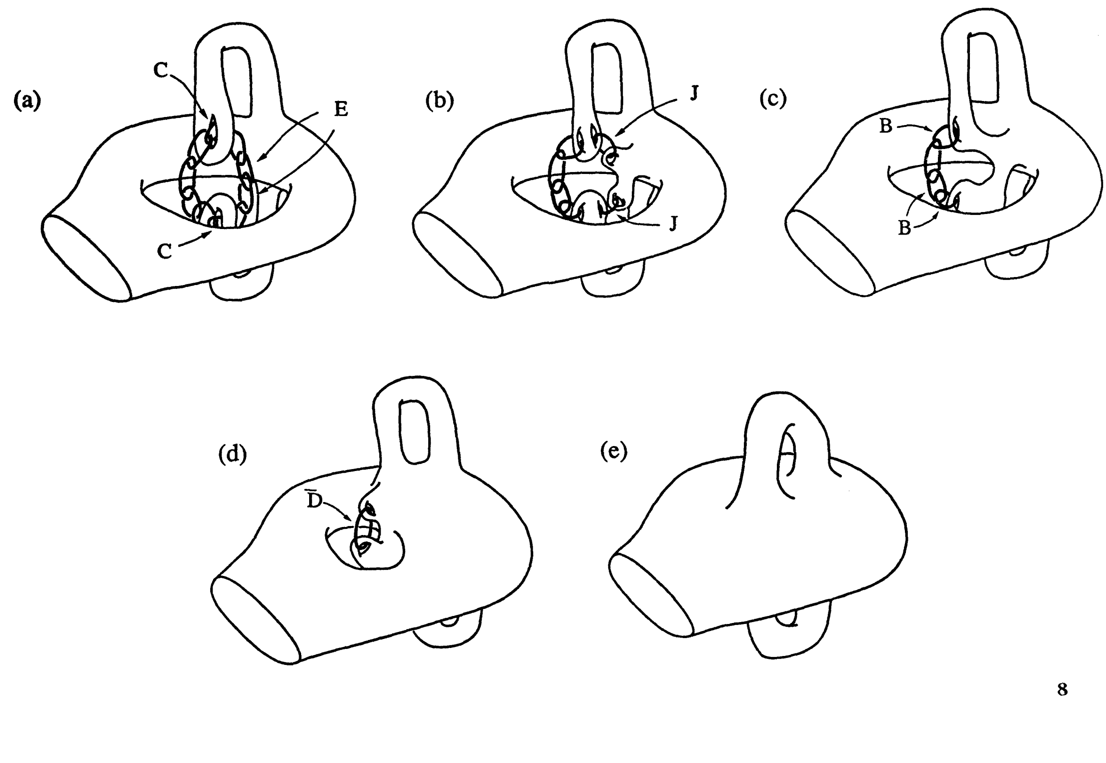
$$\bar{J} = \bar{J}\, \bar{J}\, B\, B \ldots B\, C\, C.$$



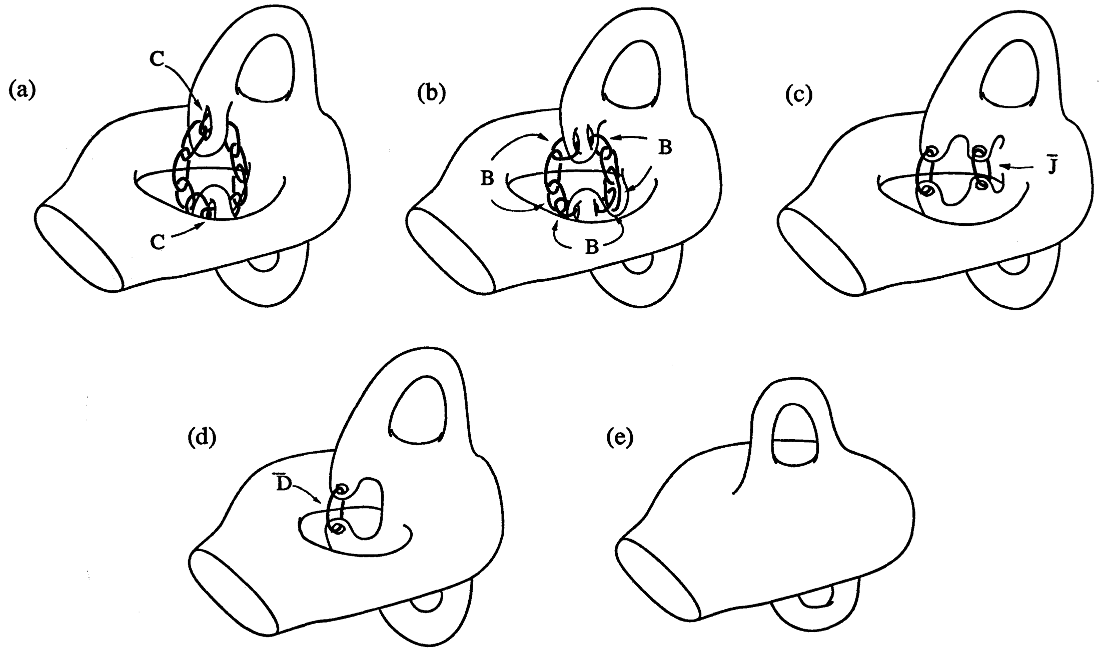
$C = D\ B\ B\ \ldots\ B\ E\ E.$



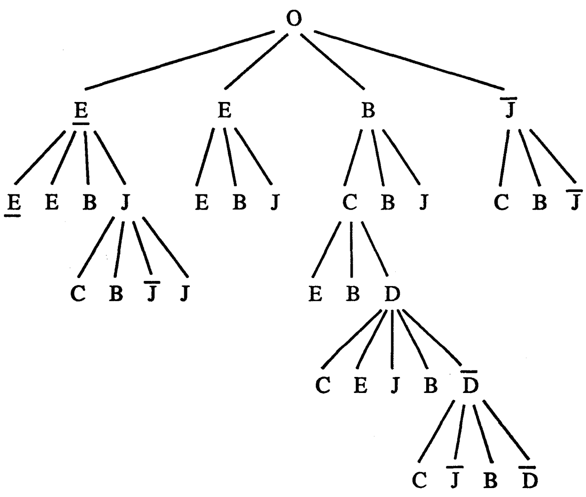
$D = \bar{D}\ B\ B\ B\ J\ J\ C\ C.$



$$\bar{D} = \bar{D}\ \bar{J}\ B\ B \ldots B\ C\ C.$$



The following chart shows the interdependence of the Antoine moves. Downward branches from the node M indicate its factors. For simplicity, we have omitted from the chart, sub-trees isomorphic to higher sub-trees. For example, we have not repeated below the three B's on the third level the large sub-tree below the B on the second level. The point of all this is that each Antoine move factors into smaller scale moves from the same finite list of Antoine moves.



O-, $\underline{\mathrm{E}}$-, E- and C-moves are used to begin a factorization, B-moves are used at intermediate stages, while J-, $\bar{\mathrm{J}}$-, D- and $\bar{\mathrm{D}}$-moves are used to end it. There is some notational ambiguity about which subrings are affected by which sub-moves, but the figures make is clear.

To refine the original filtration H_0 inductively, it is convenient to use indices in the set J_n of numbers $j \in [0, 1]$ having decimal expansions with $\leqslant n$ digits to the right of the decimal point. Thus, $\mathsf{J}_0 = \{0., 1.\}$, $\mathsf{J}_1 = \{0.0, .1, .2, \ldots, .9, 1.0\}$, etc. The order $j < j'$ on J_n is the order it inherits from being a subset of $[0, 1]$. We write $\mathsf{H}_0 = \{H_{0.}, H_{1.}\}$ where $H_{0.} = \emptyset$, $H_{1.} = \mathbf{B}^3$. Now $H_{1.}$ is the result of an O-move applied to $H_{0.}$, i.e., $H_{0.} \xrightarrow{O} H_{1.}$. Using the factorization $O = \bar{J}\ B \ldots B\ E\ \underline{E}$, we refine H_0 to $\mathsf{H}_1 = \{H_j : j \in \mathsf{J}_1\}$ where

$$\emptyset = H_{0.0} \xrightarrow{\underline{E}} H_{.1} \xrightarrow{E} H_{.2} \xrightarrow{B} \ldots \xrightarrow{B} H_{.9} \xrightarrow{\bar{J}} H_{1.0} = \mathbf{B}^3.$$

Note that H_j depends on j, not on its decimal representation; i.e., $H_{0.0} = H_{0.}$ and $H_{1.0} = H_{1.}$.

Now we repeat the process for each consecutive pair $H_{.i}, H_{.i+1}$ in H_1. If $H' = H_{.i+1}$ is the result of the Antoine move M applied to $H = H_{.i}$ then we factor M as $M_9\, M_8 \ldots M_0$ and refine $H \xrightarrow{M} H'$ accordingly:

$$H = H_{.i0} \xrightarrow{M_0} H_{.i1} \xrightarrow{M_1} \ldots \xrightarrow{M_8} H_{.i9} \xrightarrow{M_9} H_{.i+10} = H'.$$

If $i = 9$ we understand $H_{.i+10} = H_{.i+1} = H_{1.0}$. Writing out $\mathsf{H}_2 = \{H_j : j \in \mathsf{J}_2\}$ gives

($\underline{\mathrm{E}}$) $\quad \emptyset = H_{.00} \xrightarrow{\underline{E}} H_{.01} \xrightarrow{E} H_{.02} \xrightarrow{B} \ldots \xrightarrow{B} H_{.09} \xrightarrow{J} H_{.10}$

(E) $\quad H_{.10} \xrightarrow{E} H_{.11} \xrightarrow{E} H_{.12} \xrightarrow{B} \ldots \xrightarrow{B} H_{.19} \xrightarrow{J} H_{.20}$

(B) $\quad H_{.20} \xrightarrow{C} H_{.21} \xrightarrow{B} H_{.22} \xrightarrow{B} \ldots \xrightarrow{B} H_{.29} \xrightarrow{J} H_{.30}$

...

(B) $\quad H_{.80} \xrightarrow{C} H_{.81} \xrightarrow{B} H_{.82} \xrightarrow{B} \ldots \xrightarrow{B} H_{.89} \xrightarrow{J} H_{.90}$

($\bar{\mathrm{J}}$) $\quad H_{.90} \xrightarrow{C} H_{.91} \xrightarrow{C} H_{.92} \xrightarrow{B} \ldots \xrightarrow{\bar{J}} H_{.99} \xrightarrow{\bar{J}} H_{1.00} = \mathbf{B}^3.$

Again note the consistencies $H_{.i0} = H_{.i}$, $0 \leqslant i \leqslant 9$, and $H_{1.00} = H_{1.0} = H_{1.} = \mathbf{B}^3$. Applying the same process repeatedly gives $\mathsf{H}_3 = \{H_j : j \in \mathsf{J}_3\}$, $\mathsf{H}_4 = \{H_j : j \in \mathsf{J}_4\}$, etc.

We assume that the intersection of all the initial handlebodies $H_{.00\ldots01}$ is the origin.

This construction of $\mathsf{H} = \bigcup \mathsf{H}_n$ by Antoine moves gives more than just the existence of such filtrations, it gives continuity as well. The Antoine moves generate a continuous family of surfaces F_t, $0 < t \leqslant 1$, satisfying:

(i) $\quad \lim_{t \to 0} F_t = 0 \quad$ and $\quad F_1 = \partial \mathbf{B}^3$;

(ii) $\quad F_t$ bounds a region Ω_t and if $s < t$ then $\Omega_s \subset\subset \Omega_t$;

(iii) among the F_t are the handlebody boundaries ∂H_j, $j \in \mathsf{J} = \bigcup \mathsf{J}_n$. If $F_t = \partial H_j$ then the union of the nearby $F_{t'}$ forms a neighborhood of ∂H_j diffeomorphic to a neighborhood $\partial H_j \times (-\varepsilon, \varepsilon)$;

(iv) away from A, the surfaces F_t form a smooth foliation.

Because each Antoine move is a smooth self-engulfing isotopy modulo R, the ring being eaten, and the diameter of R goes exponentially to 0 with the filtration depth, it suffices to take the Cantor set of handlebody boundaries ∂H_j and fill in the complementary intervals with a smooth foliation by compact surfaces isotopic to the ∂H_j. From these properties and the way that the handlebodies H_j separate A it follows that

(v) each F_t meets A at most once and for each $a \in A$ there is a unique $t = t(a)$ such that $F_t \cap A = a$.

That is, for each ring $R \in \mathsf{R}$, no matter how small, there are $H, H' \in \mathsf{H}$ separating R from all other rings $R' \in \mathsf{R}$ of the same size. Thus, any connected surface disjoint from the handlebody boundaries ∂H_j meets A just once. On the other hand, given $a \in A$, there is a canonical nested sequence of rings $R_{j(n)} \in \mathsf{R}$ such that $a = \bigcap R_{j(n)}$. Between the corresponding handlebody pair $H_{j(n)}$, $H'_{j(n)}$ there is a surface $F_{t(n)}$. Since the $R_{j(n)}$ are nested and contained in $H'_{j(n)} \backslash H_{j(n)}$, $t(n)$ converges to limit t as $n \to \infty$. By continuity, F_t meets each $R_{j(n)}$ and therefore contains a. No other $F_{t'}$ can contain a by property (ii).

Define the map $f_0 : \mathbf{B}^3 \to [0, 1]$ according to $x \in F_{f_0(x)}$, if $x \neq 0$, and $f_0(x) = 0$ if $x = 0$. Then f_0 is continuous; on $\mathbf{B}^3 \backslash A$ it is smooth and has no critical points; on A it is injective. Thus, $f_0(A)$ is a Cantor set C_0 and f_0 sends A homeomorphically onto C_0. According to Lemma 3 of Norton and Pugh, there is a smooth homeomorphism $g : \mathbf{R} \to \mathbf{R}$ which is so flat at C_0 that $g \circ f_0$ is smooth. On $\mathbf{R} \backslash C_0$, $g' > 0$, so the critical points of $g \circ f_0$ are exactly $f_0^{-1}(C)$ where $C = g(C_0)$. All of these new critical points off A are superfluous and we can erase them by Lemma 2.1. It is then easy to extend $g \circ f_0$ from $\mathbf{B}^3$ to a smooth proper function f on $\mathbf{R}^3$ with no additional critical points.

This completes the proof of Theorem A that there exists a smooth proper function on $\mathbf{R}^3$ having Antoine's Necklace as its set of critical points.

REFERENCES

M. Brown (1962), The monotone union of open n-cells is an open n-cell, *Proc. AMS*, **12**, 812-814.

D. E. Christie (1944), Net homotopy for compacta, *Trans. AMS*, **56**, 275-308.

R. Churchill (1972), Isolated invariant sets in compact metric spaces, *JDE*, **12**, 330-352.

C. Conley (1978), *Isolated invariant sets and the Morse index*, CBMS regional conference series in math., number 38, AMS, Providence R.I.

M. E. Feign (1988), Separation properties of codimension-1 immersions, *Topology*, **27**, 319-321.

R. H. Fox and E. Artin (1948), Some wild cells and spheres in three dimensional space, *Annals of Math.*, **49**, 979-990.

F. B. Fuller (1952), A note on trajectories in the solid torus, *Annals of Math.*, **56**, 438-439.

J. Harrison and C. Pugh (1990), *A critical fractal circle*, preprint, U. of California, Berkeley.

J. Hempel (1976), *3-manifolds*, Annals of Math. Studies #86, Princeton U. Press, Princeton, N.J.

A. P. Morse (1939), The behaviour of a function on its critical set, *Annals of Math.*, **40,** 62-70.

M. Morse (1934), *The calculus of variations in the large*, AMS Colloquium Publ. XVIII, N.Y.

J. Munkres (1960), Obstructions to the smoothing of piecewise differentiable homeomorphisms, *Annals of Math.*, **72,** 521-554.

V. V. Nemytskii and V. V. Stepanov (1960), *Qualitative theory of differential equations*, Princeton U. Press, Princeton N. J.

Norton and Pugh (1991), Critical sets in the plane, *Michigan J. Math*, **38**, 441-459.

C. Robinson (1976), personal communication with C. Conley.

C. Robinson (1990), personal communication with C. Pugh.

D. Rolfsen (1976), *Knots and Links*, Publish or Perish Press, Houston Texas.

M. Saito (1990), *Links as critical sets*, preprint, U. of Texas, Austin.

A. Sard (1942), The measure of the critical values of differentiable maps, *Bull. AMS*, **48,** 883-890.

A. Shapiro and J. H. C. Whitehead (1958), A proof and extension of Dehn's Lemma, *Bull. AMS*, **64,** 174-178.

M. Shub (1987), *Global Stability of dynamical systems*, Springer-Verlag, N.Y.

W. Thurston (1990), personal communication with M. Grayson.

H. Whitney (1935), A function not constant on a connected set of critical points, *Duke Math. J.*, **1,** 514-517.

R. L. Wilder (1930), A converse to the Jordan Brouwer Separation Theorem, *Trans. AMS*, **32,** 632-657.

F. W. Wilson, Jr. (1969), Smoothing derivatives of functions and applications, *Trans. AMS*, **139,** 413-428.

F. W. Wilson, Jr. and J. Yorke (1973), Lyapunov functions and isolating blocks, *JDE*, **13,** 106-123.

M. G.
Thomas J. Watson Research Center, IBM
P.O. Box 218
Yorktown Heights, New York, 10598

C. P.
Math. Dept., U.C. Berkeley
Berkeley, California, 94720

Manuscrit reçu le 24 août 1990.

This paper was written at IBM's Thomas J. Watson Research Center, and the authors are glad to acknowledge the support they received from IBM. The pictures were executed in Adobe Illustrator, Paracompt's Swivel 3-D, WRI's Mathematica, and Hypercard.