Modeling an acceptable proof on the heuristic argument above would be quite difficult. Fortunately, a straightforward, if less visual, proof may be derived directly from Sard’s theorem.

A useful object in proving the Whitney theorem is the tangent bundle of a manifold \( X \) in \( \mathbb{R}^N \). The tangent spaces to \( X \) at various points are vector subspaces of \( \mathbb{R}^N \) that will generally overlap one another. The tangent bundle \( T(X) \) is an artifice used to pull them apart. Specifically, \( T(X) \) is the subset of \( X \times \mathbb{R}^N \) defined by

\[
T(X) = \{(x, v) \in X \times \mathbb{R}^N : v \in T_x(X)\}.
\]

\( T(X) \) contains a natural copy \( X_0 \) of \( X \), consisting of the points \((x, 0)\). In the direction perpendicular to \( X_0 \), it contains copies of each tangent space \( T_x(X) \), embedded as the sets \((x, v) : v \) fixed\).

Any smooth map \( f : X \to Y \) induces a global derivative map \( df : T(X) \to T(Y) \), defined by \( df(x, v) = (f(x), df_x(v)) \). Note that \( T(X) \) is a subset of Euclidean space; i.e., \( X = \mathbb{R}^N \), so \( T(X) = \mathbb{R}^N \times \mathbb{R}^N \). Therefore if \( Y = \mathbb{R}^M \), then \( df \) maps a subset of \( \mathbb{R}^{2M} \) into \( \mathbb{R}^{2M} \). We claim that \( df \) is smooth. For since \( f : X \to \mathbb{R}^M \) is smooth, it extends around any point to a smooth map \( F : U \to \mathbb{R}^M \), where \( U \) is an open set of \( \mathbb{R}^N \). Then \( dF : T(U) \to \mathbb{R}^{2M} \) locally extends \( df \). But \( T(U) \) is all of \( U \times \mathbb{R}^N \), an open set in \( \mathbb{R}^{2N} \), and as a map of this open set, \( df \) is obviously defined by a smooth formula. This shows that \( df : T(X) \to \mathbb{R}^{2M} \) may be locally extended to a smooth map on an open subset of \( \mathbb{R}^{2N} \), meaning that \( df \) is smooth.

The chain rule says that for smooth maps \( f : X \to Y \) and \( g : Y \to Z \), the composite \( dg \circ df : T(X) \to T(Z) \) equals \( d(g \circ f) \). Consequently, if \( f : X \to Y \) is a diffeomorphism, so is \( df : T(X) \to T(Y) \), for the chain rule implies that \( df^{-1} \circ df \) is the identity map of \( T(X) \) and \( df \circ df^{-1} \) is the identity map of \( T(Y) \). Thus diffeomorphic manifolds have diffeomorphic tangent bundles. As a result, \( T(X) \) is an object intrinsically associated to \( X \); it does not depend on the ambient Euclidean space.

Note that if \( W \) is an open set of \( X \), and hence also a manifold, then \( T(W) \) is the subset \( T(X) \cap (W \times \mathbb{R}^N) \) in \( T(X) \). Since \( W \times \mathbb{R}^N \) is open in \( X \times \mathbb{R}^N \), \( T(W) \) is open in the topology of \( T(X) \). Now suppose that \( W \) is the image of a local parametrization \( \phi : U \to W \), \( U \) being an open set in \( \mathbb{R}^k \). Then \( d\phi : T(U) \to T(W) \) is a diffeomorphism. But \( T(U) = U \times \mathbb{R}^k \) is an open subset of \( \mathbb{R}^{2k} \), so \( d\phi \) serves to parametrize the open set \( T(W) \) in \( T(X) \). Since every point of \( T(X) \) sits in such a neighborhood, we have proved

**Proposition.** The tangent bundle of a manifold is another manifold, and \( \dim T(X) = 2 \dim X \).

Now we prove a version of Whitney’s result.

**Theorem.** Every \( k \)-dimensional manifold admits a one-to-one immersion in \( \mathbb{R}^{2k+1} \).

**Proof.** In fact, if \( X \subset \mathbb{R}^N \) is \( k \)-dimensional and \( N > 2k + 1 \), we shall produce a linear projection \( \mathbb{R}^N \to \mathbb{R}^{2k+1} \) that restricts to a one-to-one immersion of \( X \). Proceeding inductively, we prove that if \( f : X \to \mathbb{R}^M \) is an injective immersion with \( M > 2k + 1 \), then there exists a unit vector \( a \in \mathbb{R}^M \) such that the composition of \( f \) with the projection map carrying \( \mathbb{R}^M \) onto the orthogonal complement of \( a \) is still an injective immersion. Now the complement \( H = \{ b \in \mathbb{R}^M : b \perp a \} \) is an \( M - 1 \) dimensional vector subspace of \( \mathbb{R}^M \), hence isomorphic to \( \mathbb{R}^{M-1} \); thus we obtain an injective immersion into \( \mathbb{R}^{M-1} \).

Define a map \( h : X \times X \times \mathbb{R} \to \mathbb{R}^M \) by \( h(x, y, t) = [f(x) - f(y)] \). Also, define a map \( g : T(X) \to \mathbb{R}^M \) by \( g(y, v) = df_x(v) \). Since \( M > 2k + 1 \), Sard’s theorem implies that there exists a point \( a \in \mathbb{R}^M \) belonging to neither image; note that \( a \neq 0 \), since \( 0 \) belongs to both images.

Let \( \pi \) be the projection of \( \mathbb{R}^M \) onto the orthogonal complement \( H \) of \( a \). Clearly \( \pi \circ f : X \to H \) is injective. For suppose that \( \pi \circ f(x) = \pi \circ f(y) \). Then the definition of \( \pi \) implies that \( f(x) - f(y) = ta \) for some scalar \( t \). If \( x \neq y \) then \( t \neq 0 \), because \( f \) is injective. But then \( h(x, y, 1/t) = a \), contradicting the choice of \( a \).

Similarly, \( \pi \circ f : X \to H \) is an immersion. For suppose that \( v \) is a nonzero vector in \( T_x(X) \) for which \( d(\pi \circ f)_x(v) = 0 \). Because \( \pi \) is linear, the chain rule yields \( \pi \circ dv_x = df_x \cdot v \). Thus \( \pi \circ dv_x = 0 \), so \( dv_x(v) = ta \) for some scalar \( t \). Because \( f \) is an immersion, \( t \neq 0 \). Thus \( g(x, 1/t) = a \), again contradicting the choice of \( a \). Q.E.D.

For compact manifolds, one-to-one immersions are the same as embeddings, so we have just proved the embedding theorem in the compact case. In general, we must modify the immersion to make it proper—a topological, not a differential problem. The situation is typical of differential topology; quite often fundamental differential concepts are most naturally and intuitively developed for compact manifolds, then extended by technical tricks to arbitrary manifolds. Rather than allow such technicalities to divert you now