A MIDDLE-DISTANCE LOOK AT ROOT THEORY

ROBERT F. BROWN

Department of Mathematics, University of California
Los Angeles, CA 90095-1555, U.S.A.
E-mail: rfb@math.ucla.edu

About 30 years ago, Robin Brooks [B1] undertook a study of the coincidence equation $f(x) = g(x)$ for maps $f, g: X \to Y$. The focus of the subject, as initiated ten years earlier by Schirmer [S], was the determination of the minimum cardinality of the set of solutions to the equation $f'(x) = g'(x)$ among all maps f', g' homotopic to f and g respectively. Brooks observed that there are two natural special cases of the coincidence equation. In addition to the well-known fixed point equation $f(x) = x$, for which $X = Y$ and g is the identity map, there is what he called the root equation $f(x) = a$ for which g is the constant map $g(x) = a \in Y$. Thus, given a map $f: X \to Y$ and a point $a \in Y$, he investigated $MR[f, a]$, the minimum cardinality of the set $f'^{-1}(a)$ for all maps f' homotopic to f. Brooks' original approach, which arose in the context of coincidence theory, was published in [BB] and subsequent papers and given an exposition in [K]. However, starting with [B2] (see also [BO]), Brooks presented an equivalent, but in some ways more convenient, approach to root theory based on covering space theory.

The goals of this paper are to give a more complete account of Brooks' covering space approach to root theory than can be found in [B2] and [BO] and to present arguments for some of the main results of the theory that are more direct than those currently in the literature. These arguments arose as the result of taking a fresh look at the subject of root theory independent of its background of coincidence theory and finding that some technical apparatus that is essential in the more general theory is unneccessary if one is concerned only with roots. It was a problem to decide how to present this material. On the one hand, because the results are already known, a detailed, formal exposition of the material didn't seem justified. On the other hand, a mere listing of definitions and results would fail to elucidate the novel features of the covering space approach and the non-traditional arguments. If we think of root theory as an unfamiliar village that I am trying to describe to the reader, the detailed exposition would be like listing every

---

street and house in the village, which gives the reader a lot of information but little idea of the overall plan of the village. On the other hand, a listing of definitions and results is like telling the reader what can be seen of the village while flying over it in a plane. This exposition will be considerably more detailed than that, including some sort of argument, generally of a pictorial nature, for each result, but leaving out many of the details. Thus, as its title indicates, this paper is like the description of the village as seen from the middle-distance, from the top of a nearby hill for instance. The general plan of the village and its major landmarks should be clear, but many of the details will remain obscure. Thus, my objective is to give the reader a pretty good overall idea of what root theory is like, and also to include enough information about proofs to show what the main ideas are, but omitting the technical details required to make those arguments entirely convincing.

We begin with connected finite polyhedra X and Y (root theory can start out in a more general setting, but we will need to specialize even further later on anyhow), a point $a \in Y$ and a map $f: X \to Y$. Choosing a base point $x_0 \in X$ and letting $y_0 = f(x_0)$, we have the induced homomorphism $f_\pi: \pi_1(X, x_0) \to \pi_1(Y, y_0)$. Denote the set of paths in Y that start at y_0 by $\mathcal{P}(Y, y_0)$ and impose an equivalence relation on it by defining paths σ and τ to be equivalent if $\sigma(1) = \tau(1)$ and the loop $\sigma \cdot \tau^{-1}$ is homotopic rel y_0 to $f\omega$ for some loop ω in X based at x_0 (see Figure 1).
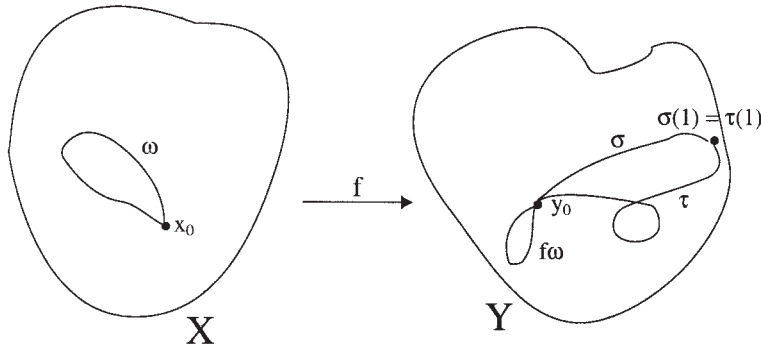

Fig. 1

We will represent the set of equivalence classes, written $\langle\sigma\rangle$, by $\hat{Y}_f$. With the usual covering space topology, we have $p_f: \hat{Y}_f \to Y$ given by $p_f\langle\sigma\rangle = \sigma(1)$, the covering space corresponding to the subgroup $f_\pi(\pi_1(X))$ of $\pi_1(Y)$, with fiber $Coker(f_\pi) = \pi_1(X)/f_\pi(\pi_1(Y))$, the cokernel of f_π. There is a natural way to lift f to a map $\hat{f}: X \to \hat{Y}_f$, namely, for $x \in X$ let α_x be any path from x_0 to x and define $\hat{f}(x) = \langle f\alpha_x\rangle$. It is easy to see that it does not matter which path we choose because, as in Figure 2, for two choices α_x and α'_x we get the same equivalence class in $\hat{Y}_f$ since $f\alpha_x \cdot (f\alpha'_x)^{-1}$ is the image of the loop $\alpha_x \cdot \alpha'^{-1}_x$.

One way to think about covering spaces is as a way of using one of the subgroups of the fundamental group of a space to spread the space out. Then the fiber over any point decomposes that point into a discrete set of factors of the point. Since the subgroup for the covering space $\hat{Y}_f$ depends on the map f, I'll call the points of $p_f^{-1}(a)$, which I denote
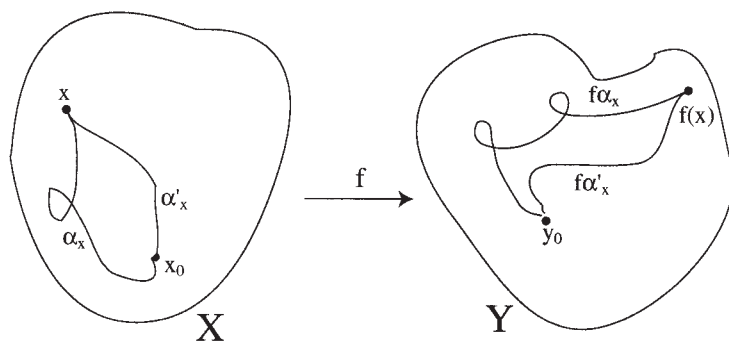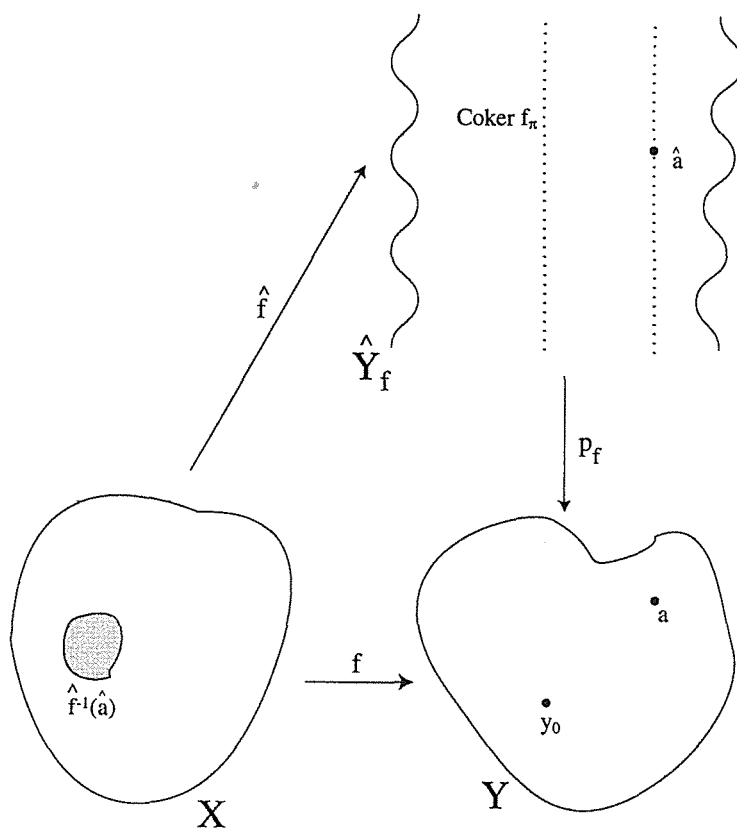
Fig. 2

Fig. 3

by $\hat{a}$, the f-*factors* of a. The closed sets $\hat{f}^{-1}(\hat{a}) \subseteq X$ are called the *root classes* of f at a (see Figure 3).

It follows from these definitions that $x, x' \in X$ are in the same root class if and only if there is a path ζ from x to x' such that the loop $f\zeta$ is homotopic to the constant loop rel a (see Figure 4).

Fig. 4
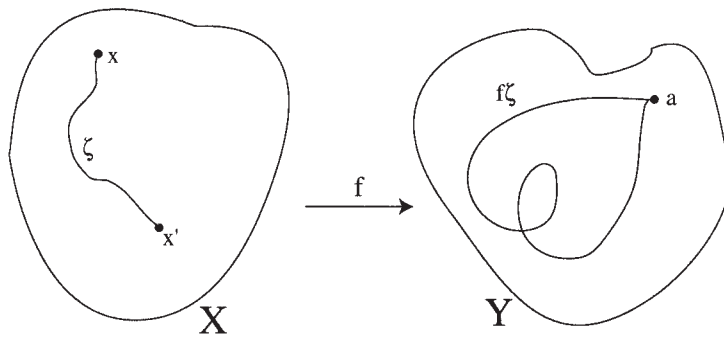

Fig. 5
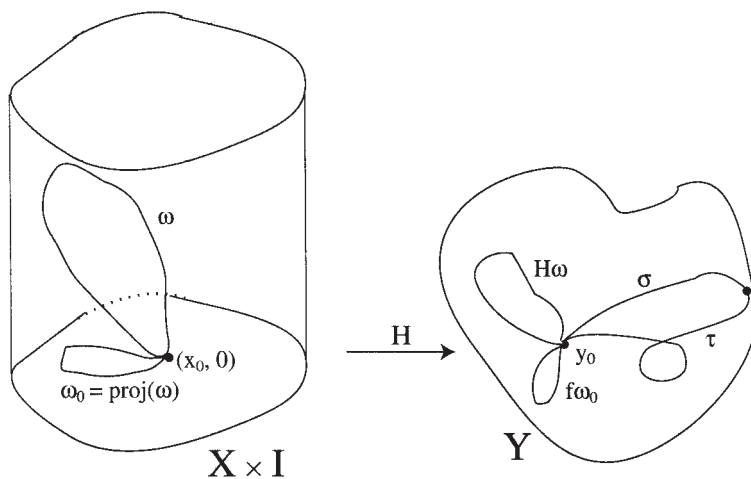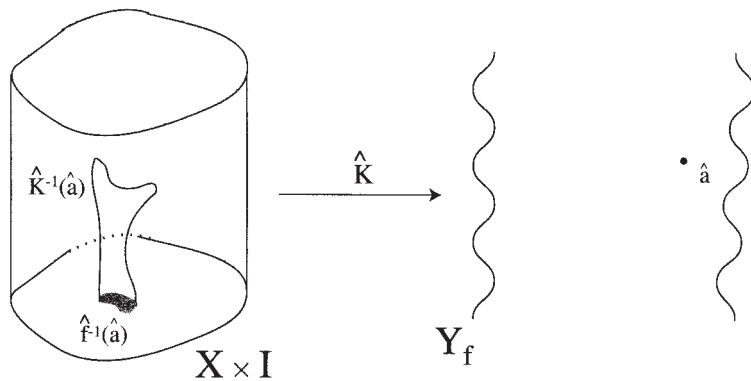

Fig. 6

Suppose $H = \{h_t\} \colon X \times I \to Y$ is a homotopy with $h_0 = f$, then the covering spaces $\hat{Y}_f$ and $\hat{Y}_H$ are both defined to be spaces of equivalence classes of $\mathcal{P}(Y, y_0)$ (take $(x_0, 0)$
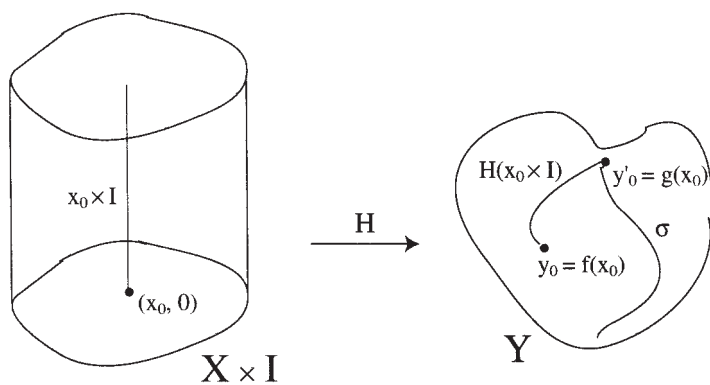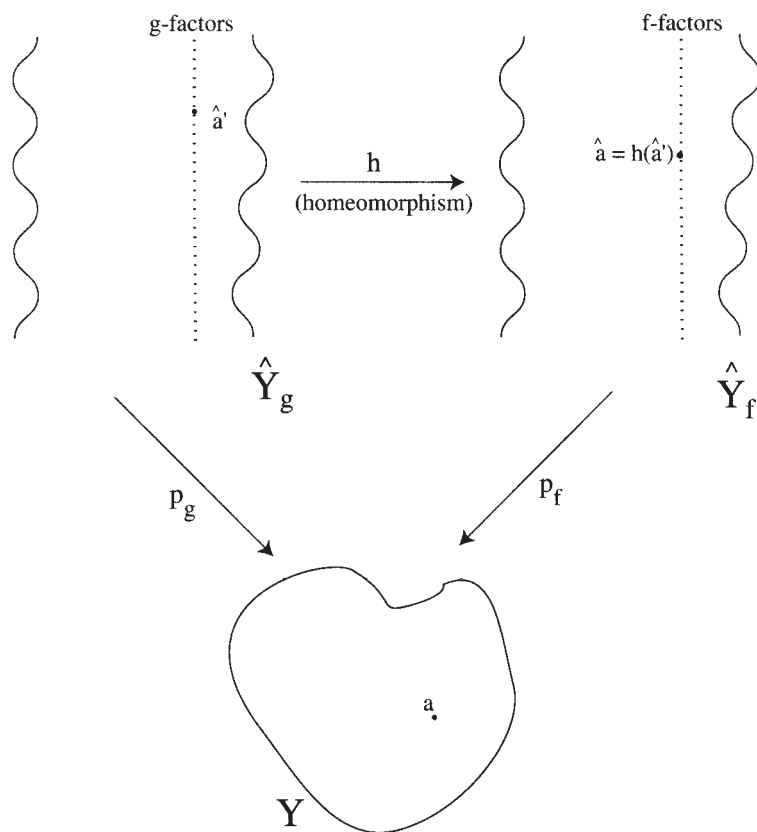
Fig. 7

Fig. 8

to be the base point of $X \times I$) but, as Figure 5 demonstrates, they are the *same* spaces, that is, $\hat{Y}_f = \hat{Y}_H$, because a loop in ω in $X \times I$ is homotopic to its projection ω_0 onto $X \times \{0\}$.
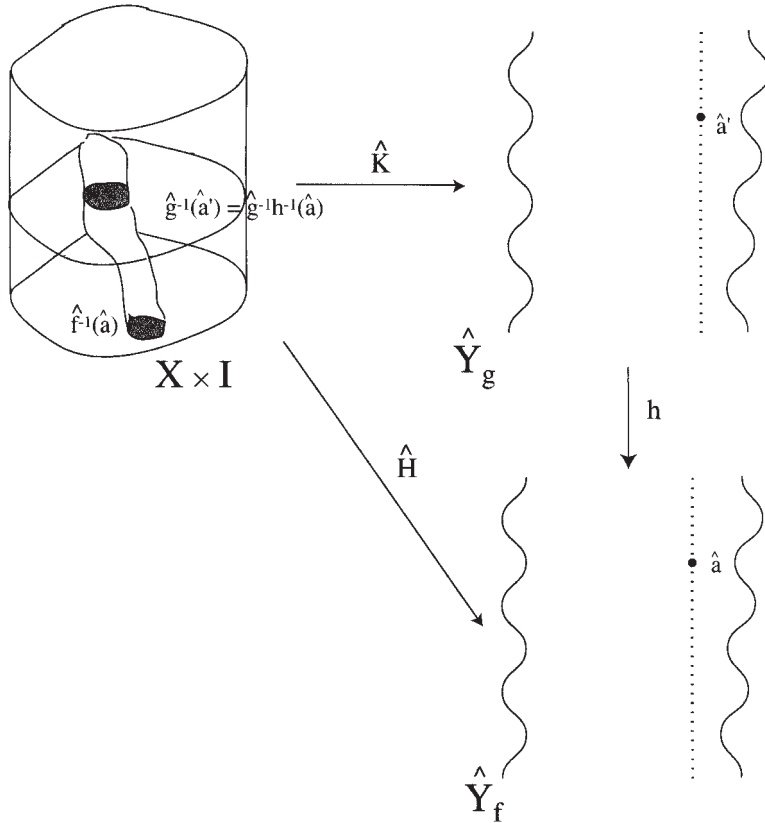
Fig. 9

Thus, when one lifts a homotopy starting at f, it doesn't matter whether the lift is thought of as a mapping to the covering space determined by f or by the entire homotopy. This observation supports the following definition. An f-factor $\hat{a} \in p_f^{-1}(a) \subseteq \hat{Y}_f$ of a is *inesssential* if there exists a homotopy $\hat{K}\colon X \times I \to \hat{Y}_f$ such that $\hat{k}_0 = \hat{f}$ and $\hat{k}_1^{-1}(\hat{a}) = \emptyset$ (see Figure 6); otherwise, $\hat{a}$ is said to be an *essential* f-factor of a.

The *Nielsen number of roots* $N(f; a)$ is defined to be the number of essential f-factors of a.

Given a homotopy $H\colon X \times I \to Y$ with $h_0 = f$ and $h_1 = g$, there is a homeomorphism $h\colon \hat{Y}_g \to \hat{Y}_f$ defined by $h\langle \sigma \rangle = \langle H(x_0 \times I) \cdot \sigma \rangle$ (note that composing a path that starts at $g(x_0)$ with the path $H(x_0 \times I)$ produces a path that starts at $f(x_0)$; see Figure 7).

As Figure 8 suggests, h is fiber-preserving.

Furthermore, as we indicate in Figure 9, h carries an inessential g-factor of a to an inessential f-factor.

Figure 9 is easy to reverse, so we can see why the following is true:

HOMOTOPY INVARIANCE THEOREM. *If $f, g\colon X \to Y$ are homotopic maps, then $N(f; a) = N(g; a)$ and therefore $N(f; a) \leq MR[f, a]$.*

From now on we will require that X and Y be connected, closed piecewise-linear (PL) manifolds of the same dimension, which we denote by n. Since Y is a manifold without boundary, so also is $\hat{Y}_f$, thus it is homogeneous and, in particular, given f-factors $\hat{a}_1$ and $\hat{a}_2$ there is a homotopy $u_t \colon \hat{Y}_f \to \hat{Y}_f$ such that u_0 is the identity map and u_1 is a homeomorphism taking $\hat{a}_1$ to $\hat{a}_2$. As Figure 10 demonstrates, if $\hat{a}_1$ is inessential, so also is $\hat{a}_2$ and therefore we conclude that either all the f-factors or essential or none are.
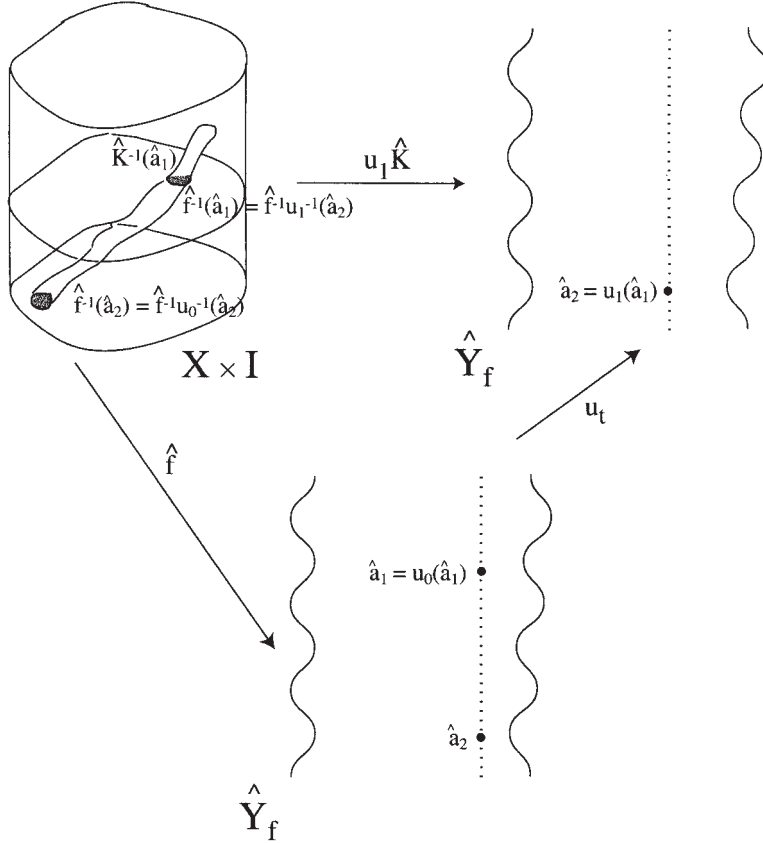

Fig. 10

To put it another way

BROOKS' THEOREM [B3]. *For any map $f \colon X \to Y$ of closed manifolds of the same dimension, either $N(f; a)$ equals the cardinality of the cokernel of f_π or else $N(f; a) = 0$.*

A major result in root theory states that, for manifolds of sufficiently high dimension, $f \colon X \to Y$ is homotopic to a map with precisely $N(f; a)$ points in the pre-image of a, that is, $N(f; a) = MR[f, a]$. The first step in proving that result is to make the pre-image finite. Since we are working with manifolds, we could appeal to transversality theory. A more elementary argument is the following. Choose a euclidean neighborhood U of a in Y and make f simplicial with respect to a triangulation of X fine enough so that if the

image of a simplex intersects a, then any simplex containing that simplex maps into U
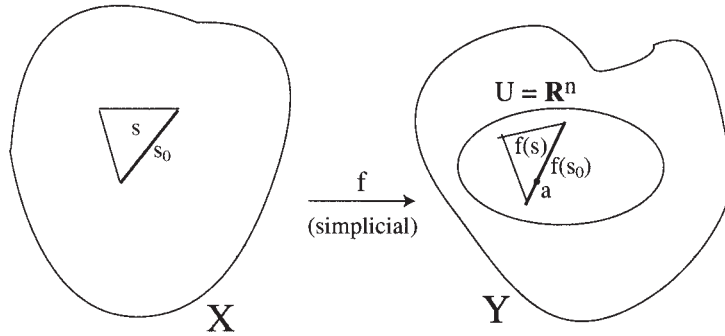(see Figure 11).

Fig. 11

Using the Homotopy Extension Theorem, we can work through the skeleta of X to
homotope the map to a map g' such that no point in the $(n-1)$-skeleton of X is mapped
to a by g'. Modifying g' over the n-simplices that it maps to a by coning, as shown in
Figure 12, we obtain a map g homotopic to f such that $g^{-1}(a)$ is finite.
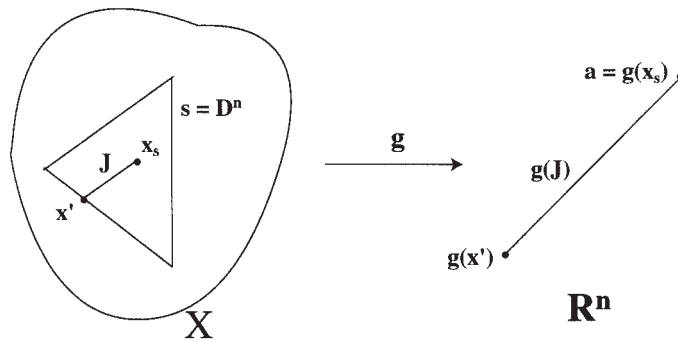

Fig. 12

Next we recall two standard results about PL manifolds (see [R]). The first comes
from the theory of general position, so we label it by GP:

GP. Let P_0 be a PL submanifold of a p-dimensional PL manifold P and let $\phi: P \to M$
be a map to an m-dimensional PL manifold M such that the restriction of ϕ to P_0 is an
embedding. If $m \geq 2p+1$ then, given $\epsilon > 0$, there is an embedding of P into M that is
ϵ-homotopic to ϕ rel P_0.

The other result comes from the theory of regular neighborhoods:

RN. If K is a subpolyhedron of an m-dimensional PL manifold M and K is homeo-
morphic to a ball, then the regular neighborhood of K in M is an m-ball and therefore
K lies in a euclidean neighborhood in M.

We will use these results to prove

REDUCTION THEOREM. *Suppose $n \geq 5$ and $f \colon X \to Y$ such that $f^{-1}(a)$ is finite. Given $x, x' \in f^{-1}(a)$ in the same root class, there is a map f' homotopic to f with $f'^{-1}(a) = f^{-1}(a) - \{x'\}$.*

We view the homotopy shrinking $f\zeta$ to the constant loop at a as a map from the disc D^2, as in Figure 13.
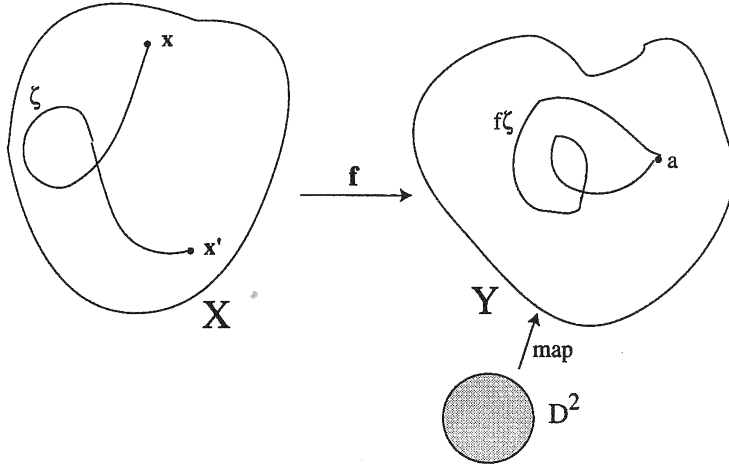

Fig. 13

Using GP, we can replace the path ζ by an arc ζ' and, after adjusting f near x and x', we can modify the map (still call it f) so that the image of ζ' is embedded, except of course at the endpoints, and the set of roots is unchanged (see Figure 14).
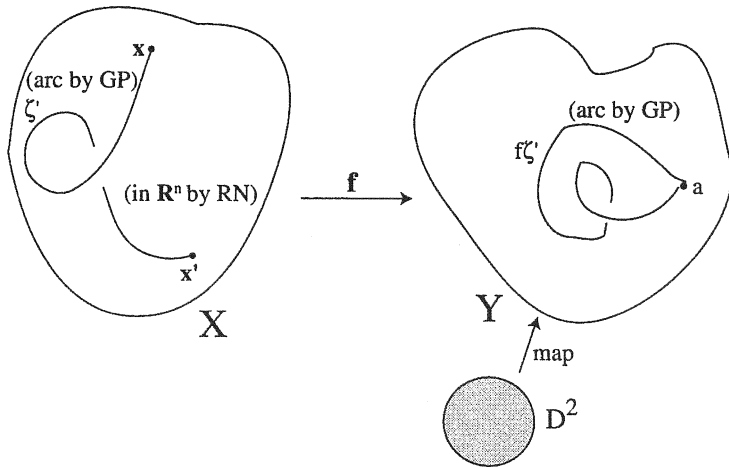

Fig. 14

Then the hypothesis $n \geq 5$ is needed to replace the map of D^2 by an embedding so that, by RN, we know that both the path ζ' and the image of D^2, and in particular the image of ζ' under the map, lie in euclidean neighborhoods, as in Figure 15.
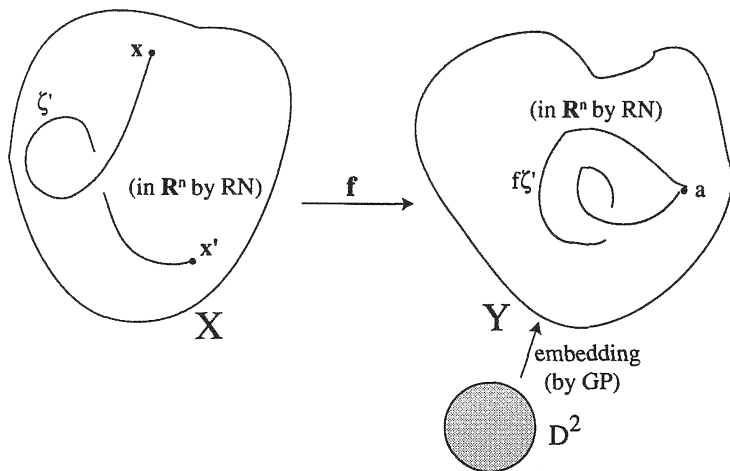

Fig. 15

The coneing argument described by Figure 16 allows us to modify f in a neighborhood of ζ in such a way that only x is mapped to a.

Fig. 16

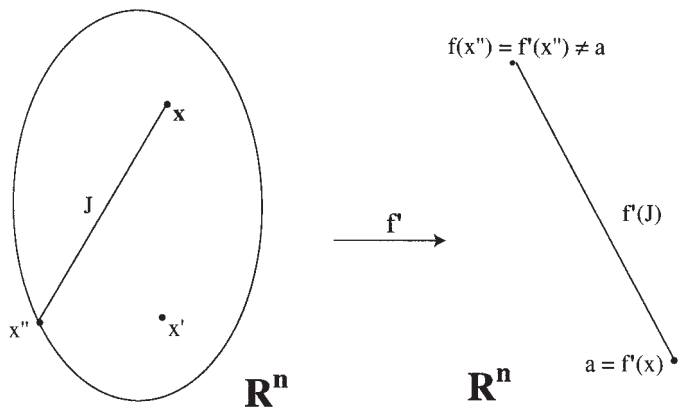The reduction argument we gave can be adjusted to hold also in the case $n = 4$ but in fact a more sophisticated argument shows that the hypothesis $n \geq 3$ is sufficient, though the result is not true in general when $n = 2$, see [L].

The Reduction Theorem allows us to replace each nonemply root class of f at a by a single point. If we further assume that the manifold X is orientable, we can go further:

REMOVAL THEOREM. *If X is orientable and $\hat{a}$ is an inessential f-factor of a such that $\hat{f}^{-1}(\hat{a}) = x$, then f may be homotoped near x to a map f' such that $f'^{-1}(a) = f^{-1}(a) - \{x\}$.*

The inessentiality of $\hat{a}$ gives us $\hat{K}: X \times I \to \hat{Y}_f$ and thus the situation described by Figure 17.

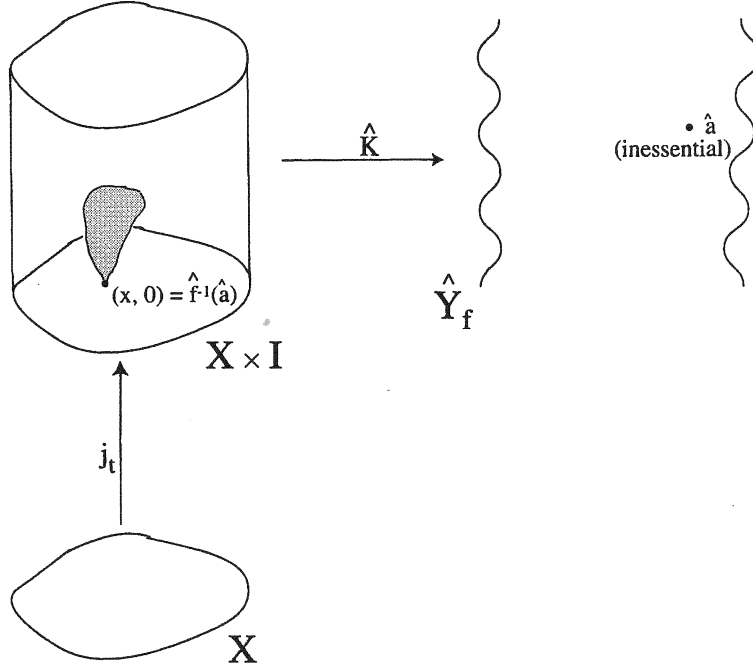Fig. 17

Thus letting

$$j_t: X \to (X \times \{t\}, (X \times \{t\}) - \hat{k}_t^{-1}(\hat{a}))$$

be inclusion, Figure 17 shows us that $\hat{f}_* j_{0*} = \hat{k}_{1*} j_{1*} = 0$. Since the orientability of X implies that $j_0: X \to (X, X - x)$ induces an isomorphism, we conclude that the local degree of $\hat{f}$ at x is zero. Thus, as Figure 18 indicates, the map f has the same property, so it can be deformed locally away from a.

The previous two theorems therefore imply, with a restriction on dimension that has allowed us to use very elementary methods,

MINIMALITY THEOREM. *If X is orientable and $n \geq 5$, then there is a map homotopic to $f: X \to Y$ with precisely $N(f; a)$ points in the pre-image of a, and so $N(f; a) = MR[f, a]$.*

An immediate consequence of this result is

DEGREE THEOREM. *Suppose X and Y are both orientable and $n \geq 5$. If the degree of $f: X \to Y$ is nonzero, then $N(f; a)$ equals the cardinality of the cokernel of f_π.*
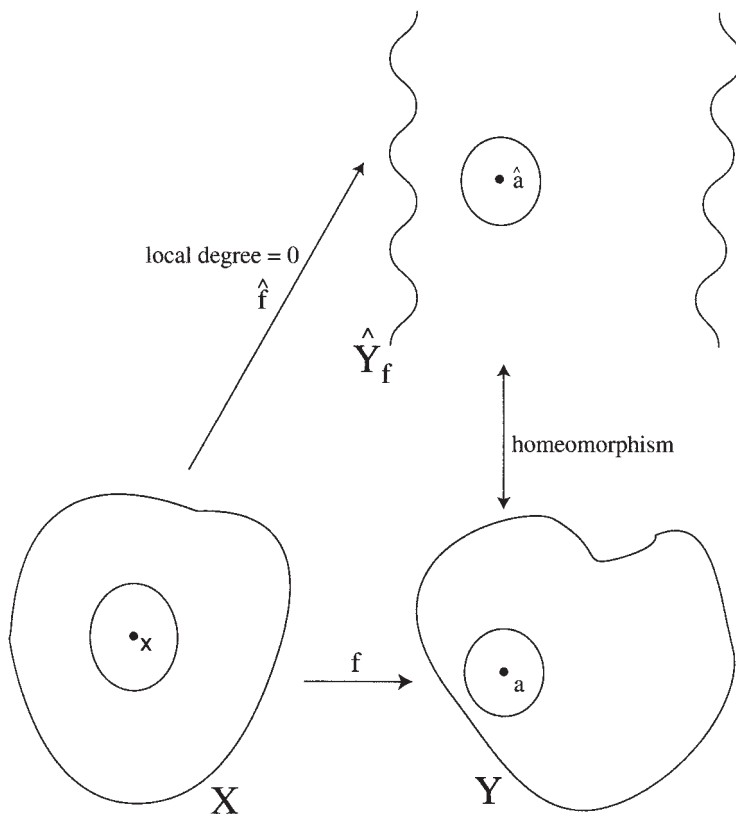
Fig. 18

The reason is that if $N(f;a)$ is not equal to the cardinality of the cokernel then, by Brooks' Theorem, $N(f;a) = 0$ and the Minimality Theorem tells us we can homotope f to a map f' with $f'^{-1}(a) = \emptyset$, that is, $f'(X) \subseteq Y - a$. But $H_n(Y - a) = 0$ so the degree of f' and thus of f, equals zero.

The Minimality Theorem does hold for dimensions $n \geq 3$, as in [L], but not for $n = 2$. However, the Degree Theorem does not really require any dimension restriction, as the following sketch of an alternative argument indicates. Without a dimension restriction, we can assume only that $\hat{f}^{-1}(\hat{a})$ is finite, but if some $\hat{a}$ is inessential, and thus all f-factors are by Brooks' Theorem, then the composition

$$ H_n(X) \xrightarrow{j_*} H_n(X, X - f^{-1}(a)) \xrightarrow{\hat{f}_*} H_n(\hat{Y}_f, \hat{Y}_f - \hat{a}) $$

is the zero homomorphism. The reason is that if that composition were not zero, then the homomorphism $\hat{f}_* j_{0*}$ that we discussed with regard to Figure 17 would have to be nonzero for at least one f-factor $\hat{a}$. The commutative diagram of Figure 19 completes this argument because the inclusion induced homomorphims j'_* is an isomorphism by the orientability of Y, so $f_*: H_n(X) \to H_n(Y)$ is again the zero homomorphism.

$$
\begin{array}{ccccc}
& & 0 & & \\
& \nearrow & & \searrow & \\
H_n(X) & \xrightarrow{\ j_* \ } & H_n(X, X - f^{-1}(a)) & \xrightarrow{\ \hat{f}_* \ } & H_n(\hat{Y}_f, \hat{Y}_f - p_f^{-1}(a)) \\
\downarrow f_* & & \searrow f_* & & \downarrow p_{f*} \\
H_n(Y) & \xrightarrow[\ \ \ \ \ \ \ \ j'_* \ \ \ \ \ \ \ \ ]{} & & & H_n(Y, Y - a)
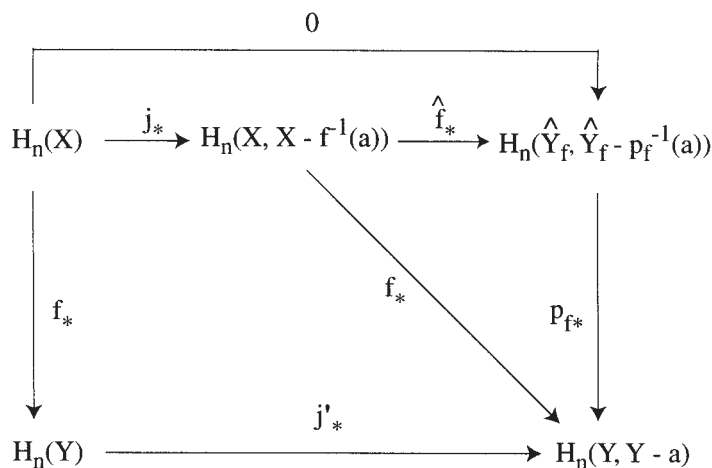\end{array}
$$

Fig. 19

References

[B1]   R. BROOKS, *Coincidences, roots and fixed points*, Doctoral Dissertation, Univ. of Calif., Los Angeles, 1967

[B2]   R. BROOKS, *On the sharpness of the Δ_2 and Δ_1 Nielsen numbers*, J. Reine Angew. Math. 259 (1973), 101–108.

[B3]   R. BROOKS, *Certain subgroups of the fundamental group and the number of roots of $f(x) = a$*, Amer. J. Math. 95 (1973), 720–728.

[BB]   R. BROOKS and R. BROWN, *A lower bound for the Δ-Nielsen number*, Trans. Amer. Math. Soc. 143 (1969), 555–564 .

[BO]   R. BROOKS and C. ODENTHAL, *Nielsen numbers for roots of maps of aspherical manifolds*, Pacific J. Math. 170 (1995), 405–419.

[K]   T. KIANG, *The Theory of Fixed Point Classes*, Springer, 1989.

[L]   X. LI, *On the root classes of mapping*, Acta Math. Sinica 2 (1986), 199–206.

[R]   B. RUSHING, *Topological Embeddings*, Academic Press, 1973.

[S]   H. SCHIRMER, *Mindestzahlen von Koinzidenzpunkten*, J. Reine Angew. Math. 194 (1955), 659–671.