Math 396. Why the universe cannot be S^4.

1. Motivation

Let $E \to M$ be a C^p vector bundle over a C^p manifold with corners M, $0 \le p \le \infty$. We have seen in the handout on metric tensor operations that, via a C^p partition of unity, E admits a Riemannian metric (of class C^p, as always). It was a crucial step in the fibral non-degeneracy of the construction that we worked throughout with a pseudo-Riemannian metric that is definite (positive or negative definite) on fibers. If we had tried to built a pseudo-Riemannian metric (such as a Lorentz metric) with some "mixed" signature, then there would have arisen an annoying point in the proof, which is to check that the bilinear form on fibers made with a partition of unity doesn't have some accidental cancellations causing the bilinear form on some fiber to be degenerate. This may seem like a minor glitch, but it is a genuine issue: if E has constant rank n and we write $n = n_+ + n_-$ with $n_+, n_- > 0$ then there are *non-trivial obstructions* to the existence of a pseudo-Riemannian metric of signature (n_+, n_-) on $E \to M$. What is the obstruction? It is a bundle analogue of the "light cone" decomposition for indefinite quadratic spaces over $\mathbf{R}$:

Theorem 1.1. *If E admits a pseudo-Riemannian metric B with signature (n_+, n_-), then there exist C^p subbundles $E^+, E^- \subseteq E$ with respective ranks n_+ and n_- such that the pseudo-Riemannian metric B has positive-definite restriction to the fibers of E^+ and negative-definite restriction to the fibers of E^-. In particular, natural map of bundles $E^+ \oplus E^- \to E$ is an isomorphism.*

The proof of this theorem is given in the next two sections. The existence of a "splitting" of E into a direct sum of two subbundles with positive rank is a serious restriction on E that need not always be satisfied. To give an example, consider the case $E = TM$ with M of class C^{p+1} (so E is of class C^p, $0 \le p \le \infty$). In this case, the theorem gives a non-trivial condition on M that it admit a pseudo-Riemannian metric with an indefinite signature (n_+, n_-), namely that the tangent bundle must be a direct sum of C^p subbundles with respective ranks n_+ and n_-. Why is this a non-trivial condition? For example, I claim that this prevents a wide class of manifolds from admitting a Lorentz metric:

Corollary 1.2. *Assume M is of class C^{p+1} with $0 \le p \le \infty$. If all closed paths in M are contractible to a point and M has no non-vanishing C^p vector fields, then TM does not admit any C^p line bundle direct summand, and so M does not admit a Lorentz metric of class C^p.*

By the hairy ball theorem and the discussion preceding Theorem 4.1 below, even-dimensional spheres satisfy the hypotheses in this corollary. It follows that the spheres S^{2n} do not admit a C^∞ Lorentz metric, nor even one of class C^0. In particular, since the General Theory of Relativity asserts that the universe is a smooth Lorentzian 4-manifold, *the universe cannot be S^4* as a C^∞ manifold. (It is crucial to the argument that the Lorentz metric really be given at *all* points. Singularities in the metric – black holes and so forth – ruin the proof. So in a sense the title of this handout is misleading.)

Proof. By Theorem 1.1, if M were to admit a Lorentz metric then TM would admit a line bundle direct summand. Hence, we just have to show such line bundles do not exist. In Theorem 4.1 below, it will be proved that under the topological hypothesis in the corollary, all C^p line bundles on M are trivial. Hence, if TM admits a C^p line bundle direct summand then it admits the trivial line bundle $M \times \mathbf{R}$ as such a direct summand. But the trivial line bundle has a non-vanishing C^p global section (the constant section 1, for example), and so when realized as a C^p-subbundle of TM this becomes a C^p vector field that is non-vanishing on fibers. That is, we have a C^p vector field that is nonzero in all tangent spaces. This contradicts the hypotheses on M. ■

2. Linear algebra

Before we take up the proof of Theorem 1.1, we consider the analogue in linear algebra in order to find the right strategy. Let V be a finite-dimensional $\mathbf{R}$-vector space endowed with a non-degenerate quadratic form q with indefinite signature (n_+, n_-). Let B be the associated non-degenerate symmetric bilinear form on V. In an old handout on quadratic spaces, we saw that although there is a decomposition $V = V_+ \oplus V_-$ into B-perpendicular subspaces such that q is positive-definite on V_+ and negative-definite on V_-, such a decomposition is *not* unique and generally depends on choices of certain bases. Hence, it cannot be invoked on the level of vector bundles without some way to make it "canonical". But there is a way to do this: we first fix an *auxiliary* inner product $\langle\cdot,\cdot\rangle$ on V. (The bundle analogue is to fix an auxiliary Riemannian metric, which we know can always be done, thanks to partitions of unity.) With this choice, we can write $B = \langle T(\cdot),\cdot\rangle$ for a unique self-adjoint $T : V \simeq V$, and then the spectral theorem provides an orthogonal eigenspace decomposition $V = \oplus_\lambda V_\lambda$ over some nonzero eigenvalues.

In terms of an eigenbasis we get a "diagonalization" of B, and more specifically of q, with coefficients of the quadratic form given by these eigenvalues taken with multiplicity equal to the dimension of the corresponding eigenspaces. Taking $V_+ = \oplus_{\lambda>0} V_\lambda$ and $V_- = \oplus_{\lambda<0} V_\lambda$, we obtain a direct sum decomposition $V = V_+ \oplus V_-$ that is orthogonal for both $\langle\cdot,\cdot\rangle$ and B, with q having positive-definite restriction to V_+ and negative-definite restriction to V_-. In particular, $\dim V_\pm = n_\pm$. This is a method of construction of the desired decomposition of V that requires no choices after we pick the auxiliary inner product. Hence, it seems promising that this approach might work at the level of vector bundles. Of course, there is a technical complication: the self-adjoint T will now be replaced with a bundle mapping that is fibrally self-adjoint with respect to the chosen auxiliary Riemannian metric and so the fibral eigenvalues will "move" (since we have replaced a single matrix with a "varying family" of matrices).

In the case that an eigenvalue in a fiber is a simple root of the characteristic polynomial on that fiber, then on the nearby fibers there is a unique eigenvalue near this one, it is simple, and it has C^p dependence in m (see Lemma 3.1). However, there is also the possibility of fibral eigenvalues with higher multiplicity. In such cases, dimensions of the fibral eigenspaces can "jump" is discontinuous ways, so individual fibral eigenspaces cannot generally fit together to make subbundles. But we need much less, just that the direct sum of the eigenspaces for eigenvalues with a fixed sign fit together to make subbundles. This is something that we will be able to construct as an application of properties of Grassmann manifolds.

3. Proof of Theorem 1.1

We now prove Theorem 1.1, inspired by our observations in the case of linear algebra (i.e., a 1-point base space). We choose an auxiliary C^p Riemannian metric $\langle\cdot,\cdot\rangle$ on $E \to M$. This gives an identification $E \simeq E^\vee$ that is the "inner product" mapping on fibers. This yields a bundle isomorphism

$$E^\vee \otimes E^\vee \simeq E \otimes E^\vee \simeq \mathrm{Hom}(E, E),$$

where the final step is uniquely characterized by the condition that it induces the natural isomorphism $E(m) \otimes E(m)^\vee \simeq \mathrm{Hom}(E(m), E(m))$ on fibers (and the smoothness of this final step is readily checked by calculation in local trivializing frames and dual frames for E over M). We may view B as a global section of $E^\vee \otimes E^\vee$, so our composite bundle isomorphism carries it to an element $T \in (\mathrm{Hom}(E, E))(M) = \mathrm{Hom}_M(E, E)$.

On fibers, $T|_m \in \mathrm{Hom}(E(m), E(m))$ likewise corresponds (via $\langle\cdot,\cdot\rangle_m$) to the non-degenerate symmetric bilinear form $B(m) \in E(m)^\vee \otimes E(m)^\vee$ with signature (n_+, n_-). That is, by chasing elementary tensors (on fibers) we see that $B(m) = \langle \cdot, T|_m(\cdot)\rangle_m$ as bilinear forms on $E(m)$, so $T|_m$ is the self-adjoint automorphism of $E(m)$ corresponding to $B(m)$ in the sense of our preceding motivational digression into linear algebra. In particular, $E(m)$ has a canonical decomposition $E(m) = E(m)^+ \oplus E(m)^-$ where $E(m)^+$ is the n_+-dimensional span of the positive eigenspaces for $T|_m$ and $E(m)^-$ is the n_--dimensional span of the negative eigenspaces for $T|_m$. These are $B(m)$-orthogonal, with $B(m)$ having positive-definite restriction to $E(m)^+$ and negative-definite restriction to $E(m)^-$. Our problem is now reduced to proving that the fibral subspaces $\{E(m)^\pm\}_{m\in M}$ in the $E(m)$'s with constant dimension $n_\pm$ fit together into the fibers of a C^p subbundle $E^\pm$ in E with constant rank $n_\pm$, as then the natural mapping $E^+ \oplus E^- \to E$ is an isomorphism on fibers and hence an isomorphism of C^p vector bundles (thereby solving our problem). Since $E(m)^-$ is the $\langle\cdot,\cdot\rangle_m$-orthogonal complement of $E(m)^+$ for each $m \in M$, if we can build a C^p subbundle E^+ gluing the $E(m)^+$'s then we can define E^- to be the orthogonal complement subbundle $(E^+)^\perp$ in E with respect to $\langle\cdot,\cdot\rangle$; this is the subbundle whose fibers are the orthogonal complements of $E^+(m)$ in $E(m)$ for all $m \in M$. (See the handout on operations with pseudo-Riemannian metrics for the verification that formation of orthogonal complement on fibers behaves well at the level of vector bundles.) Hence, we may now and do now focus our attention only on the problem of building a C^p subbundle E^+ in E with m-fiber $E(m)^+$ for every $m \in M$. We are going to build E^+ by creating a C^p map to a suitable Grassmannian manifold.

By Lemma 2.1 in the handout on subbundles and quotient bundles, the C^p subbundle E^+ is uniquely determined by its fibers if it exists, and so in view of Theorem 2.5 in that handout the problem is entirely one of proving that the union of these fibers is a closed C^p submanifold of E. This problem is local over M, and so we may work locally over M. In particular, we can assume that $E \to M$ is trivial with some rank $n > 0$. Choosing a trivializing frame $\{s_1, \dots, s_n\}$ allows us to describe the bundle mapping $T : E \simeq E$ in terms of a matrix (a_{ij}) with $a_{ij} \in C^p(M)$. With such a trivialization fixed, for each $m \in M$ we have a preferred quotient mapping $\mathbf{R}^n \simeq E(m) \twoheadrightarrow E(m)/E(m)^- \simeq E(m)^+$ with dimension n_+, so we get a point in $\mathbf{G}_{n_+}(\mathbf{R}^n)$. Our problem is *exactly* to prove that this set-theoretic mapping $M \to \mathbf{G}_{n_+}(\mathbf{R}^n)$ is C^p (as then the pullback of the universal subbundle of rank $n - n_+ = n_-$ over the Grassmannian will provide the C^p subbundle E^- in $M \times \mathbf{R}^n \simeq E$, and hence its orthogonal complement in E will be E^+).

We want to reduce our problem to the special case $n_+ = 1$. We will achieve this by using a suitable exterior power construction. Since the a_{ij}'s are continuous on M and we are allowed to work locally, the lemma on continuity of roots (Lemma 4.2 in the handout on quadratic spaces) allows us to assume that (in the sense of multiplicity) the n_+ positive eigenvalues of the diagonalizable $T|_m$'s are concentrated near specific positive numbers, and likewise (in the sense of multiplicity) for the n_- negative eigenvalues of the $T|_m$'s. By Lagrange interpolation, we can find a monic polynomial P such that P carries these positive eigenvalues to very large positive numbers and carries the negative eigenvalues to negative numbers very close to 0. By diagonalizability of the $T|_m$'s, it follows that upon replacing T with $P(T)$ we do not affect the subspaces $E(m)^+$ or $E(m)^-$ in each $E(m)$ but we have arranged that the negative eigenvalues of the $T|_m$'s are all very close to 0 and the positive eigenvalues are very large. Hence, $\wedge^{n_+}(T)$ has induced fiber maps with one (in the sense of multiplicity!) very large positive eigenvalue near some large positive number λ_0 and the rest bounded away from it from above, say less than $\lambda_0/2$.

Note that $\wedge^{n_+}(E)$ admits an induced C^p Riemannian metric for which $\wedge^{n_+}(T)$ is fibrally self-adjoint, and the fibral eigenline for the largest fibral eigenvalue is precisely the line $\wedge^{n_+}(E(m)^+)$ in $\wedge^{n_+}(E(m))$. In particular, the quotient mapping $E(m) \twoheadrightarrow E(m)/E(m)^- \simeq E(m)^+$ with n_+-dimensional target induces a 1-dimensional quotient mapping $\wedge^{n_+}(E(m)) \twoheadrightarrow \wedge^{n_+}(E(m)^+)$. By Theorem 4.1 in the old gluing handout, the "exterior power" mapping $\mathbf{G}_{n_+}(\mathbf{R}^n) \to \mathbf{P}(\wedge^{n_+}(\mathbf{R}^n))$ is a *closed embedding* of smooth manifolds. (The argument there only discusses topological aspects, but the method of proof yields the stronger results concerning the differentiable structures, in view of the standard charts for the C^∞ manifold structures.) In particular, by mapping properties of embedded submanifolds, our set-theoretic mapping $M \to \mathbf{G}_{n_+}(\mathbf{R}^n)$ is C^p if and only if its composite with the closed embedding of $\mathbf{G}_{n_+}(\mathbf{R}^n)$ into $\mathbf{P}(\wedge^{n_+}(\mathbf{R}^n))$ is C^p. But this composite is just an instance of our general problem in the special case $n_+ = 1$, namely for the bundle $\wedge^{n_+}(E)$ equipped with its induced Riemannian metric and fibrally self-adjoint bundle endomorphism $\wedge^{n_+}(T) - (\lambda_0/2)\mathrm{id}$! Thus, we now may and do assume $n_+ = 1$.

Our situation is now that there is (in the sense of multiplicity) a *unique* positive eigenvalue $\lambda(m)$ for $T|_m$ on each fiber $E(m)$, and we want to fit these to make a C^p line subbundle of E (with orthogonal complement then giving the hyperplane subbundle whose fibers are the negative eigenspaces). Since $\lambda(m)$ is a *simple* root of the characteristic polynomial, the following lemma (applied to M and the characteristic polynomial $\Lambda^n + a_{n-1}(m)\Lambda^{n-1} + \cdots + a_0(m)$ of the $T|_m$'s, with $a_i \in C^p(M)$) ensures that $\lambda(m)$ is a C^p function of $m \in M$:

Lemma 3.1. *Let X be a C^p premanifold with corners, $0 \le p \le \infty$, and let $f(x,t) = t^n + a_{n-1}(x)t^{n-1} + \cdots + a_0(x) \in C^p(X)[t]$ be a monic polynomial with coefficients in $C^p(X)$. Assume for some $x_0 \in X$ that $f(x_0,t) \in \mathbf{R}[t]$ has a simple root ρ_0. There exists $\varepsilon > 0$ and an open neighborhood U of x_0 in X such that for all $x \in U$ the polynomial $f(x,t) \in \mathbf{R}[t]$ has a unique root $\rho(x)$ in $(\rho_0 - \varepsilon, \rho_0 + \varepsilon)$ with $\rho(x)$ a simple root of $f(x,t)$. Moreover, $\rho : U \to \mathbf{R}$ is a C^p function.*

Proof. In order to handle the case $p = 0$, for which there is no implicit function theorem available, we first pass to a universal situation in the C^∞ case. Let $Y = \mathbf{R}^n$ and let F be the "universal monic degree-n polynomial" over Y: we define $F(y,t) = t^n + c_{n-1}t^{n-1} + \cdots + c_0$ where $y = (c_0, \dots, c_{n-1}) \in \mathbf{R}^n$. The $a_i \in C^p(X)$ define a C^p mapping $\phi : X \to Y$ carrying x_0 to some $y_0 \in Y$, and $F(\phi(x),t) = f(x,t)$ for all $x \in X$. In particular, $(\partial_t F)(y_0,t) = (\partial_t f)(x_0,t)$ is nonzero at $t = \rho_0$, so $F(y_0,t)$ has a simple root at ρ_0. Thus, if we can solve the problem in the C^∞ setting for F around y_0 and the simple root ρ_0, then composition with the C^p map $\phi : X \to Y$ gives a solution for f around x_0 of the desired type. Hence, we may now focus on the "universal" F over Y in the C^∞ setting.

Consider the smooth function $Y \times \mathbf{R} \to \mathbf{R}$ given by F. That is, we consider the function $h : \mathbf{R}^{n+1} \to \mathbf{R}$ defined by $(c_0, \dots, c_{n-1}, t) \mapsto t^n + c_{n-1}t^{n-1} + \cdots + c_0$. We are interested in the equation $h(\underline{c}, t) = 0$ as an implicit C^∞ equation for t in the c_i's. The assumption that ρ_0 is a simple root for $h(y_0, t)$ with some $y_0 \in Y = \mathbf{R}^n$ is exactly the condition under which the C^∞ implicit function theorem kicks in, and gives exactly the desired uniqueness of the solution to $h(y,t) = 0$ with t near ρ_0 for y near y_0, together with the smooth dependence of this solution on y. ■

Let $\lambda : M \to \mathbf{R}$ denote the C^p function extracting the unique positive eigenvalue of $T|_m$, so

$$T - \lambda\mathrm{id}_E : E \to E$$

is a C^p bundle mapping whose fibral kernel is of constant dimension 1. By Theorem 2.6 from the handout on subbundles and quotient bundles, there exists a C^p line subbundle of E whose m-fiber is the kernel of $T|_m - \lambda(m)\mathrm{id}_{E(m)}$, and this is exactly the C^p subbundle E^+ we have sought to build in this special case.

4. A topological theorem

We now turn to the proof of the serious topological input in the proof of Corollary 1.2 above. Let X be a C^p premanifold with corners, and assume that all closed paths in X (i.e., continuous maps $[0,1] \to X$ sending 0 and 1 to the same point) can be continuously contracted to a point. That is, for any closed path $\sigma : [0,1] \to X$ we assume that there exists a continuous map $\Phi : [0,1] \times [0,1] \to X$ with $\Phi(0,\cdot) = \sigma$ and $\Phi(1,\cdot)$ a constant map to a point $x_0 \in X$ (so the maps $\sigma_z = \Phi(z,\cdot)$ for $z \in [0,1]$ are to be viewed as a "continuous deformation" of σ to the constant map to x_0). For example, any finite-dimensional vector space V can be continuously contracted to a point in finite time, using $[0,1] \times V \to V$ defined by $(t,v) \mapsto (1-t)v$, and the same recipe applies to continuous contract all closed paths in V to the origin.

Somewhat more interesting (but not a surprise) is that all closed paths in S^m can be continuously contracted to a point if $m \geq 2$ (visualize the case $m = 2$). The easy case of the proof is when the image of the path $\sigma : [0,1] \to S^m$ omits some point ξ, for then since $S^m - \{\xi\}$ is homeomorphic to $\mathbf{R}^{m-1}$ and so we win by using the result for closed paths in vector spaces. The problem is therefore to continuously deform the closed path so that its image is not the entire sphere; beware of space-filling curves. But a continuous map $\sigma : [0,1] \to S^m$ is *uniformly continuous* (using whatever metric you like on S^m), so it is not difficult to find a piecewise smooth closed path that is uniformly close to σ (using "straight lines" in some coordinate charts). By considering flow along secants linking two parametric paths with the same time interval, one then gets a continuous deformation from σ to a piecewise smooth path. Piecewise smooth paths cannot be surjective since their images are a finite union of images of smooth paths and the image of a smooth map from an interval into an m-dimensional manifold has image with measure zero when $m > 1$.

Theorem 4.1. *Let X be a C^p premanifold with corners, $0 \leq p \leq \infty$, and assume that all closed paths in X are continuously contractible to a point. Every C^p line bundle $L \to X$ is trivial.*

Our proof will use one ingredient for which we shall refer the reader to topology books to save space; the result we require from those books is not at all deep, and the proof could have been included, but it is a slightly involved argument and so we will just refer to standard texts at the suitable time.

Remark 4.2. The proof of Theorem 4.1 may look like a series of miraculous tricks. In fact, the methods we use are special cases of general principles in algebraic topology (though seeing them used only in our special situation may tend to mask some of the underlying structure). What actually happened is that I translated a 2-line proof using sheaf cohomology into more elementary language, and that is how I cooked up the argument below.

To prove Theorem 4.1, we may and do assume X is connected. Let $\{U_i\}$ be an open cover such that $L|_{U_i}$ is trivial for all i. Choose a trivializing section $s_i \in L(U_i)$, so this is a nowhere-vanishing section. Hence, $s_i|_{U_i \cap U_j}$ and $s_j|_{U_i \cap U_j}$ are nowhere-vanishing sections of the line bundle $L|_{U_i \cap U_j}$, whence $s_i|_{U_i \cap U_j} = f_{ij} s_j|_{U_i \cap U_j}$ for some non-vanishing smooth function f_{ij} on $U_i \cap U_j$. Observe that we have the *cocycle relation* $f_{ij} f_{jk} = f_{ik}$ on $U_i \cap U_j \cap U_k$ since on this triple overlap we have $s_i = f_{ij} s_j = f_{ij} f_{jk} s_k$. The proof will consist of two parts: an algebraic part that tracks signs for the values of the non-vanishing functions f_{ij}, and a geometric part that uses connectivity considerations and the contractibility hypothesis on closed paths in X (!) to infer some properties of these signs.

We must exercise some caution, since $U_i \cap U_j$ may be disconnected even if the U_i's are connected. Thus, f_{ij} may not have constant sign on $U_i \cap U_j$. Let $\varepsilon_{ij} : U_i \cap U_j \to \{\pm 1\}$ be the locally constant function (constant on connected components of $U_i \cap U_j$) that encodes the sign of f_{ij} (i.e., for $x \in U_i \cap U_j$, $\varepsilon_{ij}(x) = \pm 1$ according as $f_{ij}(x) \in \mathbf{R} - \{0\}$ is positive or negative). The cocycle relation on the f_{ij}'s implies the same for their signs given by the ε_{ij}'s: on $U_i \cap U_j \cap U_k$, we have an equality of locally constant functions $\varepsilon_{ik} = \varepsilon_{ij} \varepsilon_{jk}$. Note that $\varepsilon_{ii} : U_i \to \{\pm 1\}$ is the constant function 1, and $\varepsilon_{ij} \varepsilon_{ji} = 1$.

The crucial part of the proof, and the only place where the contractibility assumption is used, is the fact that the locally constant functions $\varepsilon_{ij} : U_i \cap U_j \to \{\pm 1\}$ that satisfy the cocycle relation on triple overlaps have a special form:

Lemma 4.3. *There exist locally constant functions* $\varepsilon_i : U_i \to \{\pm 1\}$ *such that* $\varepsilon_{ij} = \varepsilon_i \varepsilon_j^{-1}$ *on* $U_i \cap U_j$ *for all* i, j.

Let us grant the lemma (which contains a beautiful geometric construction in its proof), and see why it implies the main result. Then we will prove the lemma.

Consider the trivializing section $s_i' = \varepsilon_i^{-1} s_i \in L(U_i)$ for $L|_{U_i}$. On the overlaps $U_i \cap U_j$ we have

$$s_i'|_{U_i \cap U_j} = \varepsilon_{ij}^{-1} \varepsilon_j^{-1} s_i|_{U_i \cap U_j} = \varepsilon_{ij}^{-1} f_{ij} s_j'|_{U_i \cap U_j}.$$

But the ε_{ij}'s were rigged to have the same pointwise sign as the f_{ij}'s at each point of the double overlaps $U_i \cap U_j$, so the multiplier functions $f_{ij}' = \varepsilon_{ij}^{-1} f_{ij}$ on $U_i \cap U_j$ that express the transition formula $s_i'|_{U_i \cap U_j} = f_{ij}' s_j'|_{U_i \cap U_j}$ are *everywhere positive* on $U_i \cap U_j$. In other words, we have built a collection of local trivializing sections that define consistent orientations on the fiber lines over the double overlaps. Hence, $L \to X$ is *orientable*. We may therefore choose an orientation, and upon picking a C^p Riemannian metric on L we get a unique section in $L(X)$ that is fiberwise positive (in the orientation sense) with length 1 in each fiber, and this is a nowhere-vanishing element of $L(X)$. Hence, we have trivialized the line bundle L over X.

5. Proof of Lemma 4.3

The idea is to create a space whose connectivity properties encode information concerning the "signs" ε_{ij} that satisfy the cocycle condition. This is a special case of a general technique in topology.

For each i, let $\widetilde{U}_i = U_i \times \{\pm 1\}$; this is a disjoint union of two copies of U_i labelled by ± 1. Let $\pi_1 : \widetilde{U}_i \to U_i$ be the projection. We wish to glue $\widetilde{U}_i$ and $\widetilde{U}_j$ over $U_i \cap U_j$ via the following isomorphism:

$$\phi_{ij} : \pi_i^{-1}(U_i \cap U_j) = (U_i \cap U_j) \times \{\pm 1\} \simeq (U_i \cap U_j) \times \{\pm 1\} = \pi_j^{-1}(U_i \cap U_j)$$

where the middle isomorphism is

$$(u, e) \mapsto (u, \varepsilon_{ij}(u) e).$$

Here is where the cocycle relation works its magic: since $\varepsilon_{ij} \varepsilon_{jk} = \varepsilon_{ik}$ on $U_i \cap U_j \cap U_k$, we deduce the "gluing equation" $\phi_{ij} \circ \phi_{jk} = \phi_{ik}$ as maps from $\pi_i^{-1}(U_i \cap U_j \cap U_k)$ to $\pi_k^{-1}(U_i \cap U_j \cap U_k)$. (In particular, ϕ_{ii} is the identity on $\pi_i^{-1}(U_i) = \widetilde{U}_i$ and ϕ_{ij} and inverse to ϕ_{ji}.)

We may now glue the $\widetilde{U}_i$'s as follows. Let $Y = \coprod \widetilde{U}_i$ be the disjoint union of these spaces, and impose the following equivalence relation: $y \sim y'$ if and only if $y \in \widetilde{U}_i$ and $y' \in \widetilde{U}_j$ for some i and j with $y' = \phi_{ij}(y)$. The above "gluing equation" for the ϕ's is exactly the assertion that $\sim$ is transitive, and the reflexivity and identity properties are merely the properties that ϕ_{ij} and inverse to ϕ_{ji} and that ϕ_{ii} is the identity for all i. Thus, $\widetilde{X} = Y/\sim$ makes sense, and each natural map $h_i : \widetilde{U}_i \to \widetilde{X}$ is an injective map whose image meets the image of $\widetilde{U}_j$ in exactly the subsets $\pi_i^{-1}(U_i \cap U_j)$ and $\pi_j^{-1}(U_i \cap U_j)$, with the composite bijection

$$\pi_i^{-1}(U_i \cap U_j) \simeq h_i(\widetilde{U}_i) \cap h_j(\widetilde{U}_j) \simeq \pi_j^{-1}(U_i \cap U_j)$$

given by *exactly* ϕ_{ij}.

Put in more explicit terms, in the set $\widetilde{X}$ we may say that $\widetilde{U}_i$ and $\widetilde{U}_j$ meet along the respective subsets $\pi_i^{-1}(U_i \cap U_j)$ and $\pi_j^{-1}(U_i \cap U_j)$ via the identification ϕ_{ij}. In view of the simple formula for

ϕ_{ij}, it is very easy to check the gluing conditions for topologies, so $\widetilde{X}$ admits a unique topology that makes each subset $\widetilde{U}_i$ open and induces on it the topology it has as $U_i \times \{\pm 1\}$. Moreover, we can run through the same argument with the differentiable structure, so $\widetilde{X}$ has a unique structure of C^p premanifold with corners, recovering the evident structure of this type on each open subset $\widetilde{U}_i = U_i \times \{\pm 1\}$. The maps $\pi_i : \widetilde{U}_i \to U_i \subseteq X$ "agree on overlaps in $\widetilde{X}$" in this manner, and so they glue to define a set-theoretic map $\pi : \widetilde{X} \to X$ whose restriction over $U_i \subseteq X$ is identified with $\pi_i : \widetilde{U}_i = U_i \times \{\pm 1\} \to U_i$, so π is a local isomorphism.

To summarize, $\pi : \widetilde{X} \to X$ is a "gluing" of the maps $\pi_i : \widetilde{U}_i \to U_i$, with the distinction that whereas each $\widetilde{U}_i = U_i \times \{\pm 1\}$ is trivially disconnected (even if U_i is connected), it is less evident if $\widetilde{X}$ is connected or not. The map π is an example of a degree-2 covering map: it is surjective with fibers of size 2, and (this is the key) over some opens that cover the target (such as the U_i's) it restricts to a "split covering" that is given by a product of the open target and a finite set of size 2 (in this case, $\pi^{-1}(U_i) \simeq \widetilde{U}_i = U_i \times \{\pm 1\}$ over U_i).

In fact, we claim that the C^p premanifold with corners $\widetilde{X}$ *must* be disconnected. Let us first see why this suffices to complete the proof of the lemma. Consider a non-trivial separation $\widetilde{X} = U \coprod V$ with nonempty open and closed sets U and V in $\widetilde{X}$. The map $\pi : \widetilde{X} \to X$ is both open and closed (as over each $U_i \subseteq X$ the restriction $\pi_i : \widetilde{U}_i = U_i \times \{\pm 1\} \to U_i$ is clearly open and closed), so connectivity of X forces the non-empty U and V to each surject onto X (as each has non-empty open and closed image). But U and V are disjoint and $\pi : \widetilde{X} \to X$ has all fibers of size exactly 2, so the surjective map $U \to X$ must be bijective. Since π is a local C^p isomorphism, so is its bijective restriction to the open subset U in $\widetilde{X}$, whence the restriction of π to U is a C^p-isomorphism onto X. Consider the inverse $g : X \simeq U \subseteq \widetilde{X}$. On each U_i, this is a C^p section $g_i : U_i \to U_i \times \{\pm 1\}$ to the map π_i, and so this section must be $u_i \mapsto (u_i, \varepsilon_i(u_i))$ for a locally constant function $\varepsilon_i : U_i \to \{\pm 1\}$. Since $g_i|_{U_i \cap U_j} = g|_{U_i \cap U_j} = g_j|_{U_i \cap U_j}$, it follows from the role of the ε_{ij}'s in the definition of the gluing data ϕ_{ij} used to construct $\widetilde{X}$ that these ε_i's satisfy the required identities with respect to the ε_{ij}'s on the double overlaps $U_i \cap U_j$.

It remains to prove that $\widetilde{X}$ is disconnected. We assume otherwise, so it is path-connected. Choose $x_0 \in X$, so $\pi^{-1}(x_0) = \{x, x'\}$ for distinct points $x, x' \in \widetilde{X}$. Let $\widetilde{\sigma} : [0,1] \to \widetilde{X}$ be a continuous path with $\widetilde{\sigma}(0) = x$ and $\widetilde{\sigma}(1) = x'$. Consider the continuous path $\sigma = \pi \circ \widetilde{\sigma} : [0,1] \to X$ given by projecting $\widetilde{\sigma}$ into X. Since $\pi(x) = x_0 = \pi(x')$, σ is a closed path in X. Hence, it can be continuously contracted to a constant path in X. We seek to deduce a contradiction from this property. We first make a definition that isolates an important property of π:

Definition 5.1. A continuous map of non-empty topological spaces $f : Y' \to Y$ is a *covering map* if Y has a covering by non-empty opens U_i such that there are homeomorphisms $f^{-1}(U_i) \simeq U_i \times S_i$ over U_i for a discrete non-empty set S_i. That is, $f^{-1}(U_i)$ is a disjoint (non-empty) union of open and closed subsets on which the restriction of f is a homeomorphism onto U_i. If all fibers $f^{-1}(y)$ has size n, we call f a *degree-n covering map*.

Example 5.2. The map π is a degree-2 covering map. An important example of a degree-n covering map is the nth-power map $\mathbf{C}^\times \to \mathbf{C}^\times$ given by $z \mapsto z^n$. This latter example works because for each of the n distinct nth roots $w_1, \ldots, w_n$ of each nonzero $z_0 \in \mathbf{C}$, each z near z_0 admits a unique nth root near each w_j, and these local nth-roots are continuous functions that "split" $f^{-1}(D)$ for a small disc D around z_0.

The desired contradiction follows from:

Theorem 5.3. *Let $f : Y' \to Y$ be a covering map of topological spaces, and let $\sigma' : [0,1] \to Y'$ be a continuous map such that $\sigma = f \circ \sigma' : [0,1] \to Y$ is a closed path. If σ can be continuously contracted to a point in Y then σ' must be a closed path: $\sigma'(0) = \sigma'(1)$.*

Proof. Let $\Phi : [0,1] \times [0,1] \to Y$ be the continuous map that encodes how to shrink σ to a point. That is, $\Phi(0,t) = \sigma(t)$ for all t, $\Phi(1,t) \in Y$ is a fixed point independent of t, and $\Phi(z,0) = \Phi(z,1)$ for all z. We write $\sigma_z : [0,1] \to Y$ to denote $\Phi(z,\cdot)$ for $z \in [0,1]$, so each σ_z is a closed path ($\sigma_z(0) = \sigma_z(1)$) and we visualize $\{\sigma_z\}_{z\in[0,1]}$ as a 1-parameter family of paths in Y that describe how σ is continuously deformed to a point in Y. We are given a continuous lifting $\sigma'_0 = \sigma'$ of $\sigma_0 = \sigma$ to Y' through f.

The hardest step in the proof is this: there exists a (unique) continuous $\Phi' : [0,1] \times [0,1] \to Y'$ lifting Φ (in that $f \circ \Phi' = \Phi$) with $\Phi'(0,\cdot) = \sigma'$. This follows from the important *homotopy lifting lemma* from topology, which we discuss after the proof. Using Φ', we get a continuous 1-parameter family of paths $\{\sigma'_z\}_{z\in[0,1]}$ lifting $\{\sigma_z\}$. This is proved in Theorem 5.4 below. Using Φ', or rather than paths σ'_z, consider the two continuous "paths of endpoints" $c_0 : z \mapsto \sigma'_z(0) = \Phi'(z,0)$ and $c_1 : z \mapsto \sigma'_z(1) = \Phi'(z,1)$ in Y'. For each z we have

$$f(c_0(z)) = f(\sigma'_z(0)) = \sigma_z(0) = \sigma_z(1) = f(\sigma'_z(1)) = f(c_1(z)),$$

so c_0 and c_1 lift the same path in Y. Moreover, the initial endpoints $c_0(0)$ and $c_1(0)$ are given by $c_0(0) = \Phi'(0,0) = \sigma'(0)$ and $c_1(0) = \Phi'(0,1) = \sigma'(1)$ since $\Phi'(0,\cdot) = \sigma'$, so these are the points we want to prove are equal. To prove such equality, we first explain why the final endpoints $c_0(1) = \sigma'_1(0)$ and $c_1(1) = \sigma'_1(1)$ of c_0 and c_1 are equal. Note that $f \circ \sigma'_1 = \sigma_1$ is a *constant* map (!) with image equal to some point $y \in Y$, so $\sigma'_1([0,1])$ is contained in the fiber $f^{-1}(y)$ that is a discrete space. But $\sigma'_1([0,1])$ is connected, so as a non-empty subset of a discrete space it must consist of a single point. That is, $\sigma'_1(t)$ is independent of t. In particular, $c_0(1) = c_1(1)$.

The two paths $c_0, c_1 : [0,1] \rightrightarrows Y'$ project to the same path $c = f \circ c_0 = f \circ c_1$ in Y and satisfy $c_0(1) = c_1(1)$. From this we shall now show that $c_0 = c_1$, so in particular $c_0(0) = c_1(0)$, as desired. Let $J \subseteq [0,1]$ be the subset of points $z \in [0,1]$ such that $c_0|_{[z,1]} = c_1|_{[z,1]}$, so $1 \in J$. Thus, J is a subinterval of $[0,1]$ containing 1. Consider $z \geq \inf(J)$. Let $U \subseteq Y$ be an open set around $c(z)$ over which $f^{-1}(U)$ is "split": a disjoint union of copies of U, say $f^{-1}(U) \simeq U \times S$ over U for a discrete set S. By continuity of $c : [0,1] \to Y$, for for $\varepsilon > 0$ we have $c(I) \subseteq U$ for the interval $I = (z-\varepsilon, z+\varepsilon) \cap [0,1]$, so c_0 and c_1 carry I into $f^{-1}(U) \simeq U \times S$. The images $c_0(I)$ and $c_1(I)$ are connected, and so lie in $U \times \{s_0\}$ and $U \times \{s_1\}$ for some unique $s_0, s_1 \in S$. But $c_0(z) = c_1(z)$ if $z \in J$ (e.g., if $z = 1$) and $c_0(z') = c_1(z')$ for $z' \in (z, z+\varepsilon) \cap [0,1] \subseteq J$ if $z < 1$. Either way, $c_0(I)$ meets $c_0(I)$, so $s_0 = s_1$. Hence, $c_0|_I$ and $c_0|_I$ are continuous maps from I into $U \times \{s_0\}$ such that their composites with f coincide. But f restricts to the "identity" from $U \times \{s_0\}$ onto U, so $c_0|_I = c_1|_I$. In view of the definition of I and the hypotheses on z, this forces $(z-\varepsilon, z] \cap [0,1] \subseteq J$. Hence, we conclude that J cannot have a positive infimum and that it contains its infimum. Thus, $J = [0,1]$, so $c_0(0) = c_1(0)$ as desired. ■

It remains to explain how to construct $\Phi' : [0,1] \times [0,1] \to Y'$ as used above. We first explain why a continuous lifting Φ' of Φ satisfying $\Phi'(0,0) = \sigma'(0)$ *must* satisfy $\Phi'(0,\cdot) = \sigma'$, so we may therefore ignore σ' (aside from the specification of $\Phi'(0,0) \in f^{-1}(\Phi(0,0))$) in the construction of Φ'. The continuous paths $\Phi'(0,\cdot)$ and σ' in Y' project to the same path $\Phi(0,\cdot) = \sigma$ in Y and coincide at $t = 0$, and we wish to get equality on $[0,1]$. This proceeds by *exactly* the same "interval" argument that we just used above to prove $c_0 = c_1$ given that $c_0(1) = c_1(1)$ and $f \circ c_0 = f \circ c_1$; the only difference is that instead of working across $[0,1]$ from right to left (using an infimum), we go from left to right (using a supremum). Hence, it remains to prove:

Theorem 5.4. *Let $f : Y' \to Y$ be a covering map of topological spaces, and $\Phi : [0,1] \times [0,1] \to Y$ a continuous map. Choose $y_0' \in Y'$ lying over $y_0 = \Phi(0,0)$. There exists a unique continuous map $\Phi' : [0,1] \times [0,1] \to Y'$ satisfying $\Phi'(0,0) = y_0'$ and $f \circ \Phi' = \Phi$.*

This is the important *homotopy lifting lemma* from topology, and a proof can be found in any reasonable introductory topology book. For example, see Lemmas 4.1 and 4.2 in section 8.4 of Munkres' *Topology* for one version of the proof. (Munkres lifts the map across small squares in $[0,1] \times [0,1]$ using the existence of local sections to the covering map, and he uses compact/connectedness to work his way from the lower left corner across the entire domain. Another approach is to first proving the lifting lemma for $[0,1]$, then continuously lift across the bottom edge followed by continuous lifts along all vertical directions, and then giving a direct proof that the resulting map is really continuous on the square.)