BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 7, Number 1, July 1982

THE INVERSE FUNCTION THEOREM OF NASH AND MOSER

BY RICHARD S. HAMILTON

Contents

Received by the editors November 15, 1981.
1980 *Mathematics Subject Classification.* Primary 58C15; Secondary 58C20, 58D05, 58G30.

0273-0979/82/0000-0335/$40.25

INTRODUCTION

The basic idea of this inverse function theorem was discovered by John Nash [**14**], who used it to prove his famous theorem on isometric embeddings of Riemannian manifolds. Jürgen Moser [**13**] fashioned it into an abstract theorem in functional analysis of wide applicability. Sergeraert [**20**] stated the theorem in terms of a category of maps between Fréchet spaces. Generalizations to implicit function theorems have been given by Kuranishi [**9**], Zehnder [**25**], and the author [**4**]. Applications have been made by Nash [**14**] and Jacobowitz [**8**] to isometric embeddings, by Kuranishi [**9**] to deformation of CR structures, by Moser and Zehnder [**24**] to small divisor problems, by Hörmander [**6**] to problems in gravitation, by Beale [**2**] to water waves, by Schaeffer [**18, 19**] to free boundary problems in electromagnetics, by Sergeraert [**22**] to catastrophe theory, and by the author [**5**] to foliations. These and many other examples show the power and versatility of the theorem.

We define a category of "tame" Fréchet spaces and "tame" nonlinear maps, which is essentially that of Sergeraert. The spaces carry an extra structure of a "grading", a sequence of norms $\| \ \|_n$ defining the topology. The tame condition on the Fréchet spaces guarantees that the norms satisfy some interpolation

properties. The maps are required to satisfy linear growth estimates $\|Pf\|_n \leqslant C\|f\|_{n+r}$ for a fixed r and all n. It is a surprising fact that this condition is satisfied by all nonlinear partial differential operators and by most of their inverses, including inverses of elliptic, parabolic, hyperbolic, and subelliptic operators. The Nash-Moser inverse function theorem says that if the derivatives $DP(f)h = k$ of an operator P in the category have solutions $VP(f)k = h$ in the category, then the operator P has a local inverse in the category.

We have spent a long time saying just what this means, hoping to guide the reader by many examples. We have tried to include enough material on the simple cases so that one may appeal directly to the theorem without the need to crank out extra estimates. But if these are needed, the examples included here may serve as a guide. There are by now many ways to prove the theorem, and we have chosen an algorithm that pleases us for the simplicity of its estimates and its similarity to Nash's original. Our applications in this paper are not chosen to be the most novel or the most famous, but to be instructive. Many can be proven without the Nash-Moser theorem; we hope they are easier with it.

Finally we draw the reader's attention to the counterexamples we have included, particularly a brilliant one of Lojaciewicz and Zehnder [**12**], of which we give our own version. They show that all the extra hypotheses which are not needed in the simple Banach space case are really necessary in Fréchet spaces. Indeed if we weaken our hypothesis so little as to include maps satisfying $\|Pf\|_n < C\|f\|_{2n}$ we obtain a counterexample. Thus the Nash-Moser Theorem deftly picks out from among the wide variety of maps on Fréchet space a very useful class which we can invert.

PART I. CALCULUS ON FRÉCHET SPACES

I.1. Fréchet spaces.

1.1. *Definition of a Fréchet space.* A *seminorm* on a vector space F is a real-valued function $\|\ \|: F \to R$ such that

(i) $\|f\| \geqslant 0$ for all vectors f;

(ii) $\|f + g\| \leqslant \|f\| + \|g\|$ for all vectors f and g;

(iii) $\|cf\| = |c| \cdot \|f\|$ for all scalars c and vectors f.

A collection of seminorms $\{\|\ \|_n : n \in N\}$ defines a unique topology such that a sequence or net $f_j \to f$ if and only if $\|f_j - f\|_n \to 0$ for all $n \in N$.

A *locally convex topological vector space* is a vector space with a topology which arises from some collection of seminorms. The topology is *Hausdorff* if and only if $f = 0$ when all $\|f\|_n = 0$. The topology is *metrizable* if and only if it may be defined by a countable collection of seminorms $\{\|\ \|_n\}$. In this case we may always use sequences instead of nets. A sequence f_j is *Cauchy* if $\|f_j - f_k\|_n \to 0$ as j and $k \to \infty$ for all n. The space F is (sequentially) *complete* if every Cauchy sequence converges.

1.1.1. DEFINITION. A *Fréchet* space is a complete Hausdorff metrizable locally convex topological vector space.

1.1.2. EXAMPLE. Every Banach space is a Fréchet space. The collection of norms contains only one.

1.1.3. EXAMPLE. Let R^∞ be the vector space of all sequences $\{a_j\}$ of real numbers. Put

$$\|\{a_j\}\|_n = \sum_{j=0}^{n} |a_j| \quad \text{for } n = 0, 1, 2, \ldots.$$

Then R^∞ is a Fréchet space.

1.1.4. EXAMPLE. Let $\mathcal{C}^\infty[a, b]$ be the vector space of smooth functions on $a \leqslant x \leqslant b$. Put

$$\|f\|_n = \sum_{j=0}^{n} \sup_x |D^j f(x)| .$$

Then $\mathcal{C}^\infty[a, b]$ is a Fréchet space.

1.1.5. EXAMPLE. More generally, let X be a compact manifold and V a vector bundle over X. Let $\mathcal{C}^\infty(X, V)$ be the vector space of smooth sections of the bundle over X. Choose Riemannian metrics and connections on the bundles TX and V and let $D^j f$ denote the jth covariant derivative of a section f of V. Put

$$\|f\|_n = \sum_{j=0}^{n} \sup_x |D^j f(x)| .$$

Then $\mathcal{C}^\infty(X, V)$ is a Fréchet space.

1.1.6. EXAMPLE. Let $\mathcal{C}(R)$ denote the vector space of all continuous functions on the real line. Put

$$\|f\|_n = \sup\{|f(x)| : -n \leqslant x \leqslant n\}.$$

Then $\mathcal{C}(R)$ is a Fréchet space.

1.1.7. EXAMPLE. Let $\mathcal{H}$ denote the vector space of entire holomorphic functions. Put

$$\|f\|_n = \sup\{|f(z)| : |z| \leqslant n\}.$$

Then $\mathcal{H}$ is a Fréchet space.

1.1.8. COUNTEREXAMPLE. Let $\mathcal{C}_0(R)$ denote the vector space of continuous functions on the real line with compact support. For any positive function ρ let

$$\|f\|_\rho = \sup_x \rho(x)|f(x)| .$$

Then $\mathcal{C}_0(R)$ is a complete locally convex Hausdorff topological vector space, but it is *not* a Fréchet space because it is not metrizable. For given any countable sequence of ρ_j we can find ρ so that

$$\rho/\rho_j \to \infty \text{ as } x \to \pm\infty \text{ for all } j.$$

1.2. *Properties of Fréchet spaces.* A seminorm is a *norm* if $f = 0$ whenever $\|f\| = 0$. Some Fréchet spaces admit continuous norms and others admit none. If a Fréchet space admits one norm then all the seminorms in a collection defining the topology may be taken to be norms (by adding the one norm to them all).

1.2.1. EXAMPLES. A Banach space, $\mathcal{C}^\infty[a, b]$, $\mathcal{C}^\infty(X, V)$ for X compact and $\mathcal{H}$ all admit norms, while R^∞ and $\mathcal{C}(R)$ do not.

A closed subspace of a Fréchet space is also a Fréchet space, as is a quotient of a Fréchet space by a closed subspace. The direct sum of two Fréchet spaces is a Fréchet space. A space G is a (topological) direct summand of F if there exists a third space H such that F is isomorphic to the direct sum $G \oplus H$. If a Fréchet space admits a norm then so does any closed subspace, but a quotient by a closed subspace may not.

1.2.2. EXAMPLE. Let $\mathcal{C}^\infty[0, 2\pi]$ be the space of smooth functions on $0 \leqslant x \leqslant 2\pi$, and let $\mathcal{C}^\infty_{2\pi}$ be the space of smooth functions on $-\infty < x < \infty$ which are periodic with period 2π. We can regard $\mathcal{C}^\infty_{2\pi}$ as a closed subspace of $\mathcal{C}^\infty[0, 2\pi]$, where the inclusion map i is given by restricting the function to the interval. A function on the interval comes from a periodic function if and only if all the derivatives match up at 0 and 2π. Define the projection map p from $\mathcal{C}^\infty[0, 2\pi]$ to R^∞ by letting

$$a_j = D^j f(2\pi) - D^j f(0).$$

Then there is a short exact sequence

$$0 \to \mathcal{C}^\infty_{2\pi} \xrightarrow{i} \mathcal{C}^\infty[0, 2\pi] \xrightarrow{p} R^\infty \to 0$$

since a smooth function may have an arbitrary Taylor series. Thus R^∞ is a quotient of $\mathcal{C}^\infty[0, 2\pi]$ by a closed subspace. Since R^∞ does not admit a norm, it cannot be a direct summand. Hence the sequence does not split, and $\mathcal{C}^\infty_{2\pi}$ is a closed subspace of $\mathcal{C}^\infty[0, 2\pi]$ which is not a direct summand either.

The dual of a Banach space is again a Banach space. However, the dual of a Fréchet space which is not itself a Banach space is *never* a Fréchet space. Hence in general the space of linear maps of one Fréchet space to another will not form a third Fréchet space. For this reason we shall always *avoid* taking the space of linear maps. This causes some differences from the Banach space theory as it is usually presented.

1.2.3. EXAMPLES. The dual of the space R^∞ of all sequences is the space R^∞_0 of sequences with only a finite number of nonzero terms. The dual of the space of continuous functions on R, $\mathcal{C}(R)$, is the space $\mathcal{M}_0(R)$ of measures with compact support. The dual of the space $\mathcal{C}^\infty(X)$ of smooth functions on a compact manifold is the space $\mathcal{D}_\infty(X)$ of distributions. These are all complete locally convex topological vector spaces but none of these duals are Fréchet spaces.

The Hahn-Banach theorem holds for Fréchet spaces. Thus if F is a Fréchet space and f is a nonzero vector in F we can find a continuous linear functional l: $F \to R$ (or $\mathbf{C}$) such that $l(f) = 1$. Hence if $l(f) = l(g)$ for all continuous linear functionals then $f = g$. This is a very useful tool for reducing theorems to the real-valued case. Also the open mapping theorem holds. Thus if F and G are Fréchet spaces and if L: $F \to G$ is continuous, linear and invertible then L^{-1}: $G \to F$ is also continuous.

1.3. *Families of linear maps.* Let F, G, and H denote Fréchet spaces. A linear map L: $F \to G$ is naturally defined on all of F, while a nonlinear map P: $U \subseteq F \to G$ is naturally defined only on an open subset of F. Frequently it

happens that we have a map $L(f, g)$ which is nonlinear in f but linear separately in g. Thus

$$L(f, g_1 + g_2) = L(f, g_1) + L(f, g_2).$$

We call this a family of linear maps. It will naturally be defined for some $f \in F$ and all $g \in G$ with values in a third space H, so we write

$$L: (U \subseteq F) \times G \to H.$$

Since it is natural to think of $L(f)$ as a linear operator taking G into H, we shall usually write $L(f)g$ for $L(f, g)$, but it is important to emphasize that we want to consider L as a function from an open set of the product $F \times G$ into H, and *not* as a map

$$L: U \subseteq F \to L(G, H)$$

into the space of linear maps $L(G, H)$, which in general is *not* a Fréchet space.

Even in the Banach space case there is a large difference in what it means for $L: (U \subseteq F) \times G \to H$ to be continuous, as opposed to $L: U \subseteq F \to L(G, H)$.

1.3.1. EXAMPLE. Let $\mathcal{C}_{2\pi}$ be the space of continuous functions on $-\infty < x < \infty$ which are periodic with period 2π. Define a family of linear maps $L(t)f$ by translation, so that

$$\{L(t)f\}(x) = f(x + t).$$

Thus $L(t)$ is the linear operator "translation by t". Then $L: R \times \mathcal{C}_{2\pi} \to \mathcal{C}_{2\pi}$ is continuous (jointly in the product topology) but $L: R \to L(\mathcal{C}_{2\pi}, \mathcal{C}_{2\pi})$ is not continuous (in the topology of the Banach space of linear maps). It is easy to see why this should be so. For a given continuous periodic function we know by its uniform continuity that a small translation will produce a small variation in the function. But since the collection of all continuous functions is not equicontinuous, an arbitrarily small translation will produce an arbitrarily large variation in a function with a steep enough gradient.

Let $\mathcal{C}^1_{2\pi}$ denote the Banach space of continuously differentiable periodic functions with period 2π, and let L be defined as before. Then the map $L: R \to L(\mathcal{C}^1_{2\pi}, \mathcal{C}_{2\pi})$ is continuous, but $L: R \to L(\mathcal{C}^1_{2\pi}, \mathcal{C}^1_{2\pi})$ is not. This simple example is the source of many difficulties in applying the Banach space theory to diffeomorphism groups and free boundary value problems.

We shall also have to consider families of bilinear maps $B(f, g, h)$ which are nonlinear in f and linear separately in g and h. We shall write this usually as $B(f)\{g, h\}$ to indicate the bilinearity, and regard B as being continuous if it is continuous as a function on the product

$$B: (U \subseteq F) \times G \times H \to K$$

and *not* as a map into the space of bilinear maps

$$B: U \subseteq F \to L^2(G \times H, K).$$

I.2. The Riemann integral.

2.1. *The definite integral.* Let $f(t)$ be a continuous function on $a \leqslant t \leqslant b$ with values in the Fréchet space F. We wish to define the definite integral $\int_a^b f(t)\,dt$ as an element of F.

2.1.1. THEOREM. *There exists a unique element* $\int_a^b f(t)\,dt \in F$ *such that*
(i) *for every continuous linear functional* l: $F \to R$

$$l\left(\int_a^b f(t)\,dt\right) = \int_a^b l(f(t))\,dt.$$

In addition
(ii) *for every continuous seminorm* $\|\ \|$: $F \to R$,

$$\left\|\int_a^b f(t)\,dt\right\| \leqslant \int_a^b \|f(t)\|\,dt.$$

(iii) $\int_a^b f(t)\,dt + \int_b^c f(t)\,dt = \int_a^c f(t)\,dt$.
(iv) $\int_a^b [f(t) + g(t)]\,dt = \int_a^b f(t)\,dt + \int_a^b g(t)\,dt$.
(v) $\int_a^b cf(t)\,dt = c\int_a^b f(t)\,dt$.

PROOF. Let $\mathcal{C}([a, b], F)$ be the Fréchet space of all continuous functions on $a \leqslant t \leqslant b$ with values in F with the seminorms

$$\|f\|_n = \sup_t \|f(t)\|_n.$$

We say a function $f(t)$ is linear if $f(t) = tf_1 + f_2$ for some f_1 and $f_2 \in F$, and we say $f(t)$ is piecewise linear if it is continuous and there exists a partition $a = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_k = b$ such that $f(t)$ is linear on each piece $t_{i-1} \leqslant t \leqslant t_i$ for $1 \leqslant i \leqslant k$. The vector space $\mathrm{PL}([a, b], F)$ of piecewise linear functions on $[a, b]$ with values in F is a dense linear subspace of $\mathcal{C}([a, b], F)$. For a piecewise linear function we may define the integral by the trapezoidal rule

$$\int_a^b f(t)\,dt = \sum_{i=1}^{k} \frac{1}{2}[f(t_{i-1}) + f(t_i)](t_i - t_{i-1}).$$

We may easily verify properties (i)–(v) for piecewise linear functions directly from the formula. Since the integral defines a continuous linear functional on the dense subspace $\mathrm{PL}([a, b], R)$, it extends by continuity to a continuous linear functional on all of $\mathcal{C}([a, b], R)$, where it will still satisfy (i)–(v). The uniqueness assertion follows from the Hahn-Banach theorem.

2.1.2. EXAMPLE. Let $F = R^2$ and write a vector

$$f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}.$$

Then a path

$$f(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix},$$

and

$$\int_a^b \begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix} dt = \begin{pmatrix} \int_a^b f_1(t)\,dt \\ \int_a^b f_2(t)\,dt \end{pmatrix}.$$

Hence the integral of a vector function may be obtained by ordinary integration component by component.

2.1.3. EXAMPLE. Let $F = \mathcal{C}_{2\pi}$ be the Banach space of continuous functions periodic with period 2π. Then a continuous path $f(t) \in \mathcal{C}_{2\pi}$ on $a \leqslant t \leqslant b$ is a continuous function $f(t, x)$ on $[a, b] \times (-\infty, \infty)$ periodic in x, setting $f(t)(x) = f(t, x)$. The integral of the path is given by

$$\left\{\int_a^b f(t)\, dt\right\}(x) = \int_a^b f(t, x)\, dt.$$

We have the usual formula for a change of variables.

2.1.4 THEOREM. *Let $u = \gamma(t)$ be a C^1 monotone increasing function on $a \leqslant t \leqslant b$ and let $f(u)$ be a C^1 parametrized curve on $\gamma(a) \leqslant u \leqslant \gamma(b)$ with values in a Fréchet space. Then*

$$\int_{\gamma(a)}^{\gamma(b)} f(u)\, du = \int_a^b f(\gamma(t)) \cdot \gamma'(t)\, dt.$$

PROOF. Apply an arbitrary linear functional to each side. Equality holds by the usual result for real-valued functions. Then appeal to the Hahn-Banach theorem.

2.1.5. THEOREM. *Let X be a topological space and F a Fréchet space. Let f: $X \times [a, b] \to R$ be a continuous map. Define a map g: $X \to R$ by*

$$g(x) = \int_a^b f(x, t)\, dt.$$

Then g is also continuous.

PROOF. Pick $x_0 \in X$. Since $x_0 \times [a, b]$ is compact, given any continuous seminorm $\| \ \|$ on F and any $\varepsilon > 0$ we can find a neighborhood U of x_0 in X such that, for all $x \in U$ and all t in $[a, b]$, $\| f(x, t) - f(x_0, t)\| \leqslant \varepsilon$. Then

$$\| g(x) - g(x_0)\| \leqslant \int_a^b \| f(x, t) - f(x_0, t)\|\, dt \leqslant \varepsilon(b - a).$$

This shows g is continuous.

2.2. *Parametrized curves.* Let $f(t)$ be a continuous path in a Fréchet space. We can define its derivative in the usual way.

2.2.1. DEFINITION. For a parametrized curve

$$f'(t) = \lim_{h \to 0} [f(t + h) - f(t)]/h.$$

If $f(t)$ is the position at time t then $f'(t)$ is the velocity. If the limit exists and is continuous we say f is continuously differentiable or C^1. We can now state the two fundamental theorems of calculus.

2.2.2. THEOREM. *If $f(t)$ is a C^1 curve on $a \leqslant t \leqslant b$ with values in a Fréchet space then*

$$f(b) - f(a) = \int_a^b f'(t)\, dt.$$

2.2.3. THEOREM. *If $f(t)$ is a C^0 curve on $a \leqslant t \leqslant b$ and if $g(t) = \int_a^t f(\theta)\, d\theta$ then $g(t)$ is C^1 and $g'(t) = f(t)$.*

PROOF. If l is a continuous linear functional and f is C^1 then $l \circ f$ is C^1 and $(l \circ f)' = l \circ (f')$. By the standard result for $F = R$

$$l(f(b) - f(a)) = l\left(\int_a^b f'(t)\,dt\right).$$

Then Theorem 2.2.2 follows from the Hahn-Banach theorem. To prove the other result we need a lemma.

2.2.4 LEMMA.

$$\int_t^{t+h} f(\theta)\,d\theta = h\int_0^1 f(t + hu)\,du.$$

PROOF. If $f(\theta)$ is real-valued this follows from the substitution $\theta = t + hu$. The general case comes by applying a linear functional and using the Hahn-Banach theorem.

Using the lemma, we see that

$$\frac{g(t+h) - g(t)}{h} = \int_0^1 f(t + hu)\,du.$$

Then if f is continuous we may appeal to Theorem 2.1.5 to take the limit as $h \to 0$ on the right by setting $h = 0$, so that g is differentiable and $g'(t) = f(t)$. Since f is C^0 we see g is C^1.

2.2.5. COROLLARY. *If $f(a) = g(a)$ and if f and g are C^1 with $f'(t) = g'(t)$ for $a \leqslant t \leqslant b$ then $f(b) = g(b)$.*

2.2.6. COROLLARY. *If f is C^1 on $a \leqslant t \leqslant b$ and if $\| f'(t)\| \leqslant K$ then $\| f(b) - f(a)\| \leqslant K(b - a)$.*

We say that the path $f(t)$ is $\mathcal{C}^\infty$ if all of its derivatives $f^{(n)}(t)$ exist and are continuous.

2.2.7. EXAMPLE. A smooth path $f(t)$ in $C^\infty_{2\pi}$ for $a \leqslant t \leqslant b$ is a smooth function $f(t, x)$ periodic in x, setting $f(t)(x) = f(t, x)$. We have $f'(t)(x) = D_t f(t, x)$.

I.3. The directional derivative.

3.1. *Definition of the directional derivative.* Let F and G be Fréchet spaces, U an open subset of F, and P: $U \subseteq F \to G$ a continuous nonlinear map.

3.1.1. DEFINITION. The *derivative* of P at the point $f \in U$ in the direction $h \in F$ is defined by

$$DP(f)h = \lim_{t \to 0} \frac{P(f + th) - P(f)}{t}.$$

We say P is differentiable at f in the direction h if the limit exists. We say P is continuously differentiable (or C^1) on U if the limit exists for all $f \in U$ and all $h \in F$ and if DP: $(U \subseteq F) \times F \to G$ is continuous (jointly as a function on a subset of the product).

Note well that this definition of a C^1 map does not agree with the usual definition for a Banach space. The derivative will of course be the same map, but our continuity requirement is weaker.

3.1.2. EXAMPLE. Let f: $(a, b) \subseteq R \to F$ be a parametrized curve. The simple vector space R contains the distinguished vector 1 which spans the space, and $f'(t) = Df(t)1$. Thus our two motions of differentiability coincide.

3.1.3. EXAMPLE. Let L be a continuous linear map. Then L is C^1 and $DL(f)h = Lh$.

PROOF. By linearity $L(f + th) = L(f) + tL(h)$. Then by the definition

$$DL(f)h = \lim_{t \to 0} [L(f + th) - L(f)]/t = L(h).$$

3.1.4. EXAMPLE. Let P: $U \subseteq R^2 \to R^2$ and write

$$\begin{pmatrix} z \\ w \end{pmatrix} = P\begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$DP\begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \partial z/\partial x & \partial z/\partial y \\ \partial w/\partial x & \partial w/\partial y \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}$$

and the directional derivative is given by the matrix of partial derivatives.

3.1.5. EXAMPLES. Consider the following maps P: $C^\infty([a, b]) \to C^\infty([a, b])$ and their derivatives:

(a) $P(f) = f^2$, $DP(f)h = 2fh$,

(b) $P(f) = f^3$, $DP(f)h = 3f^2h$,

(c) $P(f) = e^f$, $DP(f)h = e^f h$,

(d) $P(f) = ff'$, $DP(f)h = fh' + f'h$,

(e) $P(f) = ff'' + f'^2$, $DP(f)h = fh'' + 2f'h' + f''h$,

(f) $P(f) = \sqrt{1 + f'^2}$, $DP(f)h = f'h'/\sqrt{1 + f'^2}$.

3.1.6. EXAMPLE. Let U be a relatively open subset of $[a, b] \times R$ and let $p(x, y)$ be a smooth function on U. Let $\tilde{U}$ be the open subset of the Fréchet space $C^\infty([a, b])$ consisting of all functions of whose graph $y = f(x)$ lies in U. Define a nonlinear map P: $\tilde{U} \subseteq C^\infty[a, b] \to C^\infty[a, b]$ by setting

$$P(f)(x) = p(x, f(x)).$$

Then the map P is C^1 (and indeed C^∞) and

$$[DP(f)h](x) = D_y p(x, f(x))h(x).$$

3.1.7. EXAMPLE. Let X be a compact manifold and let V and W be vector bundles over X. We form the Fréchet spaces $C^\infty(X, V)$ and $C^\infty(X, W)$ of smooth sections of V and W over X. Let U be an open subset of V and let $\tilde{U}$ be the open subset of all sections in $C^\infty(X, V)$ whose image lies in U. Let p: $U \subseteq V \to W$ be a smooth map which takes points in the fibre of V over a point x into points in the fibre of W over the same point x (so that if π_v and π_w are the bundle projection maps then $\pi_v = p \circ \pi_w$). Define a nonlinear operator P: $\tilde{U} \subseteq C^\infty(X, V) \to C^\infty(X, W)$ by

$$P(f) = p \circ f.$$

We call such a map P a *vector bundle operator*. Choose local coordinates x on X and Y in the fibre of V and z in the fibre of W. Then p is given locally by $z = p(x, y)$. A section f of the bundle V is given locally by $y = f(x)$. The operator P is given locally by

$$[P(f)](x) = p(x, f(x)).$$

Except for a fancier setting the situation is the same as before. The derivative is given locally by

$$[DP(f)h](x) = D_y p(x, f(x))h(x)$$

where $D_y p$ is interpreted as the matrix of partial derivatives with respect to the y variable, and is applied to the vector h. It is easy to see that P is C^0. For a proof that P is C^1, see Example 3.3.3, and for a proof that P is C^∞, see Example 3.6.6.

3.1.8. EXAMPLE. If V is a vector bundle over a compact manifold X, the r-jet bundle j^rV may be defined intrinsically; but if we introduce connections and covariant derivatives then

$$j^r f = (f, Df, D^2 f, \ldots, D^r f).$$

Let p be a smooth map of an open set U in j^rV into another bundle W taking fibres into fibres over the same point, and let $\tilde{U}$ be the open set of all $f \in C^\infty(X, V)$ with the image of $j^r f$ in U. A nonlinear *differential operator* P from V into W of degree r is a map P: $\tilde{U} \subseteq C^\infty(X, V) \to C^\infty(X, W)$ given by $Pf = p \circ j^r f$. The operator P is the composition of the continuous linear map j^r with a vector bundle operator of the type before, and hence is also smooth.

3.1.9. EXAMPLE. Two circular rings of radius r are held parallel in space and a soap bubble is blown connecting one to the other. It will form in a surface of revolution having the least possible area. If the rings are positioned at $x = \pm l$ perpendicular to the x-axis and if the surface is obtained by revolving the graph of $y = f(x)$ around the x-axis, the surface will have area

$$A(f) = \int_{-l}^{l} 2\pi f \sqrt{1 + f'^2}\, dx.$$

We can regard A as a nonlinear map

$$A: C^\infty([-l, l]) \to R$$

on the Fréchet space of smooth functions with real values. It is differentiable, and its derivative is given by

$$DA(f)h = \int_{-l}^{l} 2\pi \left\{ \frac{ff'}{\sqrt{1 + f'^2}} h' + \sqrt{1 + f'^2}\, h \right\} dx$$

using the familiar rules for the variation of a product and a square root and a square. If h vanishes at the endpoints $x = \pm l$, we may integrate by parts to get

$$DA(f)h = \int_{-l}^{l} 2\pi \left\{ \frac{1 + f'^2 - ff''}{(1 + f'^2)^{3/2}} \right\} h\, dx.$$

If the surface area A is to be a minimum over all f with $f(\pm l) = r$, then the derivative $DA(f) = 0$ for all variations h with $h(\pm l) = 0$. Clearly this happens if and only if f satisfies the differential equation

$$ff'' = f'^2 + 1$$

whose symmetrical solutions are given by

$$f(x) = (\cosh sx)/s$$

for various values of the constant s. To satisfy the boundary conditions s must be chosen so that $\cosh sl = sr$. Since the problem is invariant under a change of scale, the solution depends only on the ratio r/l. Fixing attention on the case $l = 1$, we must solve

$$r = (\cosh s)/s.$$

The minimum value m of this function occurs where the secant line from the origin to the graph of $\cosh s$ becomes tangent, which happens when $s \tanh s = 1$. This equation has a solution $s \approx 1.2$ which gives $m \approx 1.5$ for the critical value. If $r/l > m$ there are two solutions, while if $r/l < m$ there are none. We shall see later that only one of the two solutions is stable in Example 3.5.12.

3.2. *Properties of the derivative.* We begin with an integral version of the fundamental theorem of calculus.

3.2.1. LEMMA. *If P is C^1 and f and h are given then $P(f + th)$ is a C^1 path in t and*

$$P(f + th)' = DP(f + th)h.$$

PROOF. We have

$$P(f + th)' = \lim_{u \to 0} [P(f + th + uh) - P(f + th)]/u = DP(f + th)h.$$

3.2.2. THEOREM. *If P: $U \subseteq F \to G$ is a C^1 map between Fréchet spaces and if the path from t to $t + h$ lies in U then*

$$P(f + h) - P(f) = \int_0^1 DP(f + th)h \, dt.$$

PROOF. By the previous lemma, $P(f + th)' = DP(f + th)h$. We need only apply Theorem 2.2.2.

Next we wish to show that if P is C^1 then $DP(f)h$ is always linear in h. We begin with scalar multiplication.

3.2.3. LEMMA. *If c is a scalar and P is C^1 then*

$$DP(f)ch = cDP(f)h.$$

PROOF. We apply the definition and then substitute $u = ch$.

$$\begin{aligned} DP(f)ch &= \lim_{t \to 0} [P(f + cth) - P(f)]/t \\ &= c \lim_{t \to 0} [P(f + cth) - P(f)]/ct \\ &= c \lim_{u \to 0} [P(f + uh) - P(f)]/u = cDP(f)h. \end{aligned}$$

3.2.4. LEMMA. *If* P: $U \subseteq F \to G$ *is* C^1, $f \in U$ *and* $h \in F$, *and* t *is small, then*

$$[P(f+th) - P(f)]/t = \int_0^1 DP(f+uth)h\,du.$$

PROOF. We apply Theorem 3.2.2 replacing h by th and t by u, so that

$$P(f+th) - P(f) = \int_0^1 DP(f+uth)th\,du.$$

Then we use Lemma 3.2.3 to factor t out of the derivative and Theorem 2.1.1(v) to factor it out of the integral.

3.2.5. THEOREM. *If* P: $U \subseteq F \to G$ *is* C^1, $f \in U$ *and* $h_1, h_2 \in F$ *then*

$$DP(f)(h_1 + h_2) = DP(f)h_1 + DP(f)h_2.$$

PROOF. First write

$$\begin{aligned} &P(f + t(h_1 + h_2)) - P(f) \\ &\quad = [P(f + th_1 + th_2) - P(f + th_1)] + [P(f + th_1) - P(f)]. \end{aligned}$$

Using the previous lemma

$$[P(f + th_1) - P(f)]/t = \int_0^1 DP(f + uth_1)h_1\,du,$$

$$[P(f + th_1 + th_2) - P(f + th_1)]/t = \int_0^1 DP(f + th_1 + uth_2)h_2\,du;$$

therefore

$$\begin{aligned} &[P(f + t(h_1 + h_2)) - P(f)]/t \\ &\quad = \int_0^1 DP(f + uth_1)h_1\,du + \int_0^1 DP(f + th_1 + uth_2)h_2\,du. \end{aligned}$$

We let $t \to 0$ on the left. We can just put $t = 0$ on the right by Theorem 2.1.5. Then

$$DP(f)(h_1 + h_2) = DP(f)h_1 + DP(f)h_2$$

as desired.

3.3. *The chain rule.* We begin with a useful characterization of C^1 functions.

3.3.1. LEMMA. *Let* P: $U \subseteq F \to G$ *be a continuous map and suppose for simplicity that* U *is convex. Then* P *is continuously differentiable if and only if there exists a continuous map* L: $(U \subseteq F) \times (U \subseteq F) \times F \to G$ *with* $L(f_0, f_1)h$ *linear in the last variable* h *such that for all* f_0 *and* f_1 *in* U

$$P(f_1) - P(f_0) = L(f_0, f_1)(f_1 - f_0).$$

In this case we always have

$$DP(f)h = L(f, f)h.$$

PROOF. Let $f_t = (1-t)f_0 + tf_1 = f_0 + t(f_1 - f_0)$. Then by Theorem 3.2.5 if P is C^1 we can put

$$L(f_0, f_1)h = \int_0^1 DP(f_t)h\,dt$$

and this gives us the desired map L. Clearly L is continuous and linear in h, and $P(f_1) - P(f_0) = L(f_0, f_1)(f_1 - f_0)$.

Conversely, if such a map L exists, then

$$[P(f + th) - P(f)]/t = L(f, f + th)h$$

using the linearity in h to pull out the scalar t. It follows letting $t \to 0$ that P is differentiable and $DP(f)h = L(f, f)h$. Since L is C^0, P is C^1.

3.3.2. EXAMPLE. Let P: $R \to R$ be defined by $P(f) = f^2$. Then

$$P(f_1) - P(f_0) = f_1^2 - f_0^2 = (f_1 + f_0)(f_1 - f_0)$$

so we must take

$$L(f_0, f_1)h = (f_1 + f_0)h.$$

If $P(f) = f^3$ then

$$P(f_1) - P(f_0) = f_1^3 - f_0^3 = (f_1^2 + f_1 f_0 + f_0^2)(f_1 - f_0)$$

so we must take

$$L(f_0, f_1)h = (f_1^2 + f_1 f_0 + f_0^2)h.$$

In one dimension L is unique, but in higher dimensions it is not.

3.3.3. EXAMPLE. Let P: $\tilde{U} \subseteq C^\infty(X, V) \to C^\infty(X, W)$ be a vector bundle operator as described in Example 3.1.7. We can use Lemma 3.3.1 to prove that P is C^1. Suppose P is induced locally by the function $p(x, y)$. Since p is smooth, we can find a function $l(x, y_0, y_1)z$ linear in the vector z so that

$$p(x, y_1) - p(x, y_0) = l(x, y_0, y_1)(y_1 - y_0).$$

We can patch these together using a partition of unity in x since X is compact. We do not need a partition of unity in y if we restrict our attention to a small enough neighborhood U of the graph of a given section. This gives us an operator

$$L: \tilde{U} \times \tilde{U} \times C^\infty(X, V) \to C^\infty(X, W)$$

given locally by

$$L(f_0, f_1)h = l(x, f_0(x), f_1(x))h(x)$$

such that

$$P(f_1) - P(f_0) = L(f_0, f_1)(f_1 - f_0).$$

Since L is clearly continuous and linear in h, it follows that P is C^1.

Now we can prove the chain rule.

3.3.4. THEOREM. *If P and Q are C^1 so is their composition $Q \circ P$ and*

$$D[Q \circ P](f)h = DQ(P(f))DP(f)h.$$

PROOF. Since the theorem is a local result we can always take P and Q to be defined on convex neighborhoods of given points and apply the previous lemma. Then we can find continuous functions $L(f_0, f_1)h$ and $M(g_0, g_1)k$ linear in h and k such that

$$P(f_1) - P(f_0) = L(f_0, f_1)(f_1 - f_0), \qquad Q(g_1) - Q(g_0) = M(g_0, g_1)(g_1 - g_0).$$

Define a function $N(f_0, f_1)h$ by

$$N(f_0, f_1)h = M(P(f_0), P(f_1))L(f_0, f_1)h.$$

Then N is continuous and linear in h, and letting $g_0 = P(f_0)$ and $g_1 = P(f_1)$ we have

$$Q(P(f_1)) - Q(P(f_0)) = N(f_0, f_1)(f_1 - f_0).$$

Then by the converse side of the previous lemma, $Q \circ P$ is C^1. Since $L(f, f)h = DP(f)h$ and $M(g, g)k = DQ(g)k$, we have

$$\begin{aligned} D[Q \circ P](f)h = N(f, f)h &= M(P(f), P(f))L(f, f)h \\ &= DQ(P(f))DP(f)h \end{aligned}$$

as claimed.

3.3.5. COROLLARY. *If $f(t)$ is a parametrized C^1 curve and P is a C^1 map then $P(f(t))$ is also a parametrized C^1 curve and*

$$P(f(t))' = DP(f(t))f'(t).$$

From this we see that a curve passing through the point f with velocity h is mapped by P into a curve passing through the point $P(f)$ with velocity $DP(f)h$. This provides a good intuitive interpretation of the directional derivative.

3.4. *Partial derivatives.* Given a function $P(f, g)$ of two or more variables we can take a partial derivative with respect to just one or the other.

3.4.1. DEFINITION.

$$D_f P(f, g)h = \lim_{t \to 0} [P(f + th, g) - P(f, g)]/t,$$

$$D_g P(f, g)k = \lim_{t \to 0} [P(f, g + tk) - P(f, g)]/t.$$

By repeating the proof of Lemma 3.3.1 we can easily establish the following result.

3.4.2. LEMMA. *The partial derivative $D_f P(f, g)h$ exists and is continuous if and only if there exists a continuous function $L(f_0, f_1, g)h$ linear in h with*

$$P(f_1, g) - P(f_0, g) = L(f_0, f_1, g)(f_1 - f_0).$$

In this case $D_f P(f, g)h = L(f, f, g)h$.

3.4.3. THEOREM. *The partial derivatives $D_f P$ and $D_g P$ exist and are continuous if and only if the total derivative DP exists and is continuous. In that case*

$$DP(f, g)(h, k) = D_f P(f, g)h + D_g P(f, g)k.$$

PROOF. By definition

$$DP(f, g)(h, k) = \lim_{t \to 0} [P(f + th, g + tk) - P(f, g)]/t.$$

If DP exists and is continuous, then putting $k = 0$ so does $D_f P(f, g)h = DP(f, g)(h, 0)$, and putting $h = 0$ so does $D_g P(f, g)k = DP(f, g)(h, k)$, and by linearity the formula is valid.

Conversely if $D_f P$ and $D_g P$ both exist and are continuous, we can find continuous functions $L(f_0, f_1, g)h$ and $M(f, g_0, g_1)k$ linear in h and k with

$$P(f_1, g) - P(f_0, g) = L(f_0, f_1, g)(f_1 - f_0),$$
$$P(f, g_1) - P(f, g_0) = M(f, g_0, g_1)(g_1 - g_0).$$

Then

$$P(f_1, g_1) - P(f_0, g_0) = L(f_0, f_1, g_1)(f_1 - f_0) + M(f_0, g_0, g_1)(g_1 - g_0).$$

Put

$$N(f_0, g_0, f_1, g_1)(h, k) = L(f_0, f_1, g_1)h + M(f_0, g_0, g_1)k.$$

Then N is continuous and linear in (h, k) and

$$P(f_1, g_1) - P(f_0, g_0) = N(f_0, g_0, f_1, g_1)(f_1 - f_0, g_1 - g_0)$$

so by Lemma 3.3.1 we have $P(f, g)$ is C^1.

If a function $L(f, g)$ is linear in g then it is only necessary to differentiate it in f, as the following theorem shows.

3.4.4. COROLLARY. *If $L(f, g)$ is jointly continuous, C^1 separately in f and linear separately in g, then it is C^1 jointly in f and g, and*

$$DL(f, g)(h, k) = D_f L(f, g)h + L(f, k).$$

PROOF. $D_f L$ exists and is continuous by hypothesis, while $D_g L(f, g)k = L(f, k)$ exists and is continuous by linearity.

For this reason we adopt the following convention. If $L(f)h$ is linear in h, we regard its derivative as the partial derivative with respect to f. If we differentiate with respect to f in the direction k, we obtain

$$DL(f)\{h, k\} = \lim_{t \to 0} [L(f + tk)h - L(f)h]/t.$$

3.4.5. THEOREM. *If $L(f)h$ is C^1 and linear in h then $DL(f)\{h, k\}$ is bilinear, i.e. linear separately in h and k.*

PROOF. Since the total derivative is linear, so is the partial derivative with respect to f. Thus

$$DL(f)\{h, k_1 + k_2\} = DL(f)\{h, k_1\} + DL(f)\{h, k_2\}.$$

To see the linearity in h, we appeal to the definition of the derivative.

$$\begin{aligned} DL(f)\{h_1 + h_2, k\} &= \lim_{t \to 0} [L(f + tk)(h_1 + h_2) - L(f)(h_1 + h_2)]/t \\ &= \lim_{t \to 0} [L(f + tk)h_1 - L(f)h_1]/t + \lim_{t \to 0} [L(f + tk)h_2 - L(f)h_2]/t \\ &= DL(f)\{h_1, k\} + DL(f)\{h_2, k\}. \end{aligned}$$

3.5. *Second derivatives.* The second derivative is the derivative of the first derivative. Following the previous convention, we differentiate $DP(f)h$ with respect to f only, in the direction k. This gives the following definition.

3.5.1. DEFINITION.

$$D^2 P(f)\{h, k\} = \lim_{t \to 0} [DP(f + tk)h - DP(f)h]/t.$$

We say P is C^2 if DP is C^1, which happens if and only if D^2P exists and is continuous. If P: $U \subseteq F \to G$ we require D^2P to be continuous jointly on the product as a map

$$D^2P\colon (U \subseteq F) \times F \times F \to G.$$

From Theorem 3.4.5. we immediately get the following result.

3.5.2. THEOREM. *If P is C^2 then $D^2P(f)\{h, k\}$ is bilinear, i.e. linear separately in h and k.*

We have the following interesting characterization of the second derivative as a limit of a second difference.

3.5.3. THEOREM. *If P is C^2 then*

$$D^2P(f)\{h, k\} = \lim_{t,u\to 0} [P(f + th + uk) - P(f + th) - P(f + uk) + P(f)]/tu.$$

PROOF. By Lemma 3.2.4 we have

$$[P(f + th) - P(f)]/t = \int_{\theta=0}^{1} DP(f + \theta th)h \, d\theta,$$

$$[P(f + th + uk) - P(f + uk)]/t = \int_{\theta=0}^{1} DP(f + \theta th + uk)h \, d\theta,$$

$$[DP(f + \theta th + uk)h - DP(f + \theta th)h]/u = \int_{\eta=0}^{1} D^2P(f + \theta th + \eta uk)\{h, k\} \, d\eta.$$

It follows that

$$[P(f + th + uk) - P(f + th) - P(f + uk) + P(f)]/tu = \int_{\theta=0}^{1}\int_{\eta=0}^{1} D^2P(f + \theta th + \eta uk)\{h, k\} \, d\eta \, d\theta.$$

Taking the limit as t and $u \to 0$ gives the desired result.

3.5.4. COROLLARY. *If P is C^2 then the second derivative is symmetric, so that*

$$D^2P(f)\{h, k\} = D^2P(f)\{k, h\}.$$

PROOF. The second difference is symmetric and the second derivative is its limit.

3.5.5. THEOREM. *If P and Q are C^2 so is their composition $Q \circ P$ and*

$$D^2[Q \circ P](f)\{h, k\} = D^2Q(P(f))\{DP(f)h, DP(f)k\} + DQ(P(f))D^2P(f)\{h, k\}.$$

PROOF. This follows from applying the chain rule

$$D[Q \circ P](f)h = DQ(P(f))DP(f)h$$

to itself.

We shall need to use Taylor's formula with integral remainder for the second derivative in estimating a certain quadratic error term.

3.5.6. THEOREM. *If P: $U \subseteq F \to G$ is C^2 and if the path connecting f and $f + h$ lies in U then*

$$P(f+h) = P(f) + DP(f)h + \int_0^1 (1-t)D^2P(f+th)\{h,h\}\,dt.$$

PROOF. The trick is to integrate by parts.

$$DP(f+th)h = (1-t)D^2P(f+th)\{h,h\} - \{(1-t)DP(f+th)h\}'.$$

Integrating this over $0 \leqslant t \leqslant 1$ and using Theorem 3.2.2 and Theorem 2.2.2

$$\begin{aligned} P(f+h) - P(f) &= \int_0^1 DP(f+th)h\,dt \\ &= \int_0^1 (1-t)D^2P(f+th)\{h,h\}\,dt + DP(f)h \end{aligned}$$

which proves the theorem.

3.5.7. EXAMPLE. If f: $(a, b) \subseteq R \to F$ is a parametrized curve then

$$f''(t) = D^2f(t)\{1,1\}.$$

3.5.8. EXAMPLE. If L is a continuous linear map then

$$D^2L(f)\{h,k\} = 0.$$

3.5.9. EXAMPLE. If P: $U \subseteq R^2 \to R$ is smooth function $z = P\binom{x}{y}$ then

$$D^2P\binom{x}{y}\left\{\binom{r}{s},\binom{u}{v}\right\} = \frac{\partial^2 z}{\partial x^2}\cdot ru + \frac{\partial^2 z}{\partial x \partial y}\cdot (rv+su) + \frac{\partial^2 z}{\partial y^2}sv.$$

Hence D^2P is the quadratic form associated with the matrix of second partial derivatives

$$\begin{pmatrix} \partial^2/\partial x^2 & \partial^2 z/\partial x \partial y \\ \partial^2 z/\partial x \partial y & \partial^2 z/\partial y^2 \end{pmatrix}.$$

If the function P has a local minimum at f then $D^2P(f)\{h,h\} \geqslant 0$ for all h, which happens if and only if the matrix of second partials is positive.

3.5.10. EXAMPLES. Here are the second derivatives of the operators in Examples 3.1.5.

(a) $D^2P(f)\{h,k\} = 2hk$,
(b) $D^2P(f)\{h,k\} = 6fhk$,
(c) $D^2P(f)\{h,k\} = e^f hk$,
(d) $D^2P(f)\{h,k\} = h'k + hk'$,
(e) $D^2P(f)\{h,k\} = h''k + 2h'k' + hk''$,
(f) $D^2P(f)\{h,k\} = h'k'/(1+f'^2)^{3/2}$.

3.5.11. EXAMPLE. If we consider the operator

$$Pf(x) = p(x, f(x))$$

of Example 3.1.6 or 3.1.7, its derivative is

$$[DP(f)h](x) = D_y p(x, f(x)) \cdot h(x)$$

and its second derivative is

$$[D^2P(f)\{h, k\}](x) = D_y^2 p(x, f(x)) \cdot h(x) \cdot k(x).$$

For Example 3.1.6 this is ordinary multiplication, while for Example 3.1.7 we regard $D_y^2 p$ as a bilinear form applied to the pair of vectors h and k.

3.5.12. EXAMPLE. We return to the soap bubble problem of Example 3.1.9. We can compute the second derivative of A, and after simplifying we get

$$D^2A(f)\{h, k\} = \int_{-l}^{l} 2\pi \left\{ \frac{f}{(1+f'^2)^{3/2}} h'k' + \frac{f'}{(1+f')^2}(hk)' \right\} dx.$$

If we consider only variations h and k which vanish at the endpoints $x = \pm l$, we can integrate by parts to obtain

$$D^2A(f)\{h, k\} = \int_{-l}^{l} 2\pi \frac{fh'k' - f''hk}{(1+f'^2)^{3/2}} dx.$$

Since all the solutions are similar under a change of scale, we focus our attention on the solution $f(x) = \cosh x$. Then

$$D^2(\cosh x)\{h, h\} = \int_{-l}^{l} 2\pi \operatorname{sech}^2 x [h'^2 - h^2]\, dx.$$

This quadratic form will be positive definite for all h vanishing at the endpoints only when l is small enough. The critical value of l occurs when there is a null eigenvector of the quadratic form, which happens for a given f when there exists an h such that for all k we have $D^2A(f)\{h, k\} = 0$. Integrating by parts for h and k vanishing at the endpoints we have

$$D^2(\cosh x)\{h, k\} = -\int_{-l}^{l} 2\pi \operatorname{sech}^2 x \{h'' - 2 \cdot \tanh x \cdot h' + h\} k\, dx$$

which vanishes for all k when h satisfies the differential equation

$$h'' - 2 \tanh x \cdot h' + h = 0.$$

If the equation has a solution vanishing at both endpoints it will have a symmetric solution, which is found to be

$$h(x) = x \sinh x - \cosh x.$$

Significantly, $h(x)$ is the derivative of the general solution $f(x) = (\cosh sx)/s$ with respect to s, evaluated at $s = 1$. For $h(x)$ to vanish at the endpoints we must have $l \tanh l = 1$. This occurs at the value $l \approx 1.2$ where $(\cosh l)/l = m \approx 1.5$ attains its minimum. Since $r = \cosh l$ when $f(x) = \cosh x$, the second derivative first fails to be positive definite at the critical value $r/l = m$. It now follows that when $r/l > m$ only one of the two solutions s of the equation $\cosh sl = sr$ produces a stable solution $f(x) = (\cosh sx)/s$, namely the one with the smaller value of s which stays further from the axis, while the other solution which comes nearer the axis is unstable. If we start with a stable soap bubble for $r/l > m$ and slowly increase l by moving the two rings apart, the

bubble will suddenly collapse when r/l reaches the critical value $m \approx 1.5$. Note that the neck of the bubble does not shrink to zero as we approach the critical position. Let n denote the radius of the neck. The critical position for the standard solution $y = \cosh x$ occurs at $l \approx 1.2$ where $(\cosh l)/l = m \approx 1.5$. Here $r = \cosh l \approx 1.8$. The neck occurs at $x = 0$ and the radius of the neck $n = 1$. Since the ratios are invariant in all solutions, we find that the critical position is reached when $l/r \approx 0.6$ and at that point $n/r \approx 0.5$. If l is increased beyond this point the neck of the bubble will suddenly collapse. The reader is invited to try the experiment.

3.6. *Higher derivatives.* The third derivative is the derivative of the second derivative. Since $D^2P(f)\{h, k\}$ is linear separately in h and k, we take only its partial derivative with respect to f in the direction l.

3.6.1. DEFINITION. If P: $U \subseteq F \to G$ then

$$D^3P(f)\{h, k, l\} = \lim_{t \to 0} \big[D^2P(f + tl)\{h, k\} - D^2P(f)\{h, k\}\big]/t.$$

Similar definitions apply to the higher derivatives. The nth derivative

$$D^nP(f)\{h_1, h_2, \ldots, h_n\}$$

will be regarded as a map

$$D^nP\colon (U \subseteq F) \times F \times \cdots \times F \to G.$$

We say P is of class C^n if D^nP exists and is continuous (jointly as a function on the product space).

3.6.2. THEOREM. *If P is C^n then $D^nP(f)\{h_1, h_2, \ldots, h_n\}$ is completely symmetric and linear separately in $h_1, h_2, \ldots, h_n$.*

PROOF. It is linear in $h_1, \ldots, h_{n-1}$ as the derivative of $D^{n-1}P$, which is already linear in these variables. It is linear in h_n as a derivative in that direction. It is symmetric in h_{n-1} and h_n as the second derivative of $D^{n-2}P$. It is symmetric in h_i and h_j for $i, j < n$ as the derivative of $D^{n-1}P$, which already has this symmetry. These transpositions generate the full symmetric group.

Sometimes it is more convenient to use the tangent functor.

3.6.3. DEFINITION. If P: $U \subseteq F \to V \subseteq G$ is a map between open subsets of Fréchet spaces, we define its tangent TP: $(U \subseteq F) \times F \to (V \subseteq G) \times G$ by

$$TP(f, h) = (P(f), DP(f)h).$$

Notice that TP is defined and continuous if and only if DP is defined and continuous. We let $T^2P = T(TP)$ and $T^nP = T(T^{n-1}P)$. Then T^nP is defined and continuous if and only if D^nP is defined and continuous (i.e. P is C^n).

3.6.4. THEOREM. *If P and Q are C^n, so is their composition $Q \circ P$, and*

$$T^n(Q \circ P) = (T^nQ) \circ (T^nP).$$

PROOF. When $n = 1$, we have $T(Q \circ P) = TQ \circ TP$ from the chain rule. The higher cases follow from induction on n.

We say a map is C^∞ if it is C^n for all n.

3.6.5. EXAMPLE. Let P: $\tilde{U} \subseteq \mathcal{C}^\infty(X,V) \to \mathcal{C}^\infty(X,V)$ be a vector bundle operator as described in Example 3.1.7. Then P is C^∞. This follows immediately from the fact that P is C^1 and TP is again a vector bundle operator. If $V \oplus V$ denotes the Whitney sum of V with itself, there is a natural isomorphism

$$\mathcal{C}^\infty(X,V) \times \mathcal{C}^\infty(X,V) \approx \mathcal{C}^\infty(X, V \oplus V).$$

Suppose P is induced by a map p: $U \subseteq V \to W$ of the vector bundles taking fibres into fibres over the same point. Locally p has the form

$$p\colon (x, y) \to (x, p(x, y)).$$

There is an intrinsically defined vertical tangent map

$$T_v p\colon U \oplus V \subseteq V \oplus V \to W \oplus W$$

given locally by

$$T_v p\colon (x, y, z) \to \big(x, p(x, y), D_y p(x, y)z\big)$$

which induces a vector bundle operator which may be identified with the tangent map

$$TP\colon \tilde{U} \times \mathcal{C}^\infty(X,V) \subseteq \mathcal{C}^\infty(X,V) \times \mathcal{C}^\infty(X,V) \to \mathcal{C}^\infty(X,W) \times \mathcal{C}^\infty(X,W).$$

3.6.6. EXAMPLE. Let P: $\tilde{U} \subseteq \mathcal{C}^\infty(X,V) \to \mathcal{C}^\infty(X,W)$ be a differential operator of degree r as described in Example 3.1.8. Then P is $\mathcal{C}^\infty$. This follows immediately from the observation that a differential operator is a composition of a vector bundle operator (which is $\mathcal{C}^\infty$) with the r-jet extension map j^r, which is a continuous linear map (and hence also $\mathcal{C}^\infty$).

I.4. Fréchet manifolds.

4.1. *Manifolds.* The usual definition of a manifold generalizes directly to Fréchet space calculus.

4.1.1. DEFINITION. A *Fréchet manifold* is a Hausdorff topological space with an atlas of coordinate charts taking their value in Fréchet spaces, such that the coordinate transition functions are all smooth maps between Fréchet spaces.

4.1.2. EXAMPLE. Let X be a compact finite dimensional manifold. A bundle over X is another finite dimensional manifold B with a smooth projection map π: $B \to X$ whose derivative is everywhere surjective. A section of the bundle is a smooth map f: $X \to B$ such that $\pi f = 1$ is the identity. The space of all sections of the bundle $\mathcal{C}^\infty(X, B)$ is a Fréchet manifold (at least if it isn't empty). Associated to each section f is a vector bundle over X called the vertical tangent bundle to B at f, which we denote by $T_v B(f)$. Its fibre at a point $x \in X$ consists of all the tangent vectors of B at $f(x)$ which lie in the null space of the derivative of π.

It is easy to construct a diffeomorphism from a neighborhood of the zero section of the vector bundle $T_v B(f)$ to a neighborhood of the image of f in B which takes fibres into fibres over the same point. This provides a one-to-one correspondence between sections near zero in the Fréchet space $\mathcal{C}^\infty(X, T_v B(f))$ and sections near f in the manifoldd $\mathcal{C}^\infty(X, B)$, and these maps serve as our

coordinate charts. The coordinate transition functions are clearly vector bundle maps in the sense of Example 3.1.7. Hence $\mathcal{C}^\infty(X, B)$ is a Fréchet manifold.

4.1.3. EXAMPLE. Let X and Y be manifolds with X compact. The space of all smooth maps of X into Y forms a Fréchet manifold $\mathcal{M}(X, Y)$. This is a special case of the preceding example, as a map of X into Y is a section of the product bundle $X \times Y$ over X.

4.1.4. EXAMPLE. As a special case, consider the manifold $\mathcal{M}(S^1, S^1)$ of maps of the circle to itself. It has an infinite number of components indexed by the degree k which may be any positive or negative integer. We let Z denote the set of all integers. Passing to the universal cover, a map of S^1 to S^1 lifts to a map of R^1 to R^1, which is just a real-valued function on the real line. If the map has degree zero its lift will be periodic, giving rise to an element of the Fréchet space $\mathcal{C}^\infty_{2\pi}$. The lift is not unique, but may vary by $2\pi n$ where $n \in Z$ is an integer. Hence the component $\mathcal{M}_0(S^1, S^1)$ of maps of degree zero is diffeomorphic to the quotient $\mathcal{C}^\infty_{2\pi}/2\pi Z$. This provides a global chart. More generally a map of degree k lifts to a function $f\colon R^1 \to R^1$ satisfying

$$f(x + 2\pi) = f(x) + 2\pi k$$

which we call k-periodic. If we let $\tilde{f}(x) = f(x) - kx$ then $\tilde{f}$ is periodic so $\tilde{f} \in \mathcal{C}^\infty_{2\pi}$. Hence the set of functions which are k-periodic is not a vector space but corresponds to the vector space $\mathcal{C}^\infty_{2\pi}$ by this affine shift. Again the lift is only unique up to a constant function $2\pi n$. The component $\mathcal{M}_k(S^1, S^1)$ of maps of degree k is also diffeomorphic to $\mathcal{C}^\infty_{2\pi}/2\pi Z$, and the above construction provides a global coordinate chart. Then the whole space

$$\mathcal{M}(S^1, S^1) \approx Z \times (\mathcal{C}^\infty_{2\pi}/2\pi Z).$$

4.1.5. EXAMPLE. Let X be a compact manifold and V a vector bundle over X. The space of all connections on V forms a Fréchet manifold $\mathcal{G}(V)$. The difference of two connections is a tensor in the bundle $L^2(TX \times V, V)$ of bilinear maps of TX and V into V. For any fixed connection we obtain a one-to-one correspondence between $\mathcal{G}(V)$ and the Fréchet space

$$\mathcal{C}^\infty(X, L^2(TX \times V, V))$$

by subtracting the reference connection from the variable one. The coordinate transition functions are just translations.

4.1.6. EXAMPLE. If V is a vector space, the Grassmann space $G_p(V)$ of all p-planes in V is a manifold. If V is a vector bundle over X then $G_p(V)$ becomes a fibre bundle over X. When $V = TX$ is the tangent bundle, a section of $G_p(TX)$ is a field of tangent p-planes. The space $\mathcal{C}^\infty(X, G_p(TX))$ of all tangent p-plane fields is a Fréchet manifold by Example 4.1.2.

4.1.7. EXAMPLE. Let X be a finite dimensional manifold and let $\mathcal{S}(X)$ denote the space of all compact smooth submanifolds of X. Then $\mathcal{S}(X)$ is a Fréchet manifold. Let $S \in \mathcal{S}(X)$ be a given submanifold, and let NS be its normal vector bundle, defined invariantly, as the quotient of the restriction of TX to S by TS. We can find a diffeomorphism between a neighborhood of the zero section in NS and a tubular neighborhood of S in X. This establishes a one-to-one correspondence between a neighborhood of 0 in the Fréchet space

$\mathcal{C}^\infty(S, NS)$ and a neighborhood of S in $\mathcal{S}(X)$, and these maps serve as our coordinate charts. We shall show later in Example 4.4.7 that the coordinate transition functions are C^∞. Note that all the submanifolds S in a connected component of $\mathcal{S}(X)$ are necessarily diffeomorphic, but there may be many components. Diffeomorphic submanifolds may lie in separate components also, as in the case of knotted circles in three space.

4.1.8. EXAMPLE. Let X be a finite dimensional manifold and let $\mathcal{R}(X)$ denote the set of all compact regions in X with smooth boundaries. Then $\mathcal{R}(X)$ is a Fréchet manifold. When X is not compact, a region is uniquely identified by its boundary, so $\mathcal{R}(X)$ is identified with the components of $\mathcal{S}(X)$ corresponding to compact submanifolds of codimension 1 which bound a compact region. When X is compact, each boundary determines two regions, so $\mathcal{R}(X)$ is a double cover of the boundary components of $\mathcal{S}(X)$. This double cover may be not just two copies. For if we consider regions diffeomorphic to a ball B^n in the sphere S^n, the complement is another ball B^n, and the first may be continuously rotated into the second.

4.2. *Submanifolds.* Let $\mathcal{M}$ be a Fréchet manifold and $\mathcal{N}$ a closed subset.

4.2.1. DEFINITION. $\mathcal{N}$ is a submanifold of $\mathcal{M}$ if every point of $\mathcal{N}$ lies in the domain of a coordinate chart on $\mathcal{M}$ with range in a product of Fréchet spaces $F \times G$ such that a point in the domain of the chart lies in the subset $\mathcal{N}$ if and only if its image under the chart lies in the subset $F \times 0$.

4.2.2. EXAMPLE. Let B be a bundle over a compact manifold X. A subbundle of B is a submanifold A of B such that the derivative of the projection map π for B is still surjective on A. Then A is itself a bundle over X, and the Fréchet manifold $\mathcal{C}^\infty(X, A)$ of sections of the subbundle A is a Fréchet submanifold of the Fréchet manifold $\mathcal{C}^\infty(X, B)$ of sections of the bundle B. Indeed, given any section f of A, we can find a tubular neighborhood of its image in the bundle B which is diffeomorphic to a tubular neighborhood of the zero section in a bundle $V \oplus W$ by a map taking fibres into fibres over the same point, such that the points in the neighborhood in B which lie in the subbundle A correspond to points in the Whitney sum $V \oplus W$ which lie in the vector subbundle $V \oplus 0$. Then we have a chart on $\mathcal{C}^\infty(X, B)$ with values in

$$\mathcal{C}^\infty(X, V \oplus W) \approx \mathcal{C}^\infty(X, V) \times \mathcal{C}^\infty(X, W)$$

such that the sections in $\mathcal{C}^\infty(X, A)$ correspond under the chart to $\mathcal{C}^\infty(X, V) \times 0$.

4.2.3. EXAMPLE. Let X be a compact manifold and Y a submanifold of Z. Then the Fréchet manifold $\mathcal{M}(X, Y)$ of smooth maps of X into Y is a smooth Fréchet submanifold of the Fréchet manifold of smooth maps of X into Z. Indeed, the maps of X into Y are the sections of the product bundle $X \times Y$ over X, and $X \times Y$ is a subbundle of $X \times Z$.

4.2.4. EXAMPLE. Let B be a bundle over a compact manifold X. Then the Fréchet manifold $\mathcal{C}^\infty(X, B)$ of sections of B is a submanifold of the Fréchet manifold $\mathcal{M}(X, B)$ of all smooth maps of X into B. Since it is sufficient to verify this locally, we may as well assume that B is an open subset of a vector bundle V over X, and consider a neighborhood of the zero section of V. The tangent bundle to V along the zero section is naturally isomorphic to $V \oplus TX$,

so we can choose a diffeomorphism between a neighborhood of the zero section in $V \oplus TX$ and a neighborhood of the zero section in the product bundle $X \times V$ such that the subbundle $V \oplus 0$ in $V \oplus TX$ corresponds to the subset of $X \times V$ which lies over the diagonal in $X \times X$. Then we get a chart on the manifold $\mathcal{M}(X, V)$ with values in the Fréchet space

$$\mathcal{C}^\infty(X, V \oplus TX) = \mathcal{C}^\infty(X, V) \times \mathcal{C}^\infty(X, TX)$$

such that the submanifold of true sections $\mathcal{C}^\infty(X, V)$ corresponds to the subset where the component in $\mathcal{C}^\infty(X, TX)$ is zero.

4.2.5. EXAMPLE. The manifold $\mathcal{E}(X, Y)$ of embeddings of a compact manifold X into a manifold Y is an open submanifold of the manifold $\mathcal{M}(X, Y)$ of all maps of X into Y. Likewise the manifold $\mathcal{D}(X)$ of all diffeomorphisms of a compact manifold to itself is an open submanifold of the manifold $\mathcal{M}(X, X)$ of all maps of X to itself.

4.2.6. EXAMPLE. We study in particular the manifold $\mathcal{D}(S^1)$ of diffeomorphisms of the circle. It has two components, those which preserve and reverse orientation, which we write as $\mathcal{D}^+(S^1)$ and $\mathcal{D}^-(S^1)$. Elements of $\mathcal{D}^+(S^1)$ have degree $+1$ and lift to functions $f(x)$ satisfying $f(x + 2\pi) = f(x) + 2\pi$. Then $f'(x + 2\pi) = f'(x)$ so the derivative $f'(x)$ is periodic. For a diffeomorphism we must have $f'(x) > 0$. To satisfy the condition on f we also need

$$\int_0^{2\pi} f'(x)\, dx = f(2\pi) - f(0) = 2\pi.$$

Let H denote the subset of the Fréchet space $\mathcal{C}^\infty_{2\pi}$ of functions which are strictly positive and have integral 2π. It is an open convex subset of a closed affine subspace of codimension 1, and hence a manifold. Then $f'(x)$ can be any element in H. It determines $f(x)$ up to the constant $f(0) \in R^1$, which is itself indeterminate up to a constant $2\pi n \in 2\pi Z$, so $f'(x)$ determines the diffeomorphism up to an element of $S^1 = R^1/2\pi Z$, namely the image of 0. This shows that $\mathcal{D}^+(S^1) \approx S^1 \times H$. Moreover H is clearly contractible, so $\mathcal{D}^+(S^1)$ has the homotopy type of S^1. The other component $\mathcal{D}^-(S^1)$ looks the same.

4.2.7. EXAMPLE. Let X be a submanifold of Y. Then the Fréchet manifold $\mathcal{S}(X)$ of compact submanifolds of X is a submanifold of the Fréchet manifold $\mathcal{S}(Y)$ of compact submanifolds of Y.

4.3. *Vector bundles.* It is natural to define Fréchet vector bundles over Fréchet manifolds in the usual way. Let $\mathcal{M}$ be a Fréchet manifold, $\mathcal{V}$ another Fréchet manifold, and π: $\mathcal{V} \to \mathcal{M}$ a projection map such that each fibre $\pi^{-1}f$ for $f \in \mathcal{M}$ has the structure of a vector space.

4.3.1. DEFINITION. We say that $\mathcal{V}$ is a Fréchet vector bundle over $\mathcal{M}$ with projection π if each point in $\mathcal{M}$ lies in the domain of a coordinate chart with values in an open subset U of a Fréchet space F and we can find a coordinate chart on $\mathcal{V}$ whose domain is the inverse image of the domain of the chart on $\mathcal{M}$ with values in the open set $(U \subseteq F) \times G$ in a product $F \times G$ for another Fréchet space G, so that the projection π of $\mathcal{V}$ on $\mathcal{M}$ corresponds to the projection of $U \times G$ on U and the vector space structure on each fibre is that induced by the vector space structure on G.

Note that for two such coordinate charts on $\mathcal{V}$ the coordinate transition function

$$(U \subseteq F) \times G \to (\tilde{U} \subseteq \tilde{F}) \times \tilde{G}$$

will always be linear from G to $\tilde{G}$. Conversely if we have an atlas on $\mathcal{V}$ of charts on products whose coordinate transitions have the form $(f, g) \to (\tilde{f}, \tilde{g})$ where $\tilde{f} = P(f)$ and $\tilde{g} = L(f)g$ is linear in g, then $\mathcal{V}$ will become a vector bundle.

4.3.2. EXAMPLE. Let $\mathfrak{M}$ be a Fréchet manifold. Then its tangent bundle $T\mathfrak{M}$ is a Fréchet vector bundle over $\mathfrak{M}$. The coordinate transition functions for $T\mathfrak{M}$ are just the tangents TP of the coordinate transition functions P for $\mathfrak{M}$. If $f(t)$ is a parametrized curve in $\mathfrak{M}$ then $f'(t)$ is a tangent vector to $\mathfrak{M}$ at $f(t)$. This sometimes provides an easy way to identify $T\mathfrak{M}$.

4.3.3. EXAMPLE. A path $f\colon (a, b) \to \mathfrak{M}(X, Y)$ is given by a map $f\colon (a, b) \times X \to Y$. Thus for each $x \in X$ the path $f(t)$ in $\mathfrak{M}(X, Y)$ gives a path $f(t)(x)$ in Y. Its t-derivative $f'(t)(x)$ is a tangent vector to Y at $f(t)(x)$. Hence $f'(t)$ is a section of the pull-back to X of the vector bundle TY under the map f. This gives the identification of the tangent space to $\mathfrak{M}(X, Y)$ at a map f as

$$T_f\mathfrak{M}(X, Y) = \mathcal{C}^\infty(X, f^*TY).$$

4.3.4. EXAMPLE. Let $\mathcal{S}(X)$ be the Fréchet manifold of compact smooth submanifolds of a finite dimensional manifold X. Its tangent bundle $T\mathcal{S}(X)$ has for its fibre at a submanifold $S \in \mathcal{S}(X)$ the Fréchet space of sections $\mathcal{C}^\infty(S, NS)$ of the normal bundle. Likewise the tangent bundle $T\mathcal{R}(X)$ to the Fréchet manifold of compact domains in X with smooth boundary has for its fibre at a domain $A \in \mathcal{R}(X)$ the Fréchet space $\mathcal{C}^\infty(\partial A, N\partial A)$ of sections over the boundary of the normal bundle. Since the normal bundle is always trivial (although not in a canonical way) we have

$$T\mathcal{R}(X)_A \approx \mathcal{C}^\infty(\partial A).$$

4.3.5. EXAMPLE. Let X be a finite dimensional manifold and $\mathcal{S}(X)$ the Fréchet manifold of all compact smooth submanifolds of X. Let $\mathcal{C}^\infty\mathcal{S}(X)$ denote the space of all smooth functions on all submanifolds $S \in \mathcal{S}(X)$. Then $\mathcal{C}^\infty\mathcal{S}(X)$ is a Fréchet vector bundle over $\mathcal{S}(X)$. The fibre over $S \in \mathcal{S}(X)$ is the Fréchet space $\mathcal{C}^\infty(S)$. A typical chart on $\mathcal{S}(X)$ at S has values in the Fréchet space $\mathcal{C}^\infty(S, NS)$ of sections of the normal bundle. Its choice provides a choice of a diffeomorphism between S and any nearby surface $\tilde{S}$. We can use this diffeomorphism to identify functions on S with functions on $\tilde{S}$. Hence we get a chart on $\mathcal{C}^\infty\mathcal{S}(X)$ with values in the product $\mathcal{C}^\infty(S, NS) \times \mathcal{C}^\infty(S)$ of the type required to make $\mathcal{C}^\infty\mathcal{S}(X)$ a Fréchet vector bundle.

4.3.6. EXAMPLE. Let X be a finite dimensional manifold and $\mathcal{R}(X)$ the Fréchet manifold of all compact regions in X with smooth boundary. Then there is a Fréchet vector bundle $\mathcal{C}^\infty\mathcal{R}(X)$ of all smooth functions on all regions. It is the bundle over $\mathcal{R}(X)$ whose fibre over any region $A \in \mathcal{R}(X)$ is the Fréchet space $\mathcal{C}^\infty(A)$ of smooth functions on the region A. The description of the coordinate transition functions involves the choice of a diffeomorphism from A to any nearby region $\tilde{A}$ which is not very canonical. The manifold

$\mathcal{R}(X)$ has a chart at A with values in the Fréchet space $\mathcal{C}^\infty(\partial A)$ obtained as follows. Choose a tubular neighborhood N of ∂A diffeomorphic to $\partial A \times [-1, 1]$. Let $y \in \partial A$ and $r \in [-1, 1]$ be local coordinates so that the part of N in A is given by $r \geqslant 0$. Then a nearby domain $\tilde{A}$ has its boundary given by $r = b(y)$ where $b \in \mathcal{C}^\infty(\partial A)$ is a small function. This provides our coordinate chart on $\mathcal{R}(X)$ near A. We define a diffeomorphism of $N \cap A$ to $N \cap \tilde{A}$ by the map $(y, r) \to (y, s)$ with

$$s = r + \phi(r)b(y)$$

where ϕ is a smooth function on $[-1, 1]$ with $\phi(r) = 1$ for $r \leqslant 0$ and $\phi(r) = 0$ for $r \geqslant \frac{1}{2}$. We extend the diffeomorphism to be the identity outside of N and inside A. A function $g \in \mathcal{C}^\infty(\tilde{A})$ corresponds to a function $f \in \mathcal{C}^\infty(A)$ by pulling back by this diffeomorphism, so that $f(y, r) = g(y, s)$ on $A \cap N$ and $f(x) = g(x)$ on the rest of A. This provides a coordinate chart on the vector bundle $\mathcal{C}^\infty\mathcal{R}(X)$ with values in the product space $\mathcal{C}^\infty(\partial A) \times \mathcal{C}^\infty(A)$.

4.4. *Maps of manifolds.* Let $\mathfrak{M}$ and $\mathfrak{N}$ be Fréchet manifolds.

4.4.1. DEFINITION. A map P: $\mathfrak{M} \to \mathfrak{N}$ is a smooth map of Fréchet manifolds if we can find charts around any point in $\mathfrak{M}$ and its image in $\mathfrak{N}$ such that the local representative of P in these charts is a smooth map of Fréchet spaces.

4.4.2. EXAMPLE. The inclusion of a submanifold is a smooth map.

4.4.3. EXAMPLE. The projection of a Fréchet vector bundle is a smooth map.

A smooth map P: $\mathfrak{M} \to \mathfrak{N}$ of Fréchet manifolds induces a *tangent map* TP: $T\mathfrak{M} \to T\mathfrak{N}$ of their tangent bundles which takes the fibre over $f \in \mathfrak{M}$ into the fibre over $P(f) \in \mathfrak{N}$ and is linear on each fibre. The local representatives for the tangent map TP are just the tangents of the local representatives for P. The *derivative* of P at f is the linear map

$$DP(f)\colon T\mathfrak{M}_f \to T\mathfrak{N}_{P(f)}$$

induced by TP on the tangent space. When the manifolds are Fréchet spaces this agrees with the previous definition.

4.4.4. EXAMPLE. Let $\mathfrak{M}$ be the Fréchet manifold of all compact connected one dimensional submanifolds of R^2, a component of the manifold $\mathcal{S}(R^2)$. Each manifold $S \in \mathfrak{M}$ is diffeomorphic to the circle. Using the Euclidean metric on R^2 we can define two smooth real valued functions on $\mathfrak{M}$

$$L\colon \mathfrak{M} \to R \quad \text{and} \quad A\colon \mathfrak{M} \to R$$

where $L(S)$ is the length of S and $A(S)$ is the area enclosed by S.

The tangent space to $\mathfrak{M}$ at S can be identified as the sections over S of its normal bundle, which may be trivialized canonically using the metric, so that $T\mathfrak{M}_S \simeq \mathcal{C}^\infty(S)$. Letting $f \in \mathcal{C}^\infty(S)$ we can compute the derivatives of L and A. If the curvature of S is given by $k \in \mathcal{C}^\infty(S)$ and if ds is the arc length then we have the formulas

$$DL(S)f = \int_S kf\,ds, \qquad DA(S)f = \int_S f\,ds.$$

If the length is to be a minimum for all curves enclosing a given area, then we must have $DL(S)f = 0$ for all f with $DA(S)f = 0$. This clearly happens if

and only if the curvature k is constant, which happens when the curve is a circle.

4.4.5. EXAMPLE. Let X, Y, and Z be finite dimensional manifolds with X and Y compact. Fix a map $g\colon X \to Y$ and define a map

$$C\colon \mathfrak{M}(Y, Z) \to \mathfrak{M}(X, Z)$$

by composition with g, so that $C(f) = f(g)$. Then C is smooth and its derivative is given by the formula

$$DC(f)h = h(g).$$

Note that if

$$h \in T\mathfrak{M}(Y, Z)_f = \mathcal{C}^\infty(Y, f^*TZ)$$

then

$$h(g) \in T\mathfrak{M}(X, Z)_{f(g)} = \mathcal{C}^\infty(X, g^*f^*TZ).$$

Alternatively we can fix a map $f\colon Y \to Z$ and define a map

$$C\colon \mathfrak{M}(X, Y) \to \mathfrak{M}(X, Z)$$

by composition with f, so that $C(g) = f(g)$. Then C is again smooth and its derivative is given by the formula

$$DC(g)k = Df(g)k.$$

Note that if

$$k \in T\mathfrak{M}(X, Y)_g = \mathcal{C}^\infty(X, g^*TY)$$

then since $Df\colon TY \to TZ$ we have

$$Df(g)k \in T\mathfrak{M}(X, Z)_{f(g)} = \mathcal{C}^\infty(X, g^*f^*TZ).$$

Finally we can do both at the same time. Define a map on the product manifold

$$C\colon \mathfrak{M}(Y, Z) \times \mathfrak{M}(X, Y) \to \mathfrak{M}(X, Z)$$

by composition, so that $C(f, g) = f(g)$. Then its derivative is the sum of the two partial derivatives, which gives the formula

$$DC(f, g)(h, k) = h(g) + Df(g)k.$$

It is important to notice that serious problems arise if we attempt to repeat this construction in Banach spaces. The space $\mathfrak{M}^r(X, Y)$ of maps of class C^r is a Banach manifold. However, the composition map

$$C\colon \mathfrak{M}^r(Y, Z) \times \mathfrak{M}^r(X, Y) \to \mathfrak{M}^r(X, Z)$$

is C^0 but not C^1, since the formula for DC involves Df. The composition

$$C\colon \mathfrak{M}^{r+1}(Y, Z) \times \mathfrak{M}^r(X, Y) \to \mathfrak{M}^r(X, Z)$$

is C^1 but not C^2. In Fréchet spaces everything can be made C^∞ and no problems of differentiability arise. Indeed the formula for DC and hence the tangent TC involves nothing more complicated than compositions and derivatives, so the existence of the higher tangents $T^2C, \ldots, T^mC$ follows from the chain rule by induction.

4.4.6. EXAMPLE. Let $\mathcal{D}(X)$ be the Fréchet manifold of diffeomorphisms of a compact manifold X of finite dimension, introduced in Example 4.2.5. There is a natural map

$$V\colon \mathcal{D}(X) \to \mathcal{D}(X)$$

which takes each diffeomorphism to its inverse, so that $V(f) = f^{-1}$. The map V is smooth. Its derivative is given by the formula

$$DV(f)h = -\big[Df(f^{-1})\big]^{-1} h(f^{-1}).$$

Note that if

$$h \in T\mathcal{D}(X)_f = \mathcal{C}^\infty(X, f^*TX)$$

and $Df\colon TX \to TX$ then

$$DV(f)h \in T\mathcal{D}(X)_{f^{-1}} = \mathcal{C}^\infty\big(X, (f^{-1})^*TX\big).$$

Again there are serious differentiability problems in Banach spaces, since Df occurs in the formula. If $\mathcal{D}^r(X)$ denotes the Banach manifold of diffeomorphisms of class C^r, then $V\colon \mathcal{D}^r(X) \to \mathcal{D}^r(X)$ is C^0 but not C^1, while $V\colon \mathcal{D}^{r+1}(X) \to \mathcal{D}^r(X)$ is C^1 but not C^2. In Fréchet spaces the map V is C^∞. Indeed we see that DV and hence TV involves nothing more complicated than derivatives, compositions and inverses, so by the chain rule the higher tangents $T^2V, \dots, T^nV$ all exist. Hence V is C^∞.

More generally, if X and Y are two distinct manifolds which are diffeomorphic, we can form the Fréchet manifold $\mathcal{D}(X, Y)$ of diffeomorphisms of X to Y, and form the inverse map

$$V\colon \mathcal{D}(X, Y) \to \mathcal{D}(Y, X)$$

as above.

4.4.7. EXAMPLE. We are now prepared to argue that the coordinate transition functions for the Fréchet manifold $\mathcal{S}(X)$ of smooth compact submanifolds of a finite dimensional manifold X are smooth, as we claimed in Example 4.1.7. Suppose we take coordinate charts of the type described there centered at two submanifolds S_1 and S_2, and that they overlap at a submanifold S. Then we can find a neighborhood U of S which is represented simultaneously as bundles over S_1 and S_2 using projections $\pi_1\colon U \to S_1$ and $\pi_2\colon U \to S_2$. The coordinate charts for $\mathcal{S}(X)$ at S_1 and S_2 are the same we used for the Fréchet manifolds of sections $\mathcal{C}^\infty(S_1, U)$ and $\mathcal{C}^\infty(S_2, U)$ of these bundles. Now these are submanifolds of the Fréchet manifolds of all maps $\mathcal{M}(S_1, U)$ and $\mathcal{M}(S_2, U)$ by Example 4.2.4. By composition with π_2 we get a map

$$\mathcal{C}^\infty(S_1, U) \subseteq \mathcal{M}(S_1, U) \to \mathcal{M}(S_1, S_2)$$

which is smooth by Example 4.4.5. The projections π_1 and π_2 induce diffeomorphisms $S \to S_1$ and $S \to S_2$, and the map in $\mathcal{M}(S_1, S_2)$ is just their composition $(\pi_2 \mid S) \circ (\pi_1 \mid S)^{-1}$ which is also a diffeomorphism. Hence the image of the above map lies in the open subset of diffeomorphisms $\mathcal{D}(S_1, S_2)$. On this subset the inverse map $V\colon \mathcal{D}(S_1, S_2) \to \mathcal{D}(S_2, S_1)$ is smooth by Example 4.4.6.

Composing the inverse map with the original section gives rise to a map

$$\mathcal{C}^\infty(S_1, U) \to \mathcal{D}(S_2, S_1) \times \mathcal{M}(S_1, U) \to \mathcal{M}(S_2, U) \supseteq \mathcal{C}^\infty(S_2, U)$$

which is smooth by Example 4.4.5. The image we get always lies in the submanifold $\mathcal{C}^\infty(S_2, U)$ of sections of U. This gives us our coordinate transition function, which is a composition of smooth maps and hence smooth.

4.4.8. DEFINITION. Let P: $\mathcal{M} \to \mathcal{N}$ be a smooth map between Fréchet manifolds. We say P is an *immersion* if for any $f \in \mathcal{M}$ we can choose coordinate charts around $f \in \mathcal{M}$ and $P(f) \in \mathcal{N}$ such that the local representative of P is the inclusion of a factor in a direct sum. We say P is a *submersion* if we can choose the local charts so that the representatives of P is the projection onto a factor in a direct sum.

4.4.9. EXAMPLES. The inclusion of a submanifold is an immersion. The projection map of a vector bundle is a submersion.

Let $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{N}$ be Fréchet manifolds and let

$$P_1\colon \mathcal{M}_1 \to \mathcal{N}, \qquad P_2\colon \mathcal{M}_2 \to \mathcal{N}$$

be smooth maps. We define the fibre product

$$\mathcal{M}_1 \times_{\mathcal{N}} \mathcal{M}_2 = \{(f_1, f_2) \in \mathcal{M}_1 \times \mathcal{M}_2 \colon P_1(f_1) = P_2(f_2)\}.$$

4.4.10. THEOREM. *If P_1 and P_2 are submersions then $\mathcal{M}_1 \times_{\mathcal{N}} \mathcal{M}_2$ is a closed submanifold of $\mathcal{M}_1 \times \mathcal{M}_2$.*

PROOF. Let Δ be the diagonal in $\mathcal{N} \times \mathcal{N}$. Then $\mathcal{M}_1 \times_{\mathcal{N}} \mathcal{M}_2 = (P_1 \times P_2)^{-1}\Delta$ is closed. If $(f_1, f_2) \in \mathcal{M}_1 \times_{\mathcal{N}} \mathcal{M}_2$ then $P_1(f_1) = P_2(f_2)$. We can find coordinate charts $U_1 \times V_1 \to \mathcal{M}_1$ and $V_1 \to \mathcal{N}$ such that P_1 is the projection on V_1, and we can find coordinate charts $U_2 \times V_2 \to \mathcal{M}_2$ and $V_2 \to \mathcal{N}$ such that P_2 is the projection on V_2. Without loss of generality we can take $V_1 = V_2 = V$. Then $U_1 \times V \times U_2 \times V$ gives a chart on $\mathcal{M}_1 \times \mathcal{M}_2$ in which the fibre product corresponds to points over the diagonal Δ in $V \times V$. Thus the fibre product is a submanifold.

4.5. *Connections.* Let $\mathcal{V}$ be a Fréchet vector bundle over a Fréchet manifold $\mathcal{M}$ with projection π. The null space of $D\pi$ is the subspace of vertical tanget vectors to $\mathcal{V}$, and it is naturally isomorphic at each point of $\mathcal{V}$ to the fibre of $\mathcal{V}$ passing through that point. It is natural to wish to pick a complementary subspace of horizontal tangent vectors.

4.5.1. DEFINITION. A *connection* on $\mathcal{V}$ is a rule which assigns to each point in $\mathcal{V}$ a complementary subspace of horizontal vectors, such that in terms of any coordinate chart on the bundle with values in $(U \subseteq F) \times G$ the subspace of horizontal vectors consists of all $(h, k) \in F \times G$ with

$$k = \Gamma(f)\{g, h\}$$

where the local representative Γ of the connection is a smooth map

$$\Gamma\colon (U \subseteq F) \times G \times F \to G$$

which is bilinear in g and h.

One easy way to visualize a connection is with parametrized curves. Given a path $v(t) \in \mathcal{V}$ we can form its covariant derivative $v'(t) \in \mathcal{V}$ using the connection. In local coordinates if

$$v(t) = (f(t), g(t))$$

then

$$v'(t) = (h(t), k(t))$$

where

$$h(t) = f'(t), \qquad k(t) = g'(t) - \Gamma(f(t))\{g(t), h(t)\}.$$

The path $v(t)$ is horizontal if $v'(t) = 0$. The path $v(t) \in \mathcal{V}$ covers the path $f(t) \in \mathfrak{M}$. In Banach spaces, given the path $f(t) \in \mathfrak{M}$ for $a \leqslant t \leqslant b$ and the initial value $v(a)$ there is a uniquely determined horizontal lifting $v(t)$ covering $f(t)$ with $v'(t) = 0$. In Fréchet spaces the horizontal lift may not exist, and if it does it may not be unique. This is because the usual existence and uniqueness theorems for ordinary differential equations fail in Fréchet spaces.

A connection on a manifold $\mathfrak{M}$ is defined to be a connection on its tangent bundle $T\mathfrak{M}$. If $f(t)$ is a path in $\mathfrak{M}$ giving the position as a function of time, then its velocity $f'(t)$ is a path in $T\mathfrak{M}$. The connection allows us to define the acceleration $f''(t)$ also as a path in $T\mathfrak{M}$. We say $f(t)$ is a geodesic if its acceleration is zero. We say a connection on $T\mathfrak{M}$ is symmetric if its local representative $\Gamma(f)\{h, k\}$ is symmetric in h and k.

4.5.2. DEFINITION. The curvature of a connection on a vector bundle $\mathcal{V}$ is the trilinear map

$$\mathcal{V} \times T\mathfrak{M} \times T\mathfrak{M} \to \mathcal{V}$$

given locally by

$$\begin{aligned} R(f)\{g, h, k\} = {} & D\Gamma(f)\{g, h, k\} - D\Gamma(f)\{g, k, h\} \\ & -\Gamma(f)\{\Gamma(f)\{g, h\}, k\} + \Gamma(f)\{\Gamma(f)\{g, k\}, h\} \end{aligned}$$

where $\Gamma(f)\{g, h\}$ is the local representative of the connection. The curvature is independent of the choice of a chart.

4.5.3. EXAMPLE. Let X and Y be finite dimensional manifolds with X compact and let $\mathfrak{M}(X, Y)$ be the Fréchet manifold of smooth maps of X into Y. A path $f(t) \in \mathfrak{M}(X, Y)$ can be evaluated at any point $x \in X$ to give a path $f(t, x) \in Y$. Suppose Y has a connection given in a local chart by $\Gamma(y)\{z, w\}$. Then $\mathfrak{M}(X, Y)$ has a connection given in local coordinates by

$$[\Gamma(f)\{h, k\}](x) = \Gamma(f(x))\{h(x), k(x)\},$$

where for each $x \in X$ we have $f(x) \in Y$ and $h(x), k(x) \in TY_{f(x)}$. A path $f(t)$ is a geodesic in $\mathfrak{M}$ if and only if the path $f(t, x)$ for any $x \in X$ is a geodesic in Y. In this case each path will have a unique lift obtained by lifting point by point in X. The curvature of the connection on $\mathfrak{M}(X, Y)$ is given in terms of the curvature of the connection on Y by

$$[R(f)\{g, h, k\}](x) = R(f(x))\{g(x), h(x), k(x)\}$$

that is, evaluating point by point.

4.5.4. EXAMPLE. Let $\mathcal{R}(X)$ be the Fréchet manifold of all compact domains with smooth boundary in a finite dimensional manifold X, and let $\mathcal{C}^\infty\mathcal{R}(X)$ be the Fréchet vector bundle whose fibre over a domain $A \in \mathcal{R}(X)$ is the Fréchet space $\mathcal{C}^\infty(A)$ of all smooth functions on A. There is a natural connection on $\mathcal{C}^\infty\mathcal{R}(X)$ of great importance in free boundary value problems. Fix a function $f \in \mathcal{C}^\infty(X)$ and choose a path of domains $A(t)$ in $\mathcal{R}(X)$. Then $f(t) = f|A(t)$ defines a path in $\mathcal{C}^\infty\mathcal{R}(X)$ by restricting the fixed function to the variable domain. The natural connection on $\mathcal{C}^\infty\mathcal{R}(X)$ is the one which makes all of these paths horizontal.

To see that this does indeed give a connection, we shall evaluate its local representative in the coordinate charet of Example 4.3.6. There a function $b \in \mathcal{C}^\infty(\partial A)$ corresponds to a domain $\tilde{A}$ with $r \geqslant b(y)$, and a function $f \in \mathcal{C}^\infty(A)$ corresponds to a function $g \in \mathcal{C}^\infty(\tilde{A})$ by the rule $f(y, r) = g(y, s)$ where

$$s = r + \phi(r)b(y).$$

Then a path $\tilde{A}(t)$ in $\mathcal{R}(x)$ will correspond to a path $b(x, t)$ in $\mathcal{C}^\infty(\partial A)$ and a path $f(x, t)$ in $\mathcal{C}^\infty(A)$ will correspond to a path $g(x, t)$ in $\mathcal{C}^\infty(\tilde{A}(t))$ by the rule

$$f(y, r, t) = g(y, s, t), \qquad s = r + \phi(r)b(y, t).$$

Differentiating in t we see that

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial r}\frac{\partial r}{\partial t} = \frac{\partial g}{\partial t}, \qquad \frac{\partial s}{\partial t} = \frac{\partial r}{\partial t} + \frac{\partial \phi}{\partial r}\frac{\partial r}{\partial t}b + \phi\frac{\partial b}{\partial t}.$$

For the path g to be horizontal in our connection we want $\partial g/\partial t = 0$ when $\partial s/\partial t = 0$. Solving, we get

$$\frac{\partial f}{\partial t} = \phi\frac{\partial f}{\partial r}\frac{\partial b}{\partial t}\Big/\left[1 + b\frac{\partial \phi}{\partial r}\right].$$

Now at the domain $\tilde{A}(t)$ given by $b(t)$ the vector $h = \partial b/\partial t$ represents the tangent vector to the path. Therefore the local representative of the connection is given by

$$\Gamma\colon (U \subseteq \mathcal{C}^\infty(\partial)) \times \mathcal{C}^\infty(A) \times \mathcal{C}^\infty(\partial A) \to \mathcal{C}^\infty(A)$$

where

$$\Gamma(b)\{f, h\} = \phi(\partial f/\partial r) \cdot h/\left[1 + b\partial\phi/\partial r\right]$$

on $N \cap A$, and outside of N we have $\phi = 0$ so $\Gamma = 0$. Since Γ is smooth and bilinear in f and h it defines a connection locally. The invariance under a change of coordinates follows immediately from its invariant description.

Any smooth function on a compact region A with smooth boundary extends to a smooth function on X, but not uniquely. It follows that any path in $\mathcal{R}(x)$ can be lifted to $\mathcal{C}^\infty\mathcal{R}(x)$ with given initial value, and the lift will be unique for a path of shrinking domains but not for a path of expanding domains. By way of contrast, we note that there is another interesting bundle $\mathcal{C}_0^\infty\mathcal{R}(x)$ whose fibre over a domain $A \in \mathcal{R}(x)$ is the Fréchet space $\mathcal{C}_0^\infty(A)$ of smooth functions on X which vanish identically outside of A. The previous formula defines a connection Γ on this subbundle also, for when f vanishes outside A so does

$\partial f/\partial r$. For this bundle the horizontal lift over a path of expanding domains exists and is unique, but the lift over a path of shrinking domains may not exist.

Finally we note that the connection Γ (for either bundle) has curvature zero. This can be shown by an explicit calculation. Differentiating $\Gamma(b)\{f, h\}$ with respect to b in the direction k gives

$$D\Gamma(b)\{f, h, k\} = -\phi \frac{\partial \phi}{\partial r} \frac{\partial f}{\partial r} hk / \left[1 + b \frac{\partial \phi}{\partial r}\right]^2$$

which is symmetric in h and k. Moreover since b, h, and k are functions of y alone they are independent of r, so $\partial h/\partial r = 0$ and

$$\begin{aligned} \Gamma(b)\{\Gamma(b)\{f, h\}, k\} &= \phi \frac{\partial}{\partial r} [\Gamma(b)\{f, h\}] \cdot k / \left[1 + b \frac{\partial \phi}{\partial r}\right] \\ &= \phi \frac{\partial}{\partial r} \left[\phi \frac{\partial f}{\partial r} \left(1 + b \frac{\partial \phi}{\partial r}\right)\right] \cdot hk / \left[1 + b \frac{\partial \phi}{\partial r}\right] \end{aligned}$$

is also symmetric in h and k. Therefore the curvature is zero. Of course it is obvious that $\mathcal{C}^\infty \mathcal{R}(X)$ must have curvature zero since there exists a horizontal section passing through every point $f \in \mathcal{C}^\infty \mathcal{R}(X)$, even though the horizontal section is not unique.

4.5.5. EXAMPLE. Let X be a Riemannian manifold, $\mathcal{S}(X)$ the Fréchet manifold of all compact smooth submanifolds of X, and $\mathcal{C}^\infty \mathcal{S}(X)$ the Fréchet vector bundle over $\mathcal{S}(X)$ of all smooth functions on compact smooth submanifolds, so that the fibre over $S \in \mathcal{S}(X)$ is the Fréchet vector space $\mathcal{C}^\infty(S)$. For each $S \in \mathcal{S}(X)$ we can form the tangent bundle TS and the normal bundle $NS = TM/TS$ over S. Likewise for each $f \in \mathcal{C}^\infty(S)$ we can define vector bundles Tf and Nf over S by letting Tf be the graph of Df,

$$Tf = \{(v, y) \in TM \times R : v \in TS \quad \text{and } y = Df(v)\}$$

and

$$Nf = TM \times R/Tf.$$

THEOREM. *There are natural isomorphisms*

$$T_S \mathcal{S}(X) = \mathcal{C}^\infty(S, NS), \qquad T_f \mathcal{C}^\infty \mathcal{S}(X) = \mathcal{C}^\infty(S, Nf).$$

PROOF. We can identify the tangent space to a manifold by identifying the tangent vectors to paths. Let $S_* = \{S_t\}$ be a smooth path in $\mathcal{S}(X)$ for $-\varepsilon < t < \varepsilon$ with $S_0 = S$. Then S_* is itself a compact smooth submanifold of $M \times (-\varepsilon, \varepsilon)$ transversal to each slice $t = \text{constant}$. If $f_* = \{f_t\}$ is a smooth path in $\mathcal{C}^\infty \mathcal{S}(X)$ for $-\varepsilon < t < \varepsilon$ with $f_0 = f$ and covering the path $S_* = \{S_t\}$ in $\mathcal{S}(X)$, then each $f_t \in \mathcal{C}^\infty(S_t)$ and f_* is a smooth function on S_*.

The path S_* defines an affine subbundle of TM over S by

$$\widetilde{TS} = \left\{v \in TM : (v, 1) \in TS_* \text{ at } t = 0\right\}$$

since $TS_* \subseteq TM \times R$ when $S_* \subseteq M \times (-\epsilon, \epsilon)$. Similarly the path f_* defines an affine subbundle of $TM \times R$ over S by

$$\widetilde{Tf} = \{(v, y) \in TM \times R\colon (v, 1) \in TS_* \quad \text{at } t = 0 \text{ and } y = Df_*(v, 1)\}.$$

It is easy to check that $\widetilde{TS}$ is a translate of TS in TM, and hence determines a section over S of the quotient bundle $NS = TM/TS$. This identifies the tangent to the path S_* as an element of $\mathcal{C}^\infty(S, NS)$. Similarly we can check that Tf is a translate of Tf in $TM \times R$, and hence determines a section over S of the quotient bundle $Nf = TM \times R/Tf$. This identifies the tangent to the path f_* as an element of $\mathcal{C}^\infty(S, Nf)$ and proves the theorem

There is an obvious short exact sequence of vector bundles over S, $0 \to R \to Nf \to NS \to 0$, which arises from the commutative diagram

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & 0 & \to & Tf & \to & TS & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & R & \to & TM \times R & \to & TM & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & R & \to & Nf & \to & NS & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

where the vector bundle isomorphism $Tf \to TS$ is obtained by forgetting the second factor in $TM \times R$. This induces a short exact sequence of Fréchet vector spaces $0 \to \mathcal{C}^\infty(S) \to \mathcal{C}^\infty(S, Nf) \to \mathcal{C}^\infty(S, NS) \to 0$ which by the previous theorem becomes

$$0 \to \mathcal{C}^\infty(S) \to T_f\mathcal{C}^\infty\mathcal{S}(X) \to T_s\mathcal{S}(X) \to 0.$$

The second map is clearly $T\pi$ where π is the projection of the vector bundle. Hence the first factor $\mathcal{C}^\infty(S)$ is the subbundle of vertical tangent vectors to the bundle $\mathcal{C}^\infty\mathcal{S}(X)$ at $f \in \mathcal{C}^\infty(S)$. These are of course just the variations where we fix the submanifold S and vary only the function f on S. They are intrinsically defined. Unlike the case of $\mathcal{R}(X)$ there is now no intrinsic way to fix the function f and vary the domain S. To do this requires the choice of a connection Γ on $\mathcal{C}^\infty\mathcal{S}(X)$, for which we need some additional data.

Suppose that X is endowed with a Riemannian metric g. Then there is a natural way to define a connection Γ on $\mathcal{C}^\infty\mathcal{S}(X)$ in terms of g. The geometric idea is that to vary the domain S while keeping the function f fixed we should keep f constant along the normal direction. It is the choice of the normal direction that requires a metric. The connection Γ may be described in the following way. We can identify the normal bundle NS with the subbundle of the bundle TM over S of vectors perpendicular to TS in the metric g. Then $NS \times R$ is a subspace of $TM \times R$ complementary to Tf, so we can identify $Nf = TM \times R/Tf$ with $NS \times R$. If $Nf = NS \times R$ we take $Hf = NS \times \{0\}$ as the horizontal subspace complementary to the vertical subspace $\{0\} \times R$. Then $\mathcal{C}^\infty(S, Hf)$ will be the horizontal subspace for the connection Γ of

$\mathcal{C}^\infty(S, Nf) = T_f\mathcal{C}^\infty\mathcal{S}(X)$ complementary to the vertical subspace $\mathcal{C}^\infty(S)$ and projecting isomorphically onto $\mathcal{C}^\infty(S, NS) = T_f\mathcal{S}(X)$ since

$$Hf = NS \times \{0\} \simeq NS.$$

It is interesting to compute the curvature of this connection. The metric g on X induces metrics on the tangent bundle TS and the normal bundle NS and covariant derivatives on sections of these bundles. If $f \in \mathcal{C}^\infty(S)$ and $h, k \in \mathcal{C}^\infty(S, NS)$ we can form the covariant derivatives ∇f and ∇h. Then using the etric on T^*S we can form the dot product $\nabla f \cdot \nabla h$ in NS, and using the metric in NS we can form another dot product $\nabla f \cdot \nabla h \cdot k$. The curvature at $S \in \mathcal{S}(X)$ is the trilinear map

$$R(S)\colon \mathcal{C}^\infty\mathcal{S}(X)_s \times T\mathcal{S}(X)_s \times T\mathcal{S}(X)_s \to \mathcal{C}^\infty\mathcal{S}(X)_s$$

which in terms of our previous identifications is a map

$$\mathcal{R}(S)\colon \mathcal{C}^\infty(S) \times \mathcal{C}^\infty(S, NS) \times \mathcal{C}^\infty(S, NS) \to \mathcal{C}^\infty(S)$$

given by

$$R(S)\{f, h, k\} = \nabla f \cdot \nabla h \cdot k - \nabla f \cdot \nabla k \cdot h,$$

We leave it as an exercise for the reader to check the formula. It is interesting to note that the curvature of this natural connection on $\mathcal{C}^\infty\mathcal{S}(X)$ is not zero, even when X is flat, and even when S is a flat submanifold of X (such as a circle in a flat torus).

4.6. *Lie groups.*

4.6.1. DEFINITION. A Fréchet Lie group is a Fréchet manifold $\mathcal{G}$ with a group structure such that the multiplication map C and the inverse map V,

$$C\colon \mathcal{G} \times \mathcal{G} \to \mathcal{G}, \quad C(g, h) = gh,$$
$$V\colon \mathcal{G} \to \mathcal{G}, \quad V(g) = g^{-1},$$

are smooth.

4.6.2. EXAMPLE. Let X be a compact manifold. Then the diffeomorphism group $\mathcal{D}(X)$ is a Fréchet Lie group.

The Lie algebra of $\mathcal{G}$ is the tangent space $G = T_1\mathcal{G}$ at the identity 1. Each element in G determines a left-invariant vector field on $\mathcal{G}$. The Lie bracket of two left-invariant vector fields is again left-invariant. This defines the Lie algebra structure of $\mathcal{G}$. There is a unique connection on $\mathcal{G}$ such that the left-invariant vector fields are horizontal. For this connection the curvature is zero and the torsion is given by the Lie bracket.

4.6.3. EXAMPLE. For the diffeomorphism group $\mathcal{D}(X)$ the Lie algebra $T_1\mathcal{D}(X)$ is the space $\mathcal{C}^\infty(X, TX)$ of vector fields on X. The Lie algebra of the group is given by the Lie bracket of the vector fields.

4.6.4. DEFINITION. We say the Lie group $\mathcal{G}$ acts on a Fréchet manifold $\mathcal{M}$ if there is a smooth map

$$A\colon \mathcal{G} \times \mathcal{M} \to \mathcal{M}, \quad A(g, h) = gh$$

such that $1h = h$ and $(g_1g_2)h = g_1(g_2h)$.

4.6.5. DEFINITION. Let $\mathcal{G}$ be a Fréchet Lie group. A *principal Fréchet $\mathcal{G}$-bundle* consists of a Fréchet manifold $\mathcal{M}$ (called the total space), an action α of $\mathcal{G}$ on

$\mathfrak{M}$, and a submersion P of $\mathfrak{M}$ onto another Fréchet manifold $\mathfrak{B}$ (called the base), such that for each $b \in \mathfrak{B}$ we can find a neighborhood V of b and a diffeomorphism P of $\mathcal{G} \times V$ onto $P^{-1}(V) \subseteq \mathfrak{M}$ such that:

(1) The action A of $\mathcal{G}$ on $\mathfrak{M}$ corresponds to the action on $\mathcal{G} \times V$ on the first factor, and

(2) the projection P onto $\mathfrak{B}$ corresponds to the projection of $\mathcal{G} \times V$ onto the second factor.

4.6.6. EXAMPLE. Let X and Y be finite dimensional manifolds with X compact. Let $\mathcal{D}(X)$ be the diffeomorphism group of X, let $\mathcal{E}(X, Y)$ be the Fréchet manifold of all embeddings of X into Y, and let $\mathcal{S}_x(Y)$ be the component of the Fréchet manifold of all submanifolds of Y consisting of those which are diffeomorphic to X. Then $\mathcal{E}(X, Y)$ is a principal $\mathcal{D}(X)$ bundle over $\mathcal{S}_x(Y)$. The action of $\mathcal{D}(X)$ on $\mathcal{E}(X, Y)$ is given by composition of the embedding and the diffeomorphism. The projection of $\mathcal{E}(X, Y)$ onto $\mathcal{S}_x(Y)$ assigns to each embedding its image. The argument in Example 4.4.7 shows how to construct the coordinate charts required.

I.5. The inverse function theorem.

5.1. *Estimates.* Let L: $H \to K$ be a continuous linear map between Fréchet spaces. Then for every seminorm $\| \ \|_K$ on K there exists a seminorm $\| \ \|_H$ on H such that $\|Lh\|_K \leqslant C\|h\|_H$. More generally we have the following result.

5.1.1. THEOREM. *Let L: $(U \subseteq F) \times H \to K$ be a family of linear maps on Fréchet spaces. If L is continuous, then for every $f_0 \in U$ and every seminorm $\| \ \|_K$ on K we can find a neighborhood $\tilde{U}$ of f_0 and a seminorm $\| \ \|_H$ on H and a constant C such that for all $f \in \tilde{U}$ and $h \in H$*

$$\|L(f)h\|_K \leqslant C\|h\|_H.$$

PROOF. The seminorms define the topology. Since L is continuous and $L(f_0)0 = 0$, we can find a neighborhood $\tilde{U}$ of f_0 and a neighborhood of 0 of the form $\|h\|_H \leqslant \varepsilon$ on which $\|L(f)h\|_K \leqslant 1$. Given any $h \in H$ we put $\tilde{h} = \varepsilon h/\|h\|_H$. Then $\|\tilde{h}\|_H \leqslant \varepsilon$ and if $f \in \tilde{U}$ we have $\|L(f)h\|_K \leqslant 1$. It follows by linearity in h that $\|L(f)h\|_K \leqslant C\|h\|_H$ with $C = 1/\varepsilon$.

5.1.2. THEOREM. *Let B: $(U \subseteq F) \times H \times K \to L$ be a family of bilinear maps on Fréchet spaces, so that $B(f)\{h, k\}$ is linear separately in h and k. If B is continuous, then for every $f_0 \in U$ and every seminorm $\| \ \|_L$ on L we can find seminorms $\| \ \|_H$ and $\| \ \|_K$ on H and K and a constant C such that for all $f \in \tilde{U}$ and all $h \in H$ and $k \in K$*

$$\|B(f)\{h, k\}\|_L \leqslant C\|h\|_H\|k\|_K.$$

PROOF. The demonstration is the same as for the preceding result.

5.1.3. THEOREM. *Let P: $U \subseteq F \to G$ be a smooth map of Fréchet spaces. Then for every $f_0 \in U$ and every seminorm $\| \ \|_G$ on G we can find a seminorm $\| \ \|_F$ on F, an $\varepsilon > 0$, and a constant C such that if $\|f_1 - f_0\|_F \leqslant \varepsilon$ and $\|f_2 - f_0\|_F \leqslant \varepsilon$ then*

$$\|P(f_1) - P(f_2)\|_G \leqslant C\|f_1 - f_2\|_F.$$

PROOF. The derivative $DP(f)h$ is continuous and linear in h. Given f_0 and $\| \ \|_G$ we can find a neighborhood $\tilde{U}$ of f_0 and a $\| \ \|_F$ with $\|DP(f)h\|_G \leq C\|h\|_F$ if $f \in \tilde{U}$. The neighborhood $\tilde{U}$ will contain a neighborhood of the form $\|f - f_0\|_F \leq \varepsilon$ for some $\| \ \|_F$ and some $\varepsilon < 0$. If this norm is different from the preceding, we can pick a third norm $\| \ \|_F$ stronger than both (for example, their sum). By the fundamental theorem of calculus

$$P(f_1) - P(f_2) = \int_0^1 DP(tf_1 + (1-t)f_2)(f_1 - f_2)\, dt.$$

If f_1 and f_2 belong to the set $\|f - f_0\|_F \leq \varepsilon$ so does the line segment joining them, so

$$\|DP(tf_1 + (1-t)f_2)(f_1 - f_2)\|_G \leq C\|f_1 - f_2\|_F.$$

Then $\|P(f_1) - P(f_2)\|_G \leq C\|f_1 - f_2\|_F$.

Next we prove a more specialized result.

5.1.4. THEOREM. *Let Q: $U \subseteq F \to G$ be a smooth map between Fréchet spaces. Suppose that $Q(0) = 0$ and $DQ(0)h = 0$ for all $h \in F$. Then for every seminorm $\| \ \|_G$ on G we can find a seminorm $\| \ \|_F$ on F, and $\varepsilon > 0$, and a constant C such that if $\|f_1\|_F \leq \varepsilon$ and $\|f_2\|_F \leq \varepsilon$ then*

$$\|Q(f_1) - Q(f_2)\|_G \leq C(\|f_1\|_F + \|f_2\|_F)\|f_1 - f_2\|_F.$$

PROOF. The second derivative $D^2Q(f)\{h, k\}$ is a smooth family of bilinear maps. By Theorem 5.1.2. for any $\| \ \|_G$ on G we can find a neighborhood of zero $\tilde{U} \subseteq F$ and a $\| \ \|_F$ on F such that if $f \in \tilde{U}$

$$\|D^2Q(f)\{h, k\}\|_G \leq C\|h\|_F\|k\|_F$$

and as before we may as well take $\tilde{U}$ of the form $\{f\colon \|f\|_F \leq \varepsilon\}$ for some $\varepsilon > 0$. By the fundamental theorem of calculus

$$DQ(f)h = \int_0^1 D^2Q(tf)\{f, h\}\, dt$$

since $DQ(0)h = 0$. Then if $\|f\|_F \leq \varepsilon$,

$$\|DQ(f)h\|_G \leq C\|f\|_F\|h\|_F.$$

We can also write

$$P(f_1) - P(f_2) = \int_0^1 DP(tf_1 + (1-t)f_2)(f_1 - f_2)\, dt$$

as before. If f_1 and f_2 belong to the ball $\tilde{U}$ so does the line segment joining them. Then

$$\|DP(tf_1 + (1-t)f_2)(f_1 - f_2)\|_G \leq C(\|f_1\|_F + \|f_2\|_F)\|f_1 - f_2\|_F$$

and the norm of the integral is bounded by the integral of the norm. This proves the estimate.

5.2. *The inverse function theorem for Banach spaces.* Let X be a metric space and P: $X \to X$ a map of X into itself. We say P is a *contraction* if there exists a $\rho < 1$ such that

$$d(Px, Py) \leq \rho d(x, y).$$

5.2.1. THEOREM. *If X is a complete metric space and P: $X \to X$ is a contraction then P has a unique fixed point x with $P(x) = x$.*

PROOF. Recall that this well-known result is proved by the following trick. Start with any $x_0 \in X$ and define a sequence $x_n \in X$ by $x_n = P(x_{n-1})$. Then $d(x_{n+1}, x_n) \leqslant \rho^n d(x_1, x_0)$. Since the geometric series $\Sigma\rho^n$ converges for $\rho < 1$, it follows that the sequence x_n is Cauchy, and if X is complete then x_n converges to an element x with $P(x) = x$. This shows the fixed point exists. If x_1 and x_2 are two fixed points then

$$d(x_1, x_2) = d(P(x_1), P(x_2)) \leqslant \rho d(x_1, x_2)$$

and if $\rho < 1$ this forces $d(x_1, x_2) = 0$. Hence the fixed point is unique.

5.2.2. COROLLARY. *Let X and Y be metric spaces with X complete and let P: $X \times Y \to X$ be a continuous map with*

$$d(P(x_1, y), P(x_2, y)) \leqslant \rho d(x_1, x_2)$$

for some $\rho < 1$. Then for every $y \in Y$ there exists a unique $x \in X$ with $P(x, y) = x$. If we let $x = S(y)$ then the map S: $Y \to X$ is continuous.

PROOF. We already know all but the continuity of S. Pick $x_0 \in X$ and let $x_{n+1} = P(x_n, y)$. Then $x_n = S_n(y)$ is a continuous map S_n: $Y \to X$, and $S_n(y) \to S(y)$ as $n \to \infty$. Since

$$d(S_{n+1}(y), S_n(y)) \leqslant C(y)\rho^n$$

where $C(y) = d(S_1(y), S_0(y))$ is a continuous function of y, it follows that the sequence $S_n(y)$ converges to $S(y)$ uniformly on a neighborhood of any $y_0 \in Y$. Therefore $S(y)$ is continuous.

We can now prove the inverse function theorem for Banach spaces.

5.2.3. THEOREM. *Let P: $U \subseteq F \to V \subseteq G$ be a smooth map between Banach spaces. Suppose that for some $f_0 \in U$ the derivative $DP(f_0)$: $F \to G$ is an invertible linear map. Then we can find neighborhoods $\tilde{U}$ of f_0 and $\tilde{V}$ of $g_0 = P(f_0)$ such that the map P gives a one-to-one map of $\tilde{U}$ onto $\tilde{V}$, and the inverse map P^{-1}: $\tilde{V} \subseteq G \to \tilde{U} \subseteq F$ is continuous. (We shall show later in Corollary 5.3.4. that P^{-1} is smooth.)*

PROOF. By replacing f by $\tilde{f} = f - f_0$ and g by $\tilde{g} = g - g_0$ we may assume that $f_0 = 0$ and $g_0 = 0$. Since $DP(0)$ is an isomorphism between F and G, we may use it to identify F and G, and assume that $DP(0)$ is the identity.

Let $Q(f) = f - P(f)$. Then $Q(0) = 0$ and $DQ(0)h = h - DP(0)h = 0$, so we may apply Theorem 5.1.4. Since F is a Banach space all norms are equivalent, and we pick one $\| \ \|$. Then

$$\|Q(f_1) - Q(f_2)\| \leqslant C(\|f_1\| + \|f_2\|)\|f_1 - f_2\|$$

if $\|f_1\| \leqslant \varepsilon$ and $\|f_2\| \leqslant \varepsilon$.

Putting $f_1 = f$ and $f_2 = 0$ we also have

$$\|Q(f)\| \leqslant C\|f\|^2$$

when $\|f\| \leqslant \varepsilon$. We put

$$R(f, g) = f - P(f) + g = Q(f) + g$$

and let

$$X = \{f: \|f\| \leqslant \varepsilon\} \quad \text{and} \quad Y = \{g: \|g\| \leqslant \delta\}.$$

Then X is a complete metric space. If $f \in X$ and $g \in Y$ then

$$\|R(f, g)\| \leqslant C\|f\|^2 + \|g\| \leqslant C\varepsilon^2 + \delta.$$

If $\varepsilon \leqslant 1/2C$ and $\delta \leqslant \varepsilon/2$ then $C\varepsilon^2 + \delta \leqslant \varepsilon$, so $R(f, g) \in X$. Thus R defines a map R: $X \times Y \to X$. Moreover $R(f_1, g) - R(f_2, g) = Q(f_1) - Q(f_2)$ so if $f_1, f_2 \in X$ and $g \in Y$

$$\|R(f_1, g) - R(f_2, g)\| \leqslant 2C\varepsilon\|f_1 - f_2\|.$$

When ε is small then $\rho = 2C\varepsilon < 1$ and R is a contradiction. To make everything work we first choose ε small compared to C, and then choose δ small compared to ε. Then it follows that from Corollary 5.2.2 for each $g \in Y$ there is a unique $f \in X$ with $R(f, g) = f$, which means $P(f) = g$. Moreover the map P^{-1}: $Y \to X$ is continuous. To finish the proof we take any open neighborhood of zero $\tilde{V} \subseteq \tilde{Y}$ and let $\tilde{U} = P^{-1}(\tilde{V})$. Since P^{-1} is continuous $\tilde{U}$ is open.

5.3. *Inverses of linear maps.* If L: $F \to G$ is a continuous linear map between Fréchet spaces and if L is invertible then L^{-1}: $G \to F$ is continuous by the open mapping theorem. If L: $(U \subseteq F) \times H \to K$ is a family of linear maps, so that $L(f)h = k$ is linear in $h \in H$ for each $f \in U \subseteq F$, and if each $L(f)$: $H \to K$ is invertible, then putting $V(f) = L(f)^{-1}$ we obtain the family of inverses V: $(U \subseteq F) \times K \to H$ where $V(f)k = h$ is the unique solution of the equation $L(f)h = k$.

5.3.1. THEOREM. *Let L: $(U \subseteq F) \times H \to K$ be a family of invertible linear maps of Fréchet spaces and let V: $(U \subseteq F) \times K \to H$ be the family of inverses. If L is smooth and V is continuous then V is smooth and*

$$DV(f)\{k, g\} = -V(f)DL(f)\{V(f)k, g\}.$$

PROOF. Recall that $DV(f)\{k, g\}$ is obtained by differentiating $V(f)k$ with respect to f in the direction g. Now the difference quotient

$$[V(f + tg)k - V(f)k]/t$$
$$= -V(f + tg)[L(f + tg)V(f)k - L(f)V(f)k]/t.$$

As $t \to 0$

$$[L(f + tg)V(f)k - L(f)V(f)k]/t \to DL(f)\{V(f)k, g\}$$

since L is C^1. If V is C^0 then by its joint continuity

$$V(f + tg)[L(f + tg)V(f)k - L(f)V(f)k]/t \to V(f)DL(f)\{V(f), g\}.$$

It follows that the difference quotient for V converges to the given limit. We see from the formula for DV that V is C^1. It now follows by the chain rule and induction that if V is C^r and L is C^{r+1} then V is C^{r+1}. If L is C^∞ then V is C^∞.

For Banach spaces there is an easy criterion for inverting a family of linear maps, since the set of invertible linear maps of H to K is open in the Banach space $L(H, K)$. We prove it in a different way.

5.3.2. THEOREM. *Let F, H and K be Banach spaces. Let L: $(U \subseteq F) \times H \to K$ be a smooth family of linear maps $L(f)h = k$ linear in h. Suppose that for some $f_0 \in U$ the map $L(f_0)$: $H \to K$ is invertible. Then there is a neighborhood $\tilde{U}$ of f_0 such that $L(f)$: $H \to K$ is invertible for all $f \in \tilde{U}$, and the family of inverses V: $(U \subseteq F) \times K \to H$ is smooth.*

PROOF. We can assume $f_0 = 0$ by replacing f by $\tilde{f} = f - f_0$. We can also assume $L(f_0)$ is the identity by identifying K with H under the isomorphism $L(f_0)^{-1}$. Then we have a map $L(f)h = k$ with $L(0)h = h$. Consider the map

$$P\colon (U \subseteq F) \times H \to F \times H$$

defined by

$$P(f, h) = (f, L(f)h).$$

Then

$$DP(f, h)(\tilde{f}, \tilde{h}) = \big(\tilde{f}, DL(f)\{h, \tilde{f}\} + L(f)\tilde{h}\big)$$

and $DP(0,0)(\tilde{f}, \tilde{h}) = (\tilde{f}, \tilde{h})$ is the identity. It follows that on a neighborhood $\tilde{U} \times \tilde{W}$ of $(0,0)$ the map P is invertible, and $L(f)h = k$ has a unique solution for all $f \in \tilde{U}$ and $k \in \tilde{W}$, which we write as $V(f)k = h$. But since $L(f)h = k$ is linear in h, it follows that $V(f)k = h$ is defined and continuous for $f \in \tilde{U}$ and all $k \in \tilde{K}$. That V is smooth follows from Theorem 5.4.1.

5.3.3. COUNTEREXAMPLE. Let $\mathcal{C}^\infty_{2\pi}$ be the Fréchet space of functions $f(x)$ which are smooth and periodic with period 2π. Define a smooth family of linear maps

$$L\colon R \times \mathcal{C}^\infty_{2\pi} \to \mathcal{C}^\infty_{2\pi}$$

by setting

$$[L(t)f](x) = \frac{1}{t}\int_x^{x+t} f(u)\,du$$

so that the value of $L(t)f$ at x is the average value of f over the interval starting at x of length t. If we make the change of variables $u = x + tv$ we get the alternative formula

$$[L(t)f](x) = \int_0^1 f(x + tv)\,dv.$$

Clearly $L(0)$ is the identity and hence invertible. But if n is an integer

$$L(2\pi/n)\sin nx = 0.$$

Indeed $L(t)$ acts term by term on the Fourier series of $f(x)$ and when $t = 2\pi/n$ it kills the terms $\sin nx$ and $\cos nx$. Thus $L(2\pi/n)$ is not invertible. Since $2\pi/n \to 0$ as $n \to \infty$ we have a counterexample to Theorem 5.3.2 and also Theorem 5.2.4 when the critical space is not a Banach space.

The reader may wonder why the same counterexample doesn't work on the Banach space $\mathcal{C}^0_{2\pi}$ of continuous periodic functions. Note that $L\colon R \times \mathcal{C}^0_{2\pi} \to \mathcal{C}^0_2$ is continuous but that $L\colon R \to L(\mathcal{C}^0_{2\pi}, \mathcal{C}^0_{2\pi})$ is not! Hence the linear maps $L(t)$ do not converge to $L(0) = I$ as $t \to 0$. Moreover consider the derivative of $L(t)f$ with respect to t, which we write as $L'(t)f$. Clearly

$$[L'(t)f](x) = \int_0^1 vf'(x + tv)\, dv$$

involves f'. Thus $L\colon R \times \mathcal{C}^\infty_{2\pi} \to \mathcal{C}^\infty_{2\pi}$ is $\mathcal{C}^\infty$ but $L\colon R \times \mathcal{C}^0_{2\pi} \to \mathcal{C}^0_{2\pi}$ is not even differentiable. The difference between $\mathcal{C}^0$ and $\mathcal{C}^\infty$ is striking.

5.3.4. COROLLARY. *The inverse function $P^{-1}(g)$ in Theorem 5.2.3 is smooth for all g in a neighborhood $\tilde{V}$ of g_0. Moreover the derivative $DP(f)h = k$ is invertible for all f in a neighborhood $\tilde{U}$ of f_0 and the family of inverses $VP(f)k = h$ is smooth, and we have*

$$DP^{-1}(g)k = VP\big(P^{-1}(g)\big)k.$$

PROOF. That the family of linear maps is invertible and that VP is smooth follows from Theorems 5.4.2 and 5.4.1.

By Lemma 3.3.1 we can find a smooth family of linear maps

$$L(f_1, f_2)h = \int_0^1 DP\big((1 - t)f_1 + tf_2\big)h\, dt$$

with $P(f_2) - P(f_1) = L(f_1, f_2)(f_2 - f_1)$, and with $L(f, f)h = DP(f)h$. Since $L(f_0, f_0) = DP(f_0)$ is invertible, it follows that $L(f_1, f_2)$ is invertible for all f_1, f_2 in a neighborhood $\tilde{U}$ of f_0, and the family of inverses $V(f_1, f_2)k = h$ solving $L(f_1, f_2)h = k$ is also smooth. Then

$$f_2 - f_1 = V(f_1, f_2)\big[P(f_2) - P(f_1)\big].$$

Now let $f_1 = P^{-1}(g_1)$ and $f_2 = P^{-1}(g_2)$. Then

$$P^{-1}(g_2) - P^{-1}(g_1) = V\big(P^{-1}(g_1), P^{-1}(g_2)\big)(g_2 - g_1).$$

Let us define

$$M(g_1, g_2)k = V\big(P^{-1}(g_1), P^{-1}(g_2)\big)k.$$

Since V is smooth and P^{-1} is continuous, we see that M is continuous and linear in k. Since

$$P^{-1}(g_2) - P^{-1}(g_1) = M(g_1, g_2)(g_2 - g_1)$$

it follows from the reverse side of Lemma 3.3.1 that P^{-1} is C^1 and we have $DP^{-1}(g)k = M(g, g)k$. Now

$$\begin{aligned} M(g, g) &= V\big(P^{-1}(g), P^{-1}(g)\big) = L\big(P^{-1}(g), P^{-1}(g)\big)^{-1} \\ &= DP\big(P^{-1}(g)\big)^{-1} = VP\big(P^{-1}(g)\big) \end{aligned}$$

so we have

$$DP^{-1}(g)k = VP\big(P^{-1}(g)\big)k.$$

If P^{-1} is C^r and VP is C^∞ then P^{-1} is C^{r+1} by the chain rule. Thus P^{-1} is C^∞.

5.3.5. EXAMPLE. Let $\mathcal{C}^\infty_{2\pi}$ be the Fréchet space of all smooth functions $f(t)$ periodic with period 2π, and let U be the open subset of all functions $f(t)$ with

$$c(f) = \int_0^{2\pi} f(t)\,dt \neq 0.$$

Define a family of linear maps

$$L\colon \left(U \subseteq \mathcal{C}^\infty_{2\pi}\right) \times \mathcal{C}^\infty_{2\pi} \to \mathcal{C}^\infty_{2\pi}$$

by letting $L(f)h = k$ where

$$k = dh/dt + fh.$$

Then the equation has a unique solution $V(f)k = h$, and V is smooth as a map

$$V\colon \left(U \subseteq \mathcal{C}^\infty_{2\pi}\right) \times \mathcal{C}^\infty_{2\pi} \to \mathcal{C}^\infty_{2\pi}.$$

To see this is true, we recall the standard method for solving a first order, linear ordinary differential equation. We begin by letting

$$F(t) = \int_0^t f(\theta)\,d\theta$$

so that $F' = f$. Note that $F(0) = 0$ and $F(2\pi) = c(f) \neq 0$ by hypothesis. Then the solution is given by

$$k(t) = e^{-F(t)}\left\{\int_0^t e^{F(\theta)}h(\theta)\,d\theta + C(f,h)\right\}$$

where C is an arbitrary constant. We wish to choose C so that $k(t)$ is periodic. This happens when $k(0) = k(2\pi)$. Thus we need

$$C(f,h) = \frac{1}{e^{c(f)} - 1}\int_0^{2\pi} e^{F(\theta)}h(\theta)\,d\theta$$

which is possible if $c(f) \neq 0$. Note that $C(f,h)$ depends smoothly on f and h and is linear in h. It is easy to check that if $f \in C^r$ and $k \in C^r$ then the solution $h \in C^{r+1}$ for $0 \leqslant r \leqslant \infty$. It is clear from the formulas that V is continuous, and then it is smooth by Theorem 5.3.1.

5.4. *Examples in Banach spaces.* We present a few of the classical examples of the inverse function theorem in finite dimensions and in Banach spaces to give the reader a feeling of how the theorem works.

5.4.1. EXAMPLE. Let $f\colon U \subseteq R^1 \to R^1$ be a real function of a real variable. Then

$$Df(x)u = f'(x)u$$

and the linear operator $Df(x)\colon R^1 \to R^1$ is multiplication by $f'(x)$. This operator is invertible if and only if $f'(x) \neq 0$. If $f'(x) > 0$ the function is increasing, while if $f'(x) < 0$ the function is decreasing. In either case it is locally invertible.

5.4.2. EXAMPLE. The function $y = x^2$ has two inverses $x = \pm\sqrt{y}$ for $y > 0$ and none for $y < 0$. The derivative $dy/dx = 2x$ is nonzero except at $x = 0$,

where the function has a turning point and the map is not invertible. Since

$$\frac{dy}{dx} = \frac{1}{\pm 2\sqrt{y}} = \frac{1}{2x}$$

we see that the derivative of the inverse is the inverse of the derivative.

5.4.3. EXAMPLE. Let $y = x^3$. Then the function has a unique solution defined everywhere given by $x = \sqrt[3]{y}$. The derivative $dy/dx = 3x^2$ is not zero except at $x = 0$. Since $dx/dy = 1/3y^{2/3}$, we see that at $y = 0$ the inverse function is continuous but not differentiable.

5.4.4. EXAMPLE. Let $y = 4x^3 - 3x$ be the third Chebychev polynomial. We have $y' = 12x^2 - 3 = 0$ at $x = \pm \frac{1}{2}$. A table of values is

x	$-\infty$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	∞
y	$-\infty$	-1	1	0	-1	1	∞

so y covers the interval $[-1, 1]$ three times while x covers it once. In this region we can make the substitution $x = \cos\theta$ and use the identity

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta$$

to obtain the three solutions

$$x = \cos \tfrac{1}{3} \arccos y.$$

In the region $x \geqslant 1$ we can make the substitution $x = \cosh t$ and use the identity

$$\cosh 3t = 4\cosh^3 t - 3\cosh t$$

to obtain the single solution

$$x = \cosh \tfrac{1}{3} \operatorname{arccosh} y.$$

Interestingly enough this latter solution is algebraic; it can also be written as

$$x = \left(\left(y + \sqrt{y^2 - 1}\right)^{1/3} + \left(y - \sqrt{y^2 - 1}\right)\right)^{1/3}/2$$

using the identity

$$\operatorname{arccosh} y = \log\left(y \pm \sqrt{y^2 - 1}\right).$$

Alternatively we can find the algebraic solution by substituting

$$x = \frac{u^2 + 1}{2u} = \frac{1}{2}\left(u + \frac{1}{u}\right)$$

which has two reciprocal solutions

$$u = x \pm \sqrt{x^2 - 1}\,.$$

We then get

$$y = \frac{u^6 + 1}{2u^3} = \frac{1}{2}\left(u^3 + \frac{1}{u^3}\right)$$

which has two reciprocal solutions

$$u = \left(y \pm \sqrt{y^2 - 1} \right)^{1/3}.$$

Knowing x as a function of u and u as a function of y gives x as a function of y.

The inverse function theorem tells us that there exists a single smooth inverse function in a neighborhood of $y = 1$ which has its values near $x = 1$. This solution is given by the algebraic formula for $y \geqslant 1$ and by the trigonometric formula for $y \leqslant 1$. That these two formulas patch together to form a single smooth (and even real analytic) function is far from obvious.

5.4.5. EXAMPLE. Let f: $U \subseteq R^2 \to R^2$ be given by $\binom{z}{w} = f\binom{x}{y}$. Then its derivative

$$Df\binom{x}{y}\binom{u}{v} = \begin{pmatrix} \partial z/\partial x & \partial z/\partial y \\ \partial w/\partial x & \partial w/\partial y \end{pmatrix}\binom{u}{v}$$

is given as a linear map by the matrix of partial derivatives. This is invertible if and only if the Jacobian determinant

$$J = \det\begin{pmatrix} \partial z/\partial x & \partial z/\partial y \\ \partial w/\partial x & \partial w/\partial y \end{pmatrix}$$

is not zero. Geometrically J is the factor by which f expands areas infinitesimally. When $J \neq 0$ the map f is locally invertible. If $J > 0$ the map preserves orientation, and if $J < 0$ it reverses orientation.

5.4.6. EXAMPLE. As a special case, consider the map f: $R^2 \to R^2$ given by

$$\begin{Bmatrix} z = x + y \\ w = xy \end{Bmatrix}.$$

This has an interesting interpretation. If a 2×2 matrix has eigenvalues x and y, then it has trace z and determinant w. We would like to recover the two eigenvalues from the trace and the determinant. The Jacobian is

$$J = \det\begin{pmatrix} 1 & 1 \\ y & x \end{pmatrix} = x - y$$

so $J = 0$ along the axis of symmetry $x = y$. When $x > y$ the map preserves orientation, while when $x < y$ it reverses orientation. Therefore it folds along the line $x = y$. Its image is the parabola $z^2 = 4w$. Since

$$z^2 - 4w = (x - y)^2 \geqslant 0$$

we see that the region $4w \leqslant z^2$ outside the parabola is covered twice, with the two solutions being obtained by interchanging x and y, while the region $4w > z^2$ inside the parabola is not covered at all. The branch of the inverse function with $x > y$ is given by

$$x = \left(z + \sqrt{z^2 - 4w} \right)/2, \qquad y = \left(z - \sqrt{z^2 - 4w} \right)/2.$$

When the eigenvalues are distinct they are smooth functions of the trace and determinant, but where the eigenvalues are equal they are not.

5.4.7. EXAMPLE. Consider the function f: $\mathbf{C} \to \mathbf{C}$ mapping the complex plane to itself given by $w = z^2$. The derivative $w' = 2z$ is zero only at the origin $z = 0$. Elsewhere the map is locally invertible, with two inverses given by the two branches of $z = \sqrt{w}$. In cartesian coordinates $z = x + iy$ and $w = u + iv$ the map f: $R^2 \to R^2$ is given by

$$\begin{Bmatrix} u = x^2 - y^2 \\ v = 2xy \end{Bmatrix}.$$

We can solve these simultaneous equations by observing that

$$u^2 + v^2 = (x^2 + y^2)^2$$

from which it follows that

$$x = \pm \left(\left(\sqrt{u^2 + v^2} + u \right)/2 \right)^{1/2},$$

$$y = \pm \left(\left(\sqrt{u^2 + v^2} - u \right)/2 \right)^{1/2}$$

where the signs of x and y are the same when $v > 0$ and opposite when $v < 0$. When $v \neq 0$ the solution is obviously smooth. By the inverse function theorem it is still smooth at $v = 0$ if $u \neq 0$. For the Jacobian determinant

$$J = \det \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} = 4(x^2 + y^2)$$

is never zero except at the origin. In polar coordinates $z = re^{i\theta}$ and $w = se^{i\psi}$ we see that

$$\begin{Bmatrix} s = r^2 \\ \psi = 2\theta \end{Bmatrix}$$

so the map squares the radius and doubles the angle. Then the inverse map

$$\begin{Bmatrix} r = \sqrt{s} \\ \theta = \psi/2 \end{Bmatrix}$$

takes the square root of the radius and halves the angle (which can be done in two ways mod 2π).

5.4.8. EXAMPLE. Consider the map f: $R^2 \to R^2$ given by $w = 2z + \bar{z}^2$ or

$$\begin{Bmatrix} u = 2x + x^2 - y^2 \\ v = 2y - 2xy \end{Bmatrix}.$$

The Jacobian determinant is

$$J = \det \begin{pmatrix} 2(1 + x) & -2y \\ -2y & 2(1 - x) \end{pmatrix} = 4(1 - x^2 - y^2)$$

so J vanishes on the unit circle $x^2 + y^2 = 1$. Inside the map preserves orientation, and outside it reverses. The map is locally invertible except on the

unit circle. If we parametrize the circle by $x = \cos\theta$ and $y = \sin\theta$ we see that its image is

$$\left\{\begin{array}{l} u = 2\cos\theta + \cos 2\theta \\ v = 2\sin\theta - \sin 2\theta \end{array}\right\}$$

which is the standard parametrization for the hypocycloid of three cusps generated by a point on a circle of radius 1 rolling around the inside of a circle of radius 3. The map folds along the circle except at the three cube roots of unity, where it makes cusps. For points $\binom{u}{v}$ inside the hypocycloid there are four smooth solutions $\binom{x}{y}$, while for points outside the hypocycloid there are only two solutions. For example, if ε is a primitive cube root of unity then the points which map to the origin $w = 0$ are $z = 0, -2, -2\varepsilon, -2\bar{\varepsilon}$. Corresponding to these there will be four smooth local inverses defined in a neighborhood of the origin.

There is an important result in the theory of singularities that any map of the plane to the plane is arbitrarily close to one having only folds and cusps. The map $w = z^2$ of Example 5.4.7 has no folds or cusps, just a double twist. But the map $w = z^2 + \varepsilon\bar{z}$ looks like the map of Example 5.4.8, folding along a circle with its image being a hypocycloid of three cusps. As $\varepsilon \to 0$ the three cusps collapse into the double twist.

5.4.9. EXAMPLE. Let B be a Banach space and let $L(B, B)$ be the Banach space of continuous linear maps of B to itself. Define a map P: $L(B, B) \to L(B, B)$ by

$$P(L) = L^2.$$

Then P is smooth and we have

$$DP(L)M = LM + ML.$$

When $L = I$ is the identity

$$DP(I)M = 2M$$

so $DP(I)$ is invertible with inverse

$$VP(I)N = N/2.$$

It follows that P is locally invertible in a neighborhood of the identity, and the inverse map $P^{-1}(L) = \sqrt{L}$ is smooth. We could also define the function $\sqrt{L}$ for L near the identity using the power series expansion

$$\sqrt{I + A} = I + \tfrac{1}{2}A - \tfrac{1}{8}A^2 + \tfrac{1}{16}A^3 - \tfrac{5}{128}A^4 + \cdots$$

whose general term is

$$\frac{(-1)^n}{2^n}\,\frac{1 \cdot 3 \cdot 5 \cdots (2n-3)}{1 \cdot 2 \cdot 3 \cdots n}A^n$$

which converges absolutely for $\|A\| < 1$ by the ratio test. This shows $P^{-1}(L) = \sqrt{L}$ is actually defined and smooth on all of the set $\{L: \|L - I\| < 1\}$.

5.4.10. EXAMPLE. The inverse function theorem can be used to prove existence of solutions for ordinary differential equations.

THEOREM. *Let $y = p(x)$ be a smooth monotone increasing function mapping the open interval $a < x < b$ onto the open interval $c < y < d$ with $p' > 0$. For every smooth periodic function g of period 2π with values in (c, d) there exists a unique smooth periodic function f of period 2π with values in (a, b) solving*

$$df/dt + p(f) = g.$$

PROOF. Let $C^r_{2\pi}$ be the Banach space of functions of class C^r periodic with period 2π. Let $\tilde{U}$ be the open subset of $C^1_{2\pi}$ of functions with range in $(a, b) = U$ and let $\tilde{V}$ be the open subset of $C^0_{2\pi}$ of functions with range in $(c, d) = V$. Define a map

$$P\colon \tilde{U} \subseteq C^1_{2\pi} \to \tilde{V} \subseteq C^0_{2\pi}$$

by letting

$$P(f) = df/dt + p(f).$$

Then P is smooth and its derivative is

$$DP(f)h = dh/dt + p'(f)h.$$

Since $p'(f) > 0$ we know from Example 5.3.5 that the equation $DP(f)h = k$ always has a unique solution $VP(f)k = h$, and if $f \in C^1_{2\pi}$ and $k \in C^0_{2\pi}$ then $h \in C^1_{2\pi}$. It follows from the inverse function theorem that P is locally invertible, and the image of P is open.

Next we observe that at the maximum or minimum of f we have $df/dt = 0$ and $p(f) = g$. Hence p maps the range of f into the range of g. Suppose moreover that we have two solutions of the equation, so that

$$p(f_1) = df_1/dt + p(f_1) = g, \qquad p(f_2) = df_2/dt + p(f_2) = g.$$

Then take the difference

$$d(f_1 - f_2)/dt + [p(f_1) - p(f_2)] = 0.$$

It follows that wherever $f_1 - f_2$ is a maximum its derivative is zero and $p(f_1) = p(f_2)$. Since p is one-to-one, $f_1 = f_2$ and the maximum of $f_1 - f_2$ is 0. Its minimum is also zero, so $f_1 = f_2$. Hence our solutions are unique and P: $\tilde{U} \to \tilde{V}$ is one-to-one.

It remains to show that P is onto. Let $P(\tilde{U})$ be the image of $\tilde{U}$ under P in $\tilde{V}$. We know $P(\tilde{U})$ is open and not empty. If we can show that $P(\tilde{U})$ is relatively closed in $\tilde{V}$, then since $\tilde{V}$ is connected we must have $P(\tilde{U}) = \tilde{V}$. To see this, let $f_j \in \tilde{U}$ be a sequence such that $P(f_j) = g_j$ converges to some $g \in \tilde{V}$. Since g is continuous and periodic its image is compact. Therefore we can find a compact subinterval $[\gamma, \delta] \subseteq (c, d)$ such that g and all the g_j have their ranges in $[\gamma, \delta]$. Let $[\alpha, \beta] = p^{-1}([\gamma, \delta])$. Then by our previous remark all the f_j have their ranges in the compact interval $[\alpha, \beta]$. Since

$$df_j/dt = g_j - p(f_j)$$

we have a uniform bound on the derivatives df_j/dt. Then by Ascolis' theorem a subsequence of the f_j will converge uniformly to a continuous function f with range in $[\alpha, \beta] \subseteq (a, b)$. By the above equation we also have $df_j/dt \to df/dt$ uniformly. Hence $f_j \to f$ in $C^1_{2\pi}$. Since $P(f_j) \to P(f)$ and $P(f_j) = g_j \to g$ we

have $P(f) = g$. This shows that the limit g belongs to $P(\tilde{U})$ also. It follows that $P\colon \tilde{U} \to \tilde{V}$ is globally invertible and the inverse is smooth.

5.4.11. EXAMPLE. We can use the inverse function theorem to solve partial differential equations on manifolds. Let X be a compact Riemannian manifold with metric

$$ds^2 = g_{ij}\,dx^i\,dx^j.$$

Let $d\mu = \mu(x)\,dx$ be the associated measure where

$$\mu = \sqrt{\det g_{ij}}$$

and let g^{ij} be the inverse matrix. There is an intrinsically defined second order linear operator on functions called the Laplacian, defined by

$$\Delta f = \operatorname{div}\operatorname{grad} f = \frac{1}{\mu}\frac{\partial}{\partial x^i}\left(\mu g^{ij}\frac{\partial f}{\partial x^j}\right).$$

If $X = S^1 \times S^1$ is the flat torus, we can regard a function on X as a function $f(x, y)$ doubly periodic in x and y, and

$$\Delta f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2$$

in the flat metric $ds^2 = dx^2 + dy^2$. By contrast if $X = S^2$ is the sphere parametrized by longitude θ and latitude ψ, then

$$ds^2 = \cos^2\psi\,d\theta^2 + d\psi^2, \qquad \mu = \cos\psi,$$

$$\Delta f = \sec^2\psi\,\frac{\partial^2 f}{\partial\theta^2} + \frac{\partial^2 f}{\partial\psi^2} - \tan\psi\,\frac{\partial f}{\partial\psi}.$$

The operator Δ on S^2 is invariant under all rotations. Any linear second order operator on S^2 invariant under rotations is a linear combination of I and Δ.

As before, let p be a smooth increasing function taking the open interval (a, b) onto the open interval (c, d) with $p' > 0$. We prove the following result.

THEOREM. *For any smooth function $g \in \mathcal{C}^\infty(X)$ with $c < g < d$ there exists a unique smooth function $f \in \mathcal{C}^\infty(X)$ with $a < f < b$ solving the equation*

$$p(f) - \Delta f = g.$$

PROOF. After we prove the Nash-Moser theorem we could work in $\mathcal{C}^\infty(X)$. Until then, we introduce the Banach spaces $\mathcal{C}^{n+\alpha}(X)$ of functions which (in local coordinates) have nth derivatives which satisfy a Hölder condition of exponent α for some α in $0 < \alpha < 1$,

$$|f(x) - f(y)| \leqslant C\,|x - y|^\alpha.$$

Let $\tilde{U} \subseteq \mathcal{C}^{2+\alpha}$ and $\tilde{V} \subseteq \mathcal{C}^\alpha$ be the open subsets of functions f and g with values in U and V respectively.

Define a map

$$P\colon \tilde{U} \subseteq \mathcal{C}^{2+\alpha}(X) \to \tilde{V} \subseteq \mathcal{C}^\alpha(X)$$

by letting $P(f) = p(f) - \Delta f$. Then P is smooth and

$$DP(f)h = p'(f)h - \Delta h = k.$$

We claim that this linear elliptic equation always has a unique solution $h \in C^{2+\alpha}$ for $f \in \tilde{U} \subseteq C^{2+\alpha}$ and $k \in C^{\alpha}$. Since $p'(f) \in C^{2+\alpha}$ we have all the regularity we need for the coefficients. Since Δ is selfadjoint, the operator $DP(f)$ has index zero, and it is one-to-one if and only if it is onto by the Fredholm alternative. To see that it is one-to-one, we observe that if $\Delta h = p'(f)h$ then integrating by parts

$$\int_X \{|\Delta h|^2 + p'(f)h^2\}\, d\mu = 0$$

and since $p' > 0$ we must have $h = 0$. Then the inverse function theorem shows that P is locally invertible everywhere, and its image $P(\tilde{U})$ is open.

As before we wish to show that $P(\tilde{U})$ is relatively closed in V. First we observe that

$$p(\text{range } f) \subseteq \text{range } g.$$

For where f has its maximum $\Delta f \leqslant 0$ and where f has its minimum $\Delta f \geqslant 0$. Now let $g_j \to g$ in $\tilde{V}$. Since g is continuous and X is compact, the range of g and all the g_j will lie in a compact subset $[\gamma, \delta]$ of (c, d). Then if $P(f_j) = g_j$, all the f_j will lie in a compact subset $[\alpha, \beta]$ of (a, b). On this compact subset p is bounded, so we have an estimate

$$\| p(f_j) \|_{C^0} \leqslant C.$$

Let $L_q^n(X)$ denote the Banach space of functions on X which (in local coordinates) have n (distributional) derivatives in L_q with $1 < q < \infty$. If we pick $q > \dim X$ then by Sobolev's embedding $L_q^n \subseteq C^{n-1}$. Moreover by Gårding's inequality

$$\| f \|_{L_q^2} \leqslant C\big(\| \Delta f \|_{L_q} + \| f \|_{L_q}\big).$$

Since $\Delta f_j = p(f_j) - g_j$ and $p(f_j)$ and g_j are bounded in $C^0 \subseteq L_q$, the f_j are bounded in $L_q^2 \subseteq C^1$. Since the inclusion $C^1 \subseteq C^{\alpha}$ is compact, by passing to a subsequence we may assume the f_j converge in C^{α}. Since $\Delta f_j = p(f_j) - g_j$, and $p(f_j)$ and g_j converge in C^{α}, we have Δf_j converging in C^{α}. Then by Gårding's inequality the f_j converge in $C^{2+\alpha}$ to a function $f \in \tilde{U}$. Clearly $P(f) = g$. This shows $P(\tilde{U})$ is relatively closed in $\tilde{V}$. Since $\tilde{V}$ is connected, $P(\tilde{U}) = \tilde{V}$.

It is also clear that P is one-to-one. For if f_1 and f_2 are two solutions of

$$p(f_1) - \Delta f_1 = g, \qquad p(f_2) - \Delta f_2 = g$$

then taking the difference

$$\Delta(f_1 - f_2) = p(f_1) - p(f_2).$$

Integrating by parts

$$\int \{|\Delta(f_1 - f_2)^2 + [f_1 - f_2][p(f_1) - p(f_2)]\}\, d\mu = 0.$$

Since p is increasing, the expression

$$[f_1 - f_2][p(f_1) - p(f_2)] \geqslant 0$$

and vanishes only when $f_1 = f_2$. Thus P is one-to-one.

Hence for every $g \in C^\alpha$ with $c < g < d$ there exists a unique $f \in C^{2+\alpha}$ with $p(f) - \Delta f = g$. Since $\Delta f = p(f) - g$, if $g \in C^{n+\alpha}$ then $f \in C^{n+2+\alpha}$ by induction. If $g \in C^\infty$ then $f \in C^\infty$.

5.4.12. EXAMPLE. We can study geodesics on a Riemannian manifold using the inverse function theorem. For simplicity we shall take the manifold X to be an open subset of R^m; the general case can be handled with Banach manifolds of maps. In local coordinates $x = \{x^i\}$ the metric has the form $ds^2 = g_{ij}\,dx^i\,dx^j$, and Γ^i_{jk} are the Christoffel symbols. A geodesic is a curve $x^i = f^i(t)$ satisfying the equation

$$P(f) = \frac{d^2 f^i}{dt^2} + \Gamma^i_{jk}(f)\frac{df^j}{dt}\frac{df^k}{dt} = 0.$$

Every curve of shortest length parametrized proportionally to the arc length is a geodesic. Also the constant curve $f \equiv a$ is a geodesic. We would like to find a geodesic with given endpoints a and b; this means we impose boundary conditions $f^i(0) = a^i$ and $f^i(1) = b^i$.

Let $C^n[0,1]$ denote the space of functions of class C^n on the interval $0 \leqslant t \leqslant 1$ and let $C^n[0,1]^m$ denote the m-fold product of the space with itself. Our curve $f(t)$ will belong to the space $C^2[0,1]^m$, with $f^i(t) \in C^2[0,1]$ for $1 \leqslant i \leqslant m$. The operator P defines a smooth map

$$P\colon C^2[0,1]^m \to C^0[0,1]^m.$$

To impose boundary conditions, let B be the map

$$B\colon C^2[0,1]^m \to C^0[0,1]^m \times R^m \times R^m, \qquad B(f) = (P(f), f(0), f(1)).$$

Then its derivative is given by

$$DB(f)h = (DP(f)h, h(0), h(1))$$

where the derivative of P may be computed as

$$\{DP(f)h\}^i = \frac{d^2 h^i}{dt^2} + 2\Gamma^i_{jk}(f)\frac{df^j}{df}\frac{dh^k}{dt} + \frac{\partial \Gamma^i_{jk}(f)}{\partial x^l}\frac{df^j}{dt}\frac{df^k}{dt}h^l.$$

This equation looks better if we introduce the covariant derivative along the curve

$$\{\partial_t h\}^i = \frac{dh^i}{dt} + \Gamma^i_{jk}(f)\frac{df^j}{dt}h^k.$$

Then the equations become

$$DP(f)h = \partial_t^2 h + R(\partial_t f, h)\partial_t f$$

where R is the curvature tensor.

To invert the operator $DB(f)$ we must solve the second order linear system of ordinary differential equations for $DP(f)h$ with given boundary values $h(0)$ and $h(1)$. By the Fredholm alternative this is possible if and only if there are no nontrivial solutions to the homogeneous equations with zero boundary values. The solutions of $DP(f)h = 0$ are known as Jacobi fields along f. There will be a solution with $h(0) = 0$ and $dh(0)/dt$ equal to any vector in R^n. If we

normalize $|dh(0)/dt| = 1$, then by a compactness argument there will be at least a finite time $T > 0$ such that the Jacobi field h cannot vanish again before $t = T$. The first point where some Jacobi field vanishes again is called the first conjugate point.

If f is a geodesic and its endpoints $a = f(0)$ and $b = f(1)$ are sufficiently close, the derivative $DB(f)$ has zero null space and hence is invertible. By the inverse function theorem B is locally invertible near f. Since f is a geodesic, $P(f) = 0$ and $B(f) = (0, a, b)$. If $\tilde{a}$ is close to a and $\tilde{b}$ is close to b, then $\tilde{f} = B^{-1}(0, \tilde{a}, \tilde{b})$ will be a geodesic, since $P(\tilde{f}) = 0$, and have endpoints $\tilde{a}$ and $\tilde{b}$. If a and b are conjugate points along a geodesic f then B fails to be locally invertible near f, as can be seen for diametrically opposite points on a sphere. When the curvature is negative there cannot be any nonzero Jacobi fields vanishing at the endpoints, as can be seen by integrating by parts to get

$$\int_0^1 \{|\partial_t h|^2 - \langle R(\partial_t f, h)\partial_t f, h\rangle\}\, dt = 0.$$

If $\langle R(u, v)u, v\rangle \leqslant 0$ then $h = 0$.

5.4.13. EXAMPLE. The classical Plateau problem calls for finding the surface of least area with a given boundary. We shall look at a special case where the surface is the graph of a function. Let D be a closed domain in the $\binom{x}{y}$-plane with a smooth boundary ∂D, and let $f\binom{x}{y}$ be a function defined on D. The graph of $z = f\binom{x}{y}$ is a surface bounding the curve which is the graph of f over ∂D. If this surface has least area among all with the same boundary then f will satisfy the nonlinear partial differential equation

$$P(f) = (1 + f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1 + f_x^2)f_{yy} = 0$$

which says geometrically that the mean curvature is zero.

Let $C^{n+\alpha}(D)$ denote the Banach space of functions on D which have nth derivatives which are Hölder continuous of exponent α $(0 < \alpha < 1)$, and let $C^{n+\alpha}(\partial D)$ be the same for functions on the boundary. Define a map

$$B\colon C^{2+\alpha}(D) \to C^{\alpha}(D) \times C^{2+\alpha}(\partial D), \qquad B(f) = (P(f), f|\partial D).$$

For a given function b on the boundary we would like to solve the boundary value problem $P(f) = 0$ and $f|\partial D = b$, which says $B(f) = (0, b)$. The graph of f will be the surface with least area bounded by the curve which is the graph of b.

The derivative of B is given by

$$DB(f)h = (DP(f)h, h|\partial D)$$

and $DP(f)h$ can be computed explicitly:

$$DP(f)h = (1 + f_y^2)h_{xx} - 2f_x f_y hxy + (1 + f_x^2)h_{yy} + 2(f_x f_{yy} - f_y f_{xy})h_x + 2(f_y f_{xx} - f_x f_{xy})h_y.$$

Observe that $DP(f)h$ is a second order linear partial differential operator in h for each given f, which we shall show is always elliptic, and that $DB(f)h$ imposes Dirichlet boundary conditions on h as well.

To see that the equation is elliptic we replace $\partial/\partial x$ and $\partial/\partial y$ by ξ and η in the highest order derivatives and compute the symbol

$$\sigma(\xi, \eta) = \left(1 + f_y^2\right)\xi^2 - 2f_x f_y \xi\eta + \left(1 + f_x^2\right)\eta^2.$$

Then it is easy to see that

$$\sigma(\xi, \eta) = \xi^2 + \eta^2 + \left(f_y\xi - f_x\eta\right)^2$$

so $\sigma(\xi, \eta) = 0 \Leftrightarrow \xi = 0$ and $\eta = 0$. Since the equation may be deformed to the Laplacian its index is zero and the Fredholm alternative holds. The homotopy is accomplished by

$$\sigma_t(\xi, \eta) = \xi^2 + \eta^2 + t\left(f_y\xi - f_x\eta\right)^2, \qquad 0 \leqslant t \leqslant 1.$$

Note that there is no zero order term in $DP(f)h$. By the maximum principle we conclude that there are no nonzero solutions of the homogeneous equation with zero boundary data. Then by the Fredholm alternative we see that $DB(f)$ is invertible for any f. Hence B is always locally invertible everywhere. As before, we can also prove that B is one-to-one and has closed range by making some a priori estimates.

We will indicate how to obtain the required a priori estimates on solutions of the Plateau equation $P(f) = 0$. First we observe if f and g are two solutions of the equation over D, and if $f \leqslant g$ on ∂D, then $f \leqslant g$ on D. This follows from applying the maximum principle to the difference $h = f - g$, which satisfies a second order elliptic equation with no constant term. Now any affine function g satisfies the Plateau equation. Thus if $P(f) = 0$ then f assumes its maximum (and minimum) on the boundary.

We can also bound the first derivatives of f in D using only the second derivatives of $f \mid \partial D$. To start we observe that the function

$$w = |\nabla f|^2 = f_x^2 + f_y^2$$

satisfies a second order elliptic inequality

$$\begin{aligned}\left(1 + f_y^2\right)w_{xx} - 2f_x f_y w_{xy} + \left(1 + f_x^2\right)w_{yy} + 2\left(f_{xx} + f_{yy}\right)\left(f_x w_x + f_y w_y\right)\\ = \left(1 + f_x^2 + f_y^2\right)\left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2\right) + \left(w_x^2 + w_y^2\right) \geqslant 0\end{aligned}$$

and hence w assumes its maximum on the boundary. The tangential derivative of f is of course known from $f \mid \partial D$, so we must estimate the normal derivative at a point on the boundary. By a rigid motion we can assume our point is at the origin and the boundary is tangent to the x-axis. Suppose f has a Taylor series

$$f\begin{pmatrix} x \\ y \end{pmatrix} = a + bx + cy + dx^2 + \cdots.$$

We must estimate the normal derivative c. We can find a constant C, depending only on the derivatives of degree at most 2 of $f \mid \partial D$, such that on ∂D

$$f\begin{pmatrix} x \\ y \end{pmatrix} - a - bx \leqslant C\left(y + x^2\right).$$

If ∂D is strictly convex, then $y \geqslant \varepsilon x^2$ for some $\varepsilon > 0$. If we let g be the affine function

$$g\begin{pmatrix} x \\ y \end{pmatrix} = a + bx + C\left(1 + \frac{1}{\varepsilon}\right) y$$

then $f \leqslant g$ on ∂D and hence on all of D. Then $c \leqslant C(1 + \frac{1}{\varepsilon})$. The same trick estimates c from below also.

Finally we must estimate the Hölder norm of f_x and f_y. Since f satisfies a quasi-linear equation in divergence form

$$\left(\frac{f_x}{\sqrt{1 + f_x^2 + f_y^2}} \right)_x + \left(\frac{f_y}{\sqrt{1 + f_x^2 + f_y^2}} \right)_y = 0$$

it follows that the derivative $g = f_x$ (or f_y) satisfies a linear elliptic equation in divergence form with coefficients depending only on f_x and f_y, namely

$$\left(\frac{\left(1 + f_y^2\right) g_x - f_x f_y g_y}{\left(1 + f_x^2 + f_y^2\right)^{3/2}} \right)_x + \left(\frac{-f_x f_y g_x + \left(1 + f_x^2\right) g_y}{\left(1 + f_x^2 + f_y^2\right)^{3/2}} \right)_y = 0.$$

Note that the matrix of the equation

$$\frac{1}{\left(1 + f_x^2 + f_y^2\right)^{3/2}} \begin{pmatrix} 1 + f_y^2 & -f_x f_y \\ -f_x f_y & 1 + f_x^2 \end{pmatrix}$$

is strictly positive-definite. Since f_x and f_y are uniformly bounded, we can appeal to standard results on linear elliptic equations in divergence form (see Ladyzenskaja and Ural'ceva [**10**, Theorem 14.1]) to estimate the C^α norm of $g = f_x$ (also f_y). It is then easy to estimate the $C^{2+\alpha}$ norm of f on D directly from the equation $P(f) = 0$ in terms of the $C^{2+\alpha}$ norm of $f \mid \partial D$.

It follows that the operator P is globally invertible. Hence for any function b on ∂D we can solve for a function f with $P(f) = 0$ on D and $f = b$ on ∂D.

5.4.14. EXAMPLE. We consider the classical problem of a round object rolling back and forth in an oscillatory motion in a smooth convex trough of general shape without friction. We shall prove the following result.

THEOREM. *For small values of the energy, there exist smooth oscillatory solutions for the equations of motion which are uniquely determined as smooth functions of their two real Fourier coefficients of order one. Moreover the period of oscillation T is a smooth function of the total energy E for $E \geqslant 0$.*

PROOF. Suppose the object has mass m, radius r, and radis of gyration $\bar{r}$. Let $\varepsilon = \bar{r}/r$ be the gyration ratio. We have the following table for the simplest cases:

$\varepsilon^2 = 1$ for a hollow cylinder,

$\varepsilon^2 = 1/2$ for a solid cylinder,

$\varepsilon^2 = 2/3$ for a hollow sphere,

$\varepsilon^2 = 2/5$ for a solid sphere.

We parametrize the motion using the distance s travelled by the center of mass from its place of rest at $s = 0$. Let $h(s)$ be the height of the center of mass above its level of rest as a function of s and let g be the gravitational constant. The total energy E is a constant, and it is the sum of the kinetic translational and rotational energies and the potential energy. This gives the equation

$$\frac{1}{2}m(1 + \varepsilon^2)\left(\frac{ds}{dt}\right)^2 + mgh(s) = E.$$

If we differentiate with respect to t and divide by the linear momentum mds/dt we get the equation

$$(1 + \varepsilon^2)d^2s/dt^2 + gh'(s) = 0.$$

It is convenient to normalize the parameters of the problem. Let c be the curvature of the trough at its lowest point. Then the curve traversed by the center of mass has curvature $\tilde{c}$ given by

$$\tilde{c} = c/(1 - cr).$$

(Note that if $cr > 1$ the object won't fit in the trough!) Then $h(s)$ will have a power series expansion (which may not converge)

$$h(s) = \tfrac{1}{2}\tilde{c}s^2 + \cdots.$$

Let $k(s) = h'(s)$. Then $k(s)$ has a power series expansion

$$k(s) = \tilde{c}s + \cdots.$$

Let T be the period of the oscillation and $\nu = 2\pi/T$ its angular frequency. We introduce dimensionless variables by the substitutions

$$\sigma = \tilde{c}s, \qquad \theta = \nu t.$$

Then the motion can be described by a smooth function $\sigma = f(\theta)$ periodic with period 2π satisfying the equation

$$(*) \qquad d^2f/d\theta^2 + (1 + z)k(f) = 0$$

where k is a smooth function with a power series

$$k(\sigma) = \sigma + k_2\sigma^2 + k_3\sigma^3 + \cdots$$

and

$$z = \frac{g\tilde{c}}{(1 + \varepsilon^2)\nu^2} - 1$$

is a dimensionless parameter that depends on the constants c, r, and ε, and is a smooth function of the angular frequency $\nu = 2\pi/T$.

We always have the rest solution $f = 0$ for any value of the frequency. At $z = 0$ we encounter resonance, and another branch of solutions bifurcates off, giving rise to the oscillatory motions. Note that at $z = 0$ the period of oscillation is given by

$$T = 2\pi\sqrt{\frac{(1 + \varepsilon^2)(1 - cr)}{gc}}$$

in terms of the radius r of the object, the curvature c of the trough, and the gyration ratio ε. For the simple case $k(\sigma) = \sigma$ we have simple harmonic motion, and the solutions of our equation $(*)$ are given by

$$f(\theta) = x\cos\theta + y\sin\theta$$

for any values of x and y, with $z = 0$. For a more general function $k(\sigma)$ it is not so easy to solve, and the solution $f(\theta)$ will contain other harmonics. Both $f(\theta)$ and the frequency parameter z will depend smoothly on the Fourier coefficients x and y.

To see this, we wish to apply the inverse function theorem. The linearization of the nonlinear operator $(*)$ at $f = 0$ and $z = 0$ is the operator $d^2f/dx^2 + f$ which of course is not invertible. Indeed this is why the solutions bifurcate. The null space of the operator contains the functions $\cos\theta$ and $\sin\theta$, while the image lacks them. Therefore we invent an operator we can invert.

Let $U \subseteq R$ be an open set on which $k(\sigma)$ is defined. Let $C^r_{2\pi}$ be the Banach space of functions $f(\theta)$ of class C^r periodic in θ with period 2π. Let $\tilde{U} \subseteq C^2_{2\pi}$ be the open subset of functions with range in U. Define a map

$$P\colon \left(U \subseteq C^2_{2\pi}\right) \times R^3 \to C^0_{2\pi} \times R^3$$

by letting $P(f, u, v, w) = (g, x, y, z)$ where

$$g = d^2f/d\theta^2 + (1 + w)k(f) - u\cos\theta - v\sin\theta,$$

$$x = \frac{1}{\pi}\int_0^{2\pi} f(\theta)\cos\theta\, d\theta, \qquad y = \frac{1}{\pi}\int_0^{2\pi} f(\theta)\sin\theta\, d\theta,$$

$$z = w.$$

Let us use a twiddle to denote a variation in a number or function. Then the derivative of P,

$$DP(f, u, v, w)(\tilde{f}, \tilde{u}, \tilde{v}, \tilde{w}) = (\tilde{g}, \tilde{x}, \tilde{y}, \tilde{z}),$$

is given by

$$\tilde{g} = d^2\tilde{f}/d\theta^2 + (1 + w)k'(f)\tilde{f} + \tilde{w}k(f) - \tilde{u}\cos\theta - \tilde{v}\sin\theta,$$

$$\tilde{x} = \frac{1}{\pi}\int_0^{2\pi} \tilde{f}(\theta)\cos\theta\, d\theta, \qquad \tilde{y} = \frac{1}{\pi}\int_0^{2\pi} \tilde{f}(\theta)\sin\theta\, d\theta,$$

$$\tilde{z} = \tilde{w}.$$

If we take the standard solution $f = 0$, $u = 0$, $v = 0$, and $w = 0$ the first equation simplifies to

$$\tilde{g} = \frac{d^2\tilde{f}}{d\theta^2} + \tilde{f} - \tilde{u}\cos\theta - \tilde{v}\sin\theta.$$

Now we claim that for each $\tilde{g}$, $\tilde{x}$, $\tilde{y}$, $\tilde{z}$ there is a unique $\tilde{f}$, $\tilde{u}$, $\tilde{v}$, $\tilde{w}$ solving these equations. This is immediately clear from considering the Fourier series of $f(\theta)$. The operator

$$Lf = d^2f/d\theta^2 + f$$

kills the terms $\sin\theta$ and $\cos\theta$ but is invertible on the others. We can choose $\tilde{u}$ and $\tilde{v}$ to match the coefficients of $\tilde{g}$. Then we can solve for $\tilde{f}$, and the solution

can be changed up to the same sort of terms, which allows us to satisfy the equations for $\tilde{x}$ and $\tilde{y}$. Of course we take $\tilde{w} = \tilde{z}$. It is a little harder to see that $\tilde{f} \in C^2$. Since $L\tilde{f} \in C^0$ we have its nth Fourier series coefficients c_n bounded. Then the coefficients for $\tilde{f}$ are $a_n = c_n/(1 - n^2)$, so the series for $\tilde{f}$ converges absolutely and $\tilde{f} \in C^0$. Then using the equation $d^2\tilde{f}/d\theta^2 \in C^0$ and $\tilde{f} \in C^2$.

Since $DP(0, 0, 0, 0)$ is invertible, it follows from the inverse function theorem that for $g = 0$ and all x, y and z sufficiently small we can find a function $f(\theta)$ and numbers u and v depending smoothly on x, y, and z solving the equation

$$(**) \qquad d^2f/d\theta^2 + (1 + z)k(f) = u\cos\theta + v\sin\theta.$$

This will be a solution of (*) if and only if $u = v = 0$. To see what happens, consider the special case $k(\sigma) = \sigma$ of harmonic motion. Then we can compute

$$u = xz, \qquad v = yz.$$

The set $u = v = 0$ is the union of the line $x = y = 0$ (corresponding to the rest solutions) and the plane $z = 0$ (corresponding to the oscillatory solutions). The reason this doesn't look like one of Thom's standard catastrophes is that the motion is special due to the conservation of energy, and the resulting set of solutions are all periodic and hence invariant under the rotation group S^1. We can exploit this symmetry.

Returning to the general case, let $f(\theta)$, u, v be a solution of (**) for some values of x, y, z. If we make the phase change $\theta \to \theta + \alpha$ then $f(\theta + \alpha)$ will also be a solution of (**) where the vectors $\binom{u}{v}$ and $\binom{x}{y}$ have been rotated through an angle α. It follows that their inner product

$$l = \begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} u \\ v \end{pmatrix} = xu + yv$$

is a smooth function of x, y, z which is invariant under rotation in the xy-plane. If $r^2 = x^2 + y^2$ then l is a smooth function of r and z. Moreover it is symmetric in r, so l is a smooth function of r^2 and z. Moreover l vanishes at $r^2 = 0$, so $l = r^2m$ for a smooth function m of r^2 and z. Thus

$$xu + yv = (x^2 + y^2)m(x^2 + y^2, z).$$

If we perform integration by parts on (**) using the expression that shows the energy is conserved in (*), we get

$$uy - vx = \int_0^{2\pi} f[u\sin\theta - v\cos\theta]\,d\theta = \int_0^{2\pi} \frac{df}{d\theta}[u\cos\theta + v\sin\theta]\,d\theta$$

$$= \int_0^{2\pi} \frac{df}{d\theta}\left[\frac{d^2f}{d\theta^2} + (1 + z)k(f)\right] d\theta = 0$$

so $uy = vx$. Then we can solve for u and v in terms of m to get

$$u = xm(x^2 + y^2, z), \qquad v = ym(x^2 + y^2, z)$$

for some smooth function m of $x^2 + y^2$ and z. Clearly $u = v = 0$ when $m = 0$. To complete the proof of the theorem we only need the following fact.

LEMMA. *At $x = y = z = 0$ we have $m = 0$ and $\partial m/\partial z = 1$.*

PROOF. Since $u = xm$ it suffices to show that $\partial u/\partial x = 0$ and $\partial^2 u/\partial x \partial z = 1$ at $x = y = 0$. Differentiating (**) with respect to x in the variables $f(\theta)$, u and v gives

$$\frac{d^2}{d\theta^2}\frac{\partial f}{\partial x} + (1+z)k'(f)\frac{\partial f}{\partial x} = \frac{\partial u}{\partial x}\cos\theta + \frac{\partial v}{\partial x}\sin\theta$$

or differentiating with respect to z gives

$$\frac{d^2}{d\theta^2}\frac{\partial f}{\partial z} + (1+z)k'(f)\frac{\partial f}{\partial z} + k(f) = \frac{\partial u}{\partial z}\cos\theta + \frac{\partial v}{\partial z}\sin\theta$$

while differentiating with respect to both gives

$$\frac{d^2}{d\theta^2}\frac{\partial^2 f}{\partial x \partial z} + (1+z)k'(f)\frac{\partial^2 f}{\partial x \partial z} + (1+z)k''(f)\frac{\partial f}{\partial x}\frac{\partial f}{\partial z} + k'(f)\frac{\partial f}{\partial x}$$
$$= \frac{\partial^2 u}{\partial x \partial z}\cos\theta + \frac{\partial^2 v}{\partial x \partial z}\sin\theta.$$

When $x = y = z = 0$ we have $f = u = v = 0$ and the equations simplify to

$$\frac{d^2}{d\theta^2}\frac{\partial f}{\partial x} + \frac{\partial f}{\partial x} = \frac{\partial u}{\partial x}\cos\theta + \frac{\partial v}{\partial x}\sin\theta,$$

$$\frac{d^2}{d\theta^2}\frac{\partial f}{\partial z} + \frac{\partial f}{\partial z} = \frac{\partial u}{\partial z}\cos\theta + \frac{\partial v}{\partial z}\sin\theta,$$

$$\frac{d^2}{d\theta^2}\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial x \partial z} + k''(0)\frac{\partial f}{\partial x}\frac{\partial f}{\partial z} + \frac{\partial f}{\partial x} = \frac{\partial^2 u}{\partial x \partial z}\cos\theta + \frac{\partial^2 u}{\partial x \partial z}\sin\theta.$$

Now the image of the operator $d^2/d\theta^2 + 1$ contains no terms $\cos\theta$ or $\sin\theta$, so we see that $\partial u/\partial x = \partial v/\partial x = \partial u/\partial z = \partial v/\partial z = 0$ at $x = y = z = 0$. Since $f(\theta)$ has for its Fourier series the coefficients x for $\cos\theta$ and y for $\sin\theta$, it follows that $\partial f/\partial x$ has coefficients 1 for $\cos\theta$ and 0 for $\sin\theta$, while $\partial f/\partial z$ has coefficients 0 for both. Therefore at $x = y = z = 0$,

$$\partial f/\partial x = \cos\theta \quad \text{and} \quad \partial f/\partial z = 0.$$

Then examining the last equation, we see that the only way to get a term $\partial f/\partial x = \cos\theta$ on the left is to have $\partial^2 u/\partial x \partial y = 1$ on the right at $x = y = z = 0$. This proves the lemma.

If we persist in the above manner we can find the (formal) power series expansion for $f(\theta)$ and m in powers of x, y, and z. Our calculations show that if

$$k(\sigma) = \alpha + k_2\sigma^2 + k_3\sigma^3 + \cdots$$

then up to terms of degree 3 in x, y and z

$$f(\theta) = x\cos\theta + y\sin\theta - k_2(x^2+y^2)$$
$$+ \tfrac{1}{6}k_2\left[(x^2-y^2)\cos 2\theta + 2xy\sin 2\theta\right] + \cdots,$$
$$m = z - \left(\tfrac{5}{6}k_2^2 - \tfrac{3}{4}k_3\right)(x^2+y^2) + \cdots.$$

At any rate, since $m = 0$ and $\partial m/\partial z = 1$ at the origin, we can solve the equation $m = 0$ in a neighborhood of the origin to obtain z as a smooth function of x and y. More precisely, since m is rotationally invariant we get that z is a smooth function of $x^2 + y^2$. Its series starts like

$$z = \left(\tfrac{5}{6}k_2^2 - \tfrac{3}{4}k_3\right)(x^2 + y^2) + \cdots$$

up to terms of order $(x^2 + y^2)^2$. Then the set of solutions to the equation (*), which is the set of solutions to (**) with $u = v = 0$, corresponds in our parameter space (x, y, z) to the line $x = y = 0$ (consisting of the stable solutions) and a smooth surface $z = \phi(x^2 + y^2)$ invariant under rotations (corresponding to the oscillations).

Finally, recall that the total energy

$$E = \frac{1}{2}m(1 + \varepsilon^2)\left(\frac{ds}{dt}\right)^2 + mgh(s)$$

is a constant of the motion. For our solutions E will be a smooth function of the parameters x and y. Since a change of phase $\theta \to \theta + \alpha$ rotates x and y but does not change the energy, we see that E is a smooth function of $x^2 + y^2$. Indeed in our expansion

$$s = [x \cos vt + y \sin vt]/\tilde{c} + \cdots$$

and hence

$$E = (mg/2\tilde{c})(x^2 + y^2) + \cdots.$$

By the inverse function theorem we can solve for $x^2 + y^2$ as a smooth function of the energy E. Now z is a smooth function of $x^2 + y^2$, v is a smooth function of z near $z = 0$, and $T = 2\pi/v$. Hence the period T is a smooth function of the energy E for $E \geqslant 0$. By an easy manipulation of series

$$T = 2\pi\sqrt{\frac{1 + \varepsilon^2}{\tilde{c}}}\left[1 + \lambda\frac{\tilde{c}}{mg}E + O(E^2)\right]$$

where $\lambda = \frac{5}{6}k_2^2 - \frac{3}{4}k_3$ is a dimensionless constant depending only on the geometry of the trough. When the trough is a circle we have

$$k(\sigma) = \sin\sigma = \sigma - \sigma^3/6 + \cdots$$

from which we see $k_2 = 0$ and $k_3 = -1/6$. This gives $\lambda = +1/8$. Since $\lambda > 0$ the period increases as the energy increases. The same formula applies to a pendulum taking $\varepsilon = 0$.

5.5. *Counterexamples in Fréchet spaces.* In this section we present some counterexamples to the inverse function theorem in Fréchet spaces. They show that any straightforward generalization will fail, and serve to justify the extra definitions and hypotheses which are necessary for the Nash-Moser inverse function theorem. We being with some counterexamples where the derivative is the identity at the origin but is not invertible at points arbitrarily close by.

5.5.1. COUNTEREXAMPLE. Let $\mathcal{C}^\infty[-1, 1]$ be the Fréchet space of smooth functions on $-1 \leqslant x \leqslant 1$ and let P be the differential operator

$$P\colon \mathcal{C}^\infty[-1, 1] \to \mathcal{C}^\infty[-1, 1], \qquad Pf = f - xfdf/dx.$$

Then P is smooth, and the derivative of P is given by

$$DP(f)g = g - xg\frac{df}{dx} - xf\frac{dg}{dx}.$$

At $f = 0$ we have $DP(0)g = g$, so that $DP(0) = I$ is the identity map. Surely nothing could be more invertible than the identity. Since $P(0) = 0$, if the inverse function theorem were true then the image of P should fill out a neighborhood of 0. But it doesn't. Let $g_n = 1/n + x^n/n!$. It is easy to see that $g_n \to 0$ in $\mathcal{C}^\infty[-1, 1]$; for the derivative of $x^n/n!$ is the same sequence again. We shall show that g_n does not lie in the image of P for any $n > 1$. Thus the image of P does not contain any neighborhood of zero.

More generally, if $b_n \neq 0$ then $1/n + b_n x^n$ does not belong to the image of P. This can be seen by examining power series. Every smooth function on $[-1, 1]$ has a formal power series expansion at 0 (which of course does not converge unless the function is analytic). By an elementary computation if

$$f = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$

then

$$Pf = a_0 + (1 - a_0)a_1 x + (a_2 - a_1^2 - 2a_0 a_2)x^2 + (a_3 - 3a_1 a_2 - 3a_0 a_3)x^3 + \cdots.$$

Suppose $Pf = 1/n + b_n x^n$. First we have $a_0 = 1/n$. If $n > 1$ then $a_0 \neq 1$ and $(1 - a_0)a_1 = 0$, so $a_1 = 0$. The next term is then $(1 - 2a_0)a_2 x^2$. If $n = 2$ then $a_0 = \frac{1}{2}$ and this term is zero, which contradicts $Pf = \frac{1}{2} + b_2 x^2$ when $b_2 \neq 0$. If $n > 2$ then $a_0 \neq \frac{1}{2}$ so we conclude $a_2 = 0$ and proceed to the next term, which is then $(1 - 3a_0)a_3 x^3$. In general, we conclude that $a_1 = a_2 = \cdots = a_{k-1} = 0$ and arrive at a term $(1 - ka_0)a_k x^k$. If $k < n$ then $a_0 \neq 1/k$; we conclude that $a_k = 0$ and proceed to the next term. When $k = n$ then since $a_0 = 1/n$ this term must be zero. Thus it cannot equal $b_n x^n$ for $b_n \neq 0$. This gives a contradiction. Hence $1/n + b_n x^n$ cannot be $P(f)$ for any smooth function f if $b_n \neq 0$ and $n > 1$.

It isn't hard to see what is going wrong if we evaluate the derivative at $f = 1/n$ applied to $h = x^k$:

$$DP(1/n)x^k = (1 - k/n)x^k.$$

Thus $DP(1/n)$ kills off the term x^n in the power series. Even though $DP(0) = I$ is the identity, the linear maps $DP(1/n)$ which are arbitrarily close by are not invertible.

Note that this counterexample does not extend to Banach spaces. If we try to write

$$P\colon C^1[-1, 1] \to C^0[-1, 1]?$$

then the derivative $DP(0)$ is the "identity" only as an inclusion of a dense subspace, which is not invertible. Likewise

$$P\colon C^1[-1, 1] \to C^1[-1, 1]?$$

does not work because P has to take a derivative, which lands us in $C^0[-1, 1]$ instead.

5.5.2. COUNTEREXAMPLE. The next example is of considerable geometric interest. If X is a compact manifold then the group of smooth diffeomorphisms $\mathcal{D}(X)$ has the structure of a Fréchet manifold and is a Fréchet Lie group (see Example 4.4.6). Its Lie algebra, the tangent space at the identity map 1, is the space $\mathcal{C}^\infty(X, TX)$ of smooth vector fields on X. The bracket operation of the Lie algebra is the usual Lie bracket of vector fields. Each vector field v on X gives rise to a one-parameter group e^{tv} of diffeomorphisms called the flow of v, obtained as the solution $y = e^{tv}(x)$ of the differential equation

$$dy/dt = v(y)$$

with initial conditions $y = x$ at $t = 0$. This equation may be rewritten as

$$de^{tv}/dt = v \circ e^{tv}$$

which justifies the exponential notation. Note that e^{tv} is the one-parameter subgroup of the Lie group of diffeomorphisms with tangent vector v at the identity.

Evaluating the flow at $t = 1$ we obtain the diffeomorphism e^v called the exponential of the vector field v. The exponential defines a smooth map

$$e^v: \mathcal{C}^\infty(X, TX) \to \mathcal{D}(X)$$

of the Lie algebra to the Lie group. The exponential of the vector field 0 is the identity map $1 = e^0$. The derivative of the exponential map at the vector field 0 is the identification of the vector fields with the tangent space of the diffeomorphisms at the identity map 1. Given a smooth map of a vector space to a manifold whose derivative at a point is the identity map of the vector space to the tangent space of the manifold, we would ordinarily expect to conclude from the inverse function theorem that the map was locally invertible, and hence filled out a neighborhood of the point on the manifold. Thus in our case, we would expect that every diffeomorphism close to the identity was the exponential of some vector field, so that a small diffeomorphism would extend to a one-parameter flow. This is false. The result fails already for the circle, where the following rather surprising result holds instead.

THEOREM. *If a diffeomorphism of the circle without fixed points is the exponential of a vector field, then it must be conjugate to a rotation.*

PROOF. If a vector field has a zero its exponential has a fixed point. Therefore a diffeomorphism without fixed points can only be the exponential of a nonzero vector field. Call the vector field v. Parametrize the circle by a parameter t, say the angle defined modulo 2π, and write $v = v(t)d/dt$. Since $v(t) \neq 0$, we may define a new parameter

$$\theta = c\int \frac{dt}{v(t)} \quad \text{where } c = 1/\int_0^{2\pi} \frac{dt}{v(t)}.$$

Then θ is also a parametrization of the circle modulo 2π, and $v = v(t)d/dt = cd/d\theta$. Thus v is a constant vector field in the parametrization θ. Its exponential is the map

$$e^v: \theta \to \theta + c$$

which is a rotation in the parametrization by θ. The change from the parameter t to the parameter θ is accomplished by some diffeomorphism. Therefore e^v is conjugate by this diffeomorphism to a rotation. This proves the theorem.

Any power of a rotation is also a rotation. If a rotation fixes one point it must be the identity. Thus if a power of a rotation fixes one point it must fix them all. The same must be true of any diffeomorphism conjugate to a rotation. Thus one way to find a diffeomorphism f not conjugate to a rotation is to make $f^k(x) = x$ for some point x but $f^k(y) \neq y$ for some other point y. The rotation $t \to t + 2\pi/k$ by an angle $2\pi/k$ has $f^k(t) \equiv t \pmod{2\pi}$ for all t. If we modify it by an extra little push just in the interval $0 < x < 2\pi/k$, so that $f(0) = 2\pi/k$ but $f(\pi/k) > 3\pi/k$, then $f^k(0) \equiv 0$ but $f^k(\pi/k) \not\equiv \pi/k \pmod{2\pi}$. Then this diffeomorphism cannot be the exponential of a vector field. By making k large and the push small, we can make f as close to the identity as we like in the $\mathcal{C}^\infty$ topology. Therefore the exponential map fails to be locally invertible at the origin, even though its derivative is the identity.

It is not hard to see what goes wrong in this example. Although the derivative of the exponential map is the identity at the origin, it fails to be invertible at nearby points. To see this, we shall compute the derivative explicitly by making an infinitesimal change w in the vector field v. Then the flow y, obtained by solving the differential equation

$$\begin{cases} dy/dt = v(y), \\ y = x \quad \text{at } t = 0 \end{cases}$$

will experience an infinitesimal change z, which will be the solution of the linear differential equation

$$\begin{cases} dz/dt = Dv(y)z + w(y), \\ z = 0 \quad \text{at } t = 0. \end{cases}$$

Just as the exponential map e^v is obtained by evaluating the flow $y = e^{tv}$ at $t = 1$, so the infinitesimal change $De^v w$ in the exponential map is obtained by evaluating the infinitesimal change in the flow z at $t = 1$.

We compute a special case. If $v(x) \neq 0$ at a point x, then $v(y) \neq 0$ all along the flow $y = e^{tv}(x)$, $0 \leqslant t \leqslant 1$, which is a compact set. We can therefore choose a coordinate chart in a neighborhood of this set in which the vector field v is a constant. In this case the exponential map is given by $e^v(x) = x + v$, since the flow is $y = e^{tv}(x) = x + tv$. Since v is constant, $Dv = 0$, and z is the solution of

$$\begin{cases} dz/dt = w(x + tv), \\ z = 0 \quad \text{at } t = 0 \end{cases}$$

which is

$$z = \int_{\tau=0}^{t} w(x + \tau v)\, d\tau.$$

Then evaluating at $t = 1$ we get

$$De^v w(x) = \int_{\tau=0}^{1} w(x + \tau v)\, d\tau$$

which says that $De^v w$, the derivative of the exponential map at v in the direction w, is the average of w over the flow of v.

Consider the case of a circle. If $v = 2\pi/k$ is a small constant vector field, its exponential e^v is a rotation through a small angle $2\pi/k$. The averaging operator

$$De^v w(t) = \int_{\tau=0}^{1} w\left(t + \frac{2\tau\pi}{k}\right) d\tau$$

annihilates the functions $w(t) = \sin kt$ or $\cos kt$. More generally each term in the Fourier series expansion of w is multiplied by some constant, and for these terms that constant is zero. Thus the derivative of the exponential map at a small rotation through an angle $2\pi/k$ is never invertible. Note that the derivative in this counterexample looks like Counterexample 5.3.3.

The preceding examples make it clear that the derivative of an operator P in Fréchet spaces may be invertible at one point but not at other points arbitrarily nearby. If however P is to be locally invertible in a neighborhood then its derivative must be invertible everywhere in that neighborhood. In Banach spaces this would have followed automatically. Therefore it seems reasonable in Fréchet spaces to add this as an extra assumption. Suppose then that $P(f) = g$ defines a smooth map between open sets in Fréchet spaces

$$P\colon U \subseteq F \to V \subseteq G$$

and that for each $f \in U$ the derivative $DP(f)$ is an invertible linear map of F to G. (Note U may be made smaller if necessary.) We write its inverse $DP(f)^{-1}$ as $VP(f)$; this notation avoids confusion between $D(P^{-1})$ and $(DP)^{-1}$. Then $VP(f)k = h$ is the unique solution of the equation $DP(f)h = k$. We regard VP as a function of two variables, that is as a map on the product space

$$VP\colon U \times G \subseteq F \times G \to F$$

the same as we did for DP. Thus we avoid the space of linear maps $L(F, G)$, which is not a Fréchet space. Of course $VP(f)k$ is linear in k for each f. It is reasonable to ask that VP should be continuous and even smooth. We may then ask if, with all these assumptions, P will be locally invertible? The answer is no, as the following example shows.

5.5.3. COUNTEREXAMPLE. Let H be the Fréchet space of functions holomorphic in the entire complex plane, with the topology of uniform convergence on compact sets. Define a map

$$P\colon H \to H, \qquad P(f) = e^f.$$

Then P is smooth and its derivative is given by

$$DP\colon H \times H \to H, \qquad k = DP(f)h = e^f h.$$

The derivative is clearly invertible for all f, and the family of inverses

$$VP\colon H \times H \to H, \qquad h = VP(f)k = e^{-f}k$$

is also smooth. Therefore it is rather surprising that the map P is not locally invertible. Indeed the image of P does not even contain an open set. We can even prove the following precise result on the image of P.

THEOREM. *Let S be the subset of all functions in H which are never zero. Then S is relatively closed in $H - \{0\}$ and contains no open set. The fundamental group $\pi_1(S)$ is the integers Z. The map P: $H \to H$ has image S, and is a local homeomorphism onto S. Moreover P is a covering map and identifies H as the universal cover of S.*

PROOF. If $f(z)$ is an entire holomorphic function which has a zero but is not identically zero, then the zeroes of f are isolated and of finite multiplicity. If γ is a curve on which $f(z)$ is never zero, the number of zeroes of $f(z)$ inside γ counting multiplicities is given by

$$(1/2\pi i)\int_\gamma (f'(z)/f(z))\,dz$$

by Rouché's theorem. The integral is a continuous function in a neighborhood of $f \in H$, and since it is integer-valued it is locally constant. Therefore any function close to f in H still has a zero. This shows that $H - \{0\} - S$ is open, so S is relatively closed in $H - \{0\}$.

Every nonconstant polynomial has a zero by the fundamental theorem of algebra. The polynomials are dense in H, as we see by expanding in a power series, and so the nonconstant polynomials are also dense. Thus S contains no open set.

Since $P(f) = e^f$ is never zero, the image of P lies in S. If $g \in S$, we can choose a branch of the logarithm at $g(0)$ and define $\log g(z)$ in a neighborhood of 0. We can extend $\log g$ by analytic continuation to the entire plane since the plane is simply connected. Then $f = \log g$ solves $e^f = g$. This shows the image of P is all of S.

Suppose $g_j \to g$ in H. Then $g_j \to g$ uniformly on every closed disc. It follows that $\log g_j \to \log g$ uniformly on a disc around 0, and also on each subsequent disc in the analytic continuation. Hence the $f_j = P^{-1}(g_j) = \log g_j$ converge to $f = P^{-1}(g) = \log g$ uniformly on each compact set, so $f_j \to f$ in H. This shows P is a local homeomorphism onto S. Moreover if g_t is a continuous path in S for $0 \leqslant t \leqslant 1$, then g_t is also continuous when restricted to any closed disc. If $P(f_0) = g_0$, we can then choose the $f_t = \log g_t$ to vary continuously on each closed disc by analytic continuation for $0 \leqslant t \leqslant 1$. Thus g_t lifts to a continuous path $f_t \in H$ with $P(f_t) = g_t$. Hence P: $H \to S$ is a covering map.

Since H is simply connected, H is the universal cover of S. For any $g \in S$ the choice of $P^{-1}(g) = \log g$ is unique up to a constant $2\pi ni$ for integer $n \in Z$. This shows that the fundamental group $\pi_1(S) = Z$. Let $\mathbf{C}^* = \mathbf{C} - \{0\}$ be the nonzero complex numbers. Then $\mathbf{C}^* \subseteq S$. There is a projection $S \to \mathbf{C}^*$ given by $f \to f(0)$. We can define a homotopy

$$\phi\colon S \times [0, 1] \to S$$

where

$$\phi f(t, z) = f(tz).$$

Then $\phi f(0, z) = f(0)$ and $\phi f(1, z) = f(z)$. Thus $\mathbf{C}^*$ is a deformation retract of S. This also shows $\pi_1(S) = \pi_1(\mathbf{C}^*) = Z$.

5.5.4. COUNTEREXAMPLE. Let E^n be the Banach space of continuous functions $f(x)$ on the real line which vanish for $x \leqslant 1$ and have

$$\| f \|_n = \sup_x e^{nx} | f(x) | < \infty$$

and let E^∞ be the Fréchet space of functions in E^n for all n. For a product we have the estimate

$$\| fg \|_{p+q} \leqslant \| f \|_p \| g \|_q .$$

Define a linear map $Lf = g$ by letting $g(x) = f(x/2)$. Then we have the estimate

$$\| Lf \|_n \leqslant \| f \|_{2n}$$

so L defines a continuous linear map of E^{2n} into E^n or of E^∞ into E^∞. Define a bilinear map $B\{f, h\} = k$ by letting $k(x) = f(x/2)h(x/2)$. Then we have the estimate

$$\| B\{f, h\} \|_n \leqslant \| f \|_n \| h \|_n$$

so B defines a continuous bilinear map of $E^n \times E^n \to E^n$ for $0 \leqslant n \leqslant \infty$.

Define a nonlinear map $P(f) = g$ by letting

$$g(x) = f(x) - f(x/2)^2$$

so that $P(f) = f - B\{f, f\}$. Then P is a smooth map of E^n to itself for $0 \leqslant n \leqslant \infty$. Its derivative $DP(f)h = k$ is given by

$$k(x) = h(x) - 2f(x/2)h(x/2)$$

so $DP(f)h = h - 2B\{f, h\}$. Note that $DP(0)h = h$ is the identity map. Therefore for $n < \infty$ the map P is invertible as a map of a neighborhood of 0 to a neighborhood of 0 by the inverse function theorem.

We can be more precise. We can solve the equation $P(f) = g$ recursively for f using the formula

$$f(x) = g(x) + f(x/2)^2$$

since the functions all vanish for $x \leqslant 1$. Indeed $f(x) = g(x)$ for $x \leqslant 2$. Let us introduce the norms

$$\| f \|_{n,t} = \sup_{x \leqslant t} e^{nx} | f(x) | .$$

We prove the following result.

THEOREM. *If $g \in E^n$ and $\| g \|_n < \frac{1}{4}$ then the solution $f \in E^n$ and $\| f \|_n < \frac{1}{2}$.*

PROOF. We have $\| g \|_{n,t} < \frac{1}{4}$ for all t. Then $\| f \|_{n,2} < \frac{1}{2}$. Suppose $\| f \|_{n,t/2} < \frac{1}{2}$ for some t. Since

$$f(x) = g(x) + f(x/2)^2$$

we have

$$\|f\|_{n,t} \leqslant \|g\|_{n,t} + \|f\|^2_{n,t/2} < \tfrac{1}{2}$$

also. Thus $\|f\|_n < \frac{1}{2}$.

Observe however that the neighborhoods $\|g\|_n < \frac{1}{4}$ on which P is invertible are shrinking as $n \to \infty$. Therefore it is not so surprising that P is not invertible on E^∞ at all. To see this let

$$f_m(x) = \begin{cases} 0 & \text{if } x \leqslant 1, \\ (x-1)e^{-2m} & \text{if } 1 \leqslant x \leqslant 2, \\ e^{-mx} & \text{if } 2 \leqslant x < \infty. \end{cases}$$

Then f_m is continuous and $f_m \in E^m$ but $f_m \notin E^{m+1}$. Let $g_m = P(f_m)$. Then $g_m(x) = 0$ for $x \geqslant 4$ so $g_m \in E^\infty$ for all m. Now we can compute

$$\|f_m\|_n = e^{2(n-m)}$$

for $m \geqslant n$, so $f_m \to 0$ as $m \to \infty$ in E^n. Since P is continuous as a map of E^n to itself, $g_m \to 0$ in E^n for all n, so $g_m \to 0$ in E^∞. But $P^{-1}(g_m) = f_m \notin E^\infty$ so P is not invertible on any neighborhood of zero.

Now by contrast the derivative of P is always invertible on a neighborhood of zero. Let $U = \{f \in E^\infty : \|f\|_0 < \frac{1}{4}\}$.

THEOREM. *If $f \in U$ then for each $k \in E^\infty$ there exists a unique $h \in E^\infty$ with $DP(f)h = k$. Letting $h = VP(f)k$, the solution map VP: $(U \subseteq E^\infty) \times E^\infty \to E^\infty$ is smooth. Moreover we have the estimate*

$$\|h\|_n \leqslant \|k\|_n + 4\|f\|_{2n}\|k\|_0.$$

PROOF. We begin with the estimate. We can always solve for h recursively from the formula

$$h(x) = k(x) + 2f(x/2)h(x/2)$$

which gives the estimate

$$\|h\|_n \leqslant \|k\|_n + 2\|f\|_{2n}\|h\|_0.$$

When $n = 0$ and $\|f\|_0 < \frac{1}{4}$, we get $\|h\|_0 \leqslant 2\|k\|_0$. Substituting this for the last term gives the desired estimate. To prove that the solution exists, we can repeat the previous argument on $1 \leqslant x \leqslant t$ to get

$$\|h\|_{n,t} \leqslant \|k\|_{n,t} + 4\|f\|_{2n,t}\|k\|_{0,t}$$

and let $t \to \infty$. To prove VP is continuous, choose sequences $f_j \in U$ and $k_j \in E^\infty$ with $f_j \to f \in U$ and $k_j \to k$ in E^∞, and let $h_j = VP(f_j)k_j$ and $h = VP(f)k$. Since

$$h_j(x) = k_j(x) + 2f_j(x/2)h_j(x/2), \qquad h(x) = k(x) + 2f(x/2)h(x/2)$$

we see by subtracting that

$$\|h_j - h\|_0 \leqslant \|k_j - k\|_0 + \tfrac{1}{2}\|h_j - h\|_0 + 2\|f_j - f\|_0\|h\|_0$$

so $\|h_j - h\|_0 \to 0$ as $j \to \infty$. Now

$$h_j = k_j + 2B\{f_j, h_j\}$$

and if $f_j \to f$ in E^∞ and $h_j \to h$ in E^0 then $B\{f_j, h_j\} \to B\{f, h\}$ in E^∞. Thus h_j converges in E^∞ to an element $\bar{h}$ with

$$\bar{h} = k + 2B\{f, \bar{h}\}$$

so $\bar{h} = h$. If VP is continuous then it is smooth by Theorem 5.3.1.

5.6. *Differential equations in Fréchet spaces.* Let U be an open subset of a Fréchet space F and

$$P\colon U \subseteq F \to F$$

a smooth map. We can try to solve the ordinary differential equation

$$df/dt = P(f)$$

for a smooth path $f(t)$ on $a \leqslant t \leqslant b$ with values in F, with given initial condition $f(a) \in F$. For a Banach space the solution will always exist and be unique. For a Fréchet space it may neither exist nor be unique, as the following counterexamples show.

5.6.1. COUNTEREXAMPLES. Let $\mathcal{C}^\infty[0, 1]$ be the Fréchet space of smooth functions $f(x)$ on $0 \leqslant x \leqslant 1$ and consider the linear differential equation

$$\partial f/\partial t = \partial f/\partial x.$$

A path $f(t) \in \mathcal{C}^\infty[0, 1]$ for $0 \leqslant t \leqslant 1$ is a function $f(x, t)$ on the square $0 \leqslant x \leqslant 1$, $0 \leqslant t \leqslant 1$. The above equation has solutions $f(x, t) = f(x + t)$.

Given any smooth function in $\mathcal{C}^\infty[0, 1]$ at $t = 0$, we can extend it to $0 \leqslant x \leqslant 2$. Then the above formula gives a solution for $0 \leqslant t \leqslant 1$. Since the extension is not unique; neither is the solution.

By way of contrast, if we consider the same equation on the space $\mathcal{C}_c^\infty[0, 1]$ of smooth functions of x which vanish outside the interval $0 \leqslant x \leqslant 1$, we see that the solution exists for $0 \leqslant t \leqslant \varepsilon$ if and only if $f(x) = 0$ for $0 \leqslant x \leqslant \varepsilon$. In this case the solution is unique.

Of course, on the space $\mathcal{C}_\sharp^\infty[0, 1]$ of smooth functions on $x \geqslant 0$ which vanish on $x \geqslant 1$, the solution exists for all time and is unique. And on the space $\mathcal{C}_b^\infty[0, 1]$ of smooth functions on $x \leqslant 1$ which vanish for $x \leqslant 0$, the solution does not exist on $0 \leqslant t \leqslant \varepsilon$ unless $f(x) = 0$ for $x \leqslant \varepsilon$, and even in that case it is not unique.

There is an interesting class of differential equations on Fréchet spaces which we can solve.

5.6.2. DEFINITION. Let $P\colon U \subseteq F \to V \subseteq G$ be a smooth map between Fréchet spaces F and G. We say P is a smooth *Banach map* if we can factor $P = Q \circ R$ where $R\colon U \subseteq F \to W \subseteq B$ and $Q\colon W \subseteq B \to V \subseteq G$ are smooth maps and B is a Banach space.

5.6.3. THEOREM. *Let $P\colon U \subseteq F \to F$ be a smooth Banach map of a Fréchet space F into itself. For every $f_0 \in U$ we can find a neighborhood $\tilde{U}$ of f_0 in F and an $\varepsilon > 0$ such that for all $\bar{f} \in \tilde{U}$ the differential equation*

$$df/dt = P(f)$$

has a unique solution with $f(0) = \bar{f}$ on $0 \leqslant t \leqslant \varepsilon$ depending smoothly on t and $\bar{f}$.

PROOF. We can factor $P = Q \circ R$ where R has range and Q has domain in a Banach space B, and Q and R are smooth. Without loss of generality we may assume $f_0 = 0$ and $R(0) = 0$. In the following discussion C will denote various constants independent of the parameters ε, δ, η, and θ.

By Theorem 5.1.3, we can find a seminorm $\| \ \|_F$ on F, an $\eta > 0$ and a constant C such that if $\| f_1 \|_F$, $\| f_2 \|_F \leqslant \eta$ then

$$\| R(f_1) - R(f_2) \|_B \leqslant C \| f_1 - f_2 \|_F$$

for the norm $\| \ \|_B$ on the Banach space B. Since B has only one norm up to equivalence, we can find a $\theta > 0$ and a constant C such that if $\| g_1 \|_B$, $\| g_2 \|_B \leqslant \theta$ then

$$\| Q(g_1) - Q(g_2) \|_F \leqslant C \| g_1 - g_2 \|_B.$$

As special cases, since $R(0) = 0$ and $\| Q(0) \|_F \leqslant C$ and we can take $\theta \leqslant 1$, we get

$$\begin{aligned} \| R(f) \|_B &\leqslant C \| f \|_F \quad \text{if } \| f \|_F \leqslant \eta, \\ \| Q(g) \|_F &\leqslant C \qquad \text{if } \| g \|_B \leqslant \theta. \end{aligned}$$

These are the estimates we shall use.

Let $\tilde{U} = \{ f \in F \colon \| \bar{f} \|_F \leqslant \delta \}$, and let $C([0, \varepsilon], B)$ be the space of continuous paths $g(t) \in B$ on $0 \leqslant t \leqslant \varepsilon$, where $\varepsilon > 0$ and $\delta > 0$ will be chosen small. We set up the map $M(\bar{f}, g(t)) = k(t)$ by letting

$$R(t) = R\Big(\bar{f} + \int_0^t Q(g(\theta))\, d\theta \Big).$$

We claim the following.

5.6.4. LEMMA. *Let ε and δ be small compared to η, and η small compared to θ. Let $\tilde{V}$ be the set of paths $g(t)$ in $C([0, \varepsilon], B)$ with $\| g(t) \|_B \leqslant \theta$. Then M defines a continuous map*

$$M \colon (\tilde{U} \subseteq F) \times \big(\tilde{V} \subseteq C([0, \varepsilon], B) \big) \to \tilde{V} \subseteq C([0, \varepsilon], B)$$

which is a contraction in g for each $\bar{f}$.

PROOF. By our estimates if $g(t) \in \tilde{V}$

$$\sup_t \| \bar{f} + \int_0^t Q(g(\theta))\, d\theta \|_B \leqslant \delta + C\varepsilon \leqslant \eta$$

if δ and ε are small compared to η, and

$$\sup \| k(t) \|_B \leqslant C\eta \leqslant \theta$$

if η is small compared to θ. Hence $k(t) \in \tilde{V}$. If $M(\bar{f}, g_1(t)) = k_1(t)$ and $M(\bar{f}, g_2(t)) = k_2(t)$ then

$$\begin{aligned} \sup_t \| k_1(t) - k_2(t) \|_B &\leqslant C \in \sup_t \| Q(g_1(t)) - Q(g_2(t)) \|_F \\ &\leqslant C \in \sup_t \| g_1(t) - g_2(t) \|_B \end{aligned}$$

so M is a contraction of $C\varepsilon < 1$.

It follows from the contraction mapping theorem that for each $\bar{f} \in \tilde{U}$ there exists a unique $g(t) \in \tilde{V}$ with $M(\bar{f}, g(t)) = g(t)$. Moreover the solution $g(t)$ depends continuously on $\bar{f}$. We let

$$f(t) = \bar{f} + \int_0^t Q(g(\theta))\, d\theta.$$

Then $g(t) = R(f(t))$. Since $P = Q \circ R$,

$$f(t) = \bar{f} + \int_0^t P(f(\theta))\, d\theta$$

which implies that, for $0 \leqslant t \leqslant \varepsilon$,

$$df/dt = P(f)$$

and $f = \bar{f}$ at $t = 0$.

We write $f(t) = S(\bar{f})$ for the solution. We know that $g(t) \in C([0, \varepsilon], B)$ depended continuously on $\bar{f}$. Now the space of continuous maps of $[0, \varepsilon]$ into F is a Fréchet space $C([0, \varepsilon], F)$ with norms

$$\| f \|_n = \sup_t \| f(t) \|_n \quad \text{for } \| \ \|_n \text{ on } F.$$

We see that the solution map

$$S\colon \tilde{U} \subseteq F \to C([0, \varepsilon], F)$$

is continuous. Moreover from the equation we see that the solution $f(t)$ is smooth in t, and S defines a continuous map

$$S\colon \tilde{U} \subseteq F \to C^\infty([0, \varepsilon], F)$$

into the Fréchet space of smooth paths on $0 \leqslant t \leqslant \varepsilon$ with values in F, which has seminorms

$$\| f \|_{n,k} = \sup_t \| D^k f(t) \|_n \quad \text{for } \| \ \|_n \text{ on } F.$$

To see that $f(t)$ depends smoothly on $\bar{f}$ we have recourse to the following strategem. In Lemma 3.3.1 we saw that we can find a smooth map $L(f_0, f_1)h$ linear in h with

$$P(f_1) - P(f_0) = L(f_0, f_1)(f_1 - f_0).$$

Since L is obtained by averaging DP and P factors through a Banach space, so does L. Therefore we can solve the system

$$\begin{aligned} df_0/dt &= P(f_0), \\ df_1/dt &= P(f_1), \\ dh/dt &= L(f_0, f_1)h \end{aligned}$$

with $f_0 = \bar{f}_0$, $f_1 = \bar{f}_1$, and $h = \bar{h}$ at $t = 0$. Let us write $f_0 = S(\bar{f}_0)$, $f_1 = S(\bar{f}_1)$, and $h = M(\bar{f}_0, \bar{f}_1)\bar{h}$. These all exist if $\bar{f}_0$ and $\bar{f}_1$ are near a given point and $\bar{h}$ is near zero. Since the solution h is homogeneous of degree 1 in $\bar{h}$, it will exist for all $\bar{h} \in F$ and be linear in $\bar{h}$. Moreover S and M are continuous in $\bar{f}_0, \bar{f}_1$, and h.

Subtracting the first two equations we see that

$$d(f_1 - f_0)/dt = L(f_0, f_1)(f_1 - f_0)$$

so that

$$S(\bar{f}_1) - S(\bar{f}_0) = M(\bar{f}_0, \bar{f}_1)(\bar{f}_1 - \bar{f}_0).$$

Then by the reverse side of Lemma 3.3.1 we see that the map $S(\bar{f})$ is C^1 in $\bar{f}$. In addition

$$DS(\bar{f})\bar{h} = M(\bar{f}, \bar{f})\bar{h} = h$$

is the solution of the equation

$$dh/dt = DP(f)h$$

with $h = \bar{h}$ at $t = 0$.

In terms of the tangent functor

$$TP(f, h) = (P(f), DP(f)h)$$

we see that if $f = S(\bar{f})$ is the solution of

$$df/dt = P(f)$$

with $f = \bar{f}$ at $t = 0$, then $(f, h) = TS(\bar{f}, \bar{h})$ is the solution of

$$d(f, h)/dt = TP(f, h)$$

with $(f, h) = (\bar{f}, \bar{h})$ at $t = 0$. Now if each solution S is C^r, then TS is also C^r so S is C^{r+1}. It follows that S is C^∞ in $\bar{f}$.

5.6.5. COROLLARY. *Let $P(f, g)$ be a smooth Banach map and solve the differential equation*

$$df/dt = P(f, g)$$

with $f = 0$ at $t = 0$. Then the solution $f = S(g)$ is a smooth function of g.

PROOF. We adjoin a new variable h in the same Fréchet space as g, and solve the equation for the pair (f, h)

$$df/dt = P(f, h), \qquad dh/dt = 0$$

with $f = 0$ and $h = g$ at $t = 0$. Then we apply our previous result. The same trick works for time-dependent equations

$$df/dt = P(f, g, t)$$

where we adjoin a new variable x and solve

$$df/dt = P(f, g, x), \qquad dx/dt = 1$$

with $x = 0$ at $t = 0$. This proves the following.

5.6.6. COROLLARY. *If $P(f, g, t)$ is a smooth Banach map, the differential equation*

$$df/dt = P(f, g, t)$$

has a unique solution with $f = \bar{f}$ at $t = \alpha$, and on a neighborhood of a given $\bar{f}_0$ and g_0 it exists for a fixed time $\varepsilon > 0$ and depends smoothly on $\bar{f}$ and g.

PART II. THE NASH-MOSER CATEGORY

II.1. Tame Fréchet spaces.

1.1. *Graded Fréchet spaces.* The Nash-Moser inverse function theorem works in the category of tame Fréchet spaces and smooth tame maps, which we shall define in this part. Unfortunately the definitions involve a structure which is more refined than the topology of the Fréchet space. Our counterexamples in I.5.5 show that this extra structure is actually necessary, and that the current results are in some sense the best possible.

The topology of a Fréchet space is defined by a countable collection of seminorms $\{\| \ \|_n\}$ with the condition that a sequence $f_j \to f$ if and only if $\| f_j - f \|_n \to 0$ for all n as $j \to \infty$. We say that the seminorms define the topology. Given a collection of seminorms, the topology is uniquely defined. But clearly there may be many collections of seminorms which would define the same topology. We wish to distinguish one collection.

1.1.1. DEFINITION. A *grading* on a Fréchet space is a collection of seminorms $\{\| \ \|_n : n \in J\}$ indexed by the integers $J = \{0, 1, 2, \dots\}$ which are increasing in strength, so that

$$\| f \|_0 \leqslant \| f \|_1 \leqslant \| f \|_2 \leqslant \cdots$$

and which define the topology. A *graded Fréchet space* is one with a choice of grading.

We can of course insure that a collection increases in strength by adding to each seminorm all the lower ones.

1.1.2. EXAMPLES. (1) Let B be a Banach space with norm $\| \ \|_B$. Then B is a graded space with $\| f \|_n = \| f \|_B$ for all n.

(2) Let $\Sigma(B)$ denote the space of all sequences $\{f_k\}$ of elements in a Banach space B such that

$$\| \{f_k\} \|_n = \sum_{k=0}^{\infty} e^{nk} \| f_k \|_B < \infty$$

for all $n \geqslant 0$. Then $\Sigma(B)$ is a graded space with the norms above.

(3) Let X be a measure space with measure μ, let $w \geqslant 0$ be a positive weight function, and let $L_1^\infty(X, \mu, w)$ be the space of all measurable functions f such that

$$\| f \|_n = \int_X e^{nw} |f| \, d\mu < \infty$$

for all $n \geqslant 0$. Then $L_1^\infty(X, \mu, w)$ is a graded space with the norms above.

(4) Let X be a compact manifold. Then $\mathcal{C}^\infty(X)$ is a graded space with

$$\| f \|_n = \| f \|_{C^n(X)}$$

where $C^n(X)$ is the Banach space of functions with continuous partial derivatives of degree $\leqslant n$. If V is a vector bundle over X then the space $\mathcal{C}^\infty(X, V)$ of smooth sections of V is also a graded space.

(5) Let $\mathcal{P}$ be the space of entire holomorphic functions which are periodic with period $2\pi i$ and bounded in each left half-plane. Then $\mathcal{P}$ is a graded space with norms

$$\| f \|_n = \sup\{|f(z)| : \operatorname{Re} z = n\}.$$

A closed subspace of a graded space is again a graded space with the induced norms. A cartesian product $F \times G$ of two graded spaces F and G is a graded space with norms

$$\|(f, g)\|_n = \| f \|_n + \| g \|_n.$$

1.1.3. DEFINITION. We say that two gradings $\{\| \|_n\}$ and $\{\| \|'_n\}$ are *tamely equivalent* of degree r and base b if

$$\| f \|_n \leqslant C \| f \|'_{n+r} \quad \text{and} \quad \| f \|'_n \leqslant C \| f \|_{n+r}$$

for all $n \geqslant b$ (with a constant C which may depend on n). All of the definitions and theorems in this paper will remain valid when a grading is replaced by a tamely equivalent one (except for a few remarks on precise norm estimates).

1.1.4. EXAMPLES. (1) If B is a Banach space, then the space $\Sigma(B)$ of exponentially decreasing sequences in B has the following equivalent gradings:

(a) $\| \{f_k\} \|_n = \| \{f_k\} \|_{l_1^n(B)} = \Sigma_k e^{nk} \| f_k \|_B$,

(b) $\| \{f_k\} \|_n = \| \{f_k\} \|_{l_q^n(B)} = \{\Sigma_k e^{qnk} \| f_k \|_B^q\}^{1/q}$, or

(c) $\| \{f_k\} \|_n = \| \{f_k\} \|_{l_\infty^n(B)} = \sup_k e^{nk} \| f_k \|_B$.

If $r > 0$ and $C = \Sigma_k e^{-rk} < \infty$ then

$$\| \{f_k\} \|_{l_\infty^n(B)} \leqslant \| \{f_k\} \|_{l_q^n(B)} \leqslant \| \{f_k\} \|_{l_1^n(B)} \leqslant C \| \{f_k\} \|_{l_\infty^{n+r}(B)}.$$

(2) The space $\mathcal{P}$ of entire holomorphic functions periodic with period $2\pi i$ and bounded in each left half-plane has the following equivalent gradings:

(a) $$\| f \|_n = \| f \|_{L_1^n} = \frac{1}{2\pi} \int_n^{n+2\pi i} |f(z)| \, dz,$$

(b) $$\| f \|_n = \| f \|_{L_q^n} = \left\{ \frac{1}{2\pi} \int_n^{n+2\pi i} |f(z)|^q \, dz \right\}^{1/q},$$

(c) $$\| f \|_n = \| f \|_{L_\infty^n} = \sup\{|f(z)| : \operatorname{Re} z = n\}$$

where the integrals are taken over a path with $\operatorname{Re} z = n$. We have estimates for $1 < q < \infty$ and $r > 0$

$$\| f \|_{L_1^n} \leqslant \| f \|_{L_q^n} \leqslant \| f \|_{L_\infty^n} \leqslant C \| f \|_{L_1^{n+r}}.$$

The final estimate follows from the Cauchy integral formula

$$f(z) = \frac{1}{2\pi i} \int_{n+r}^{n+r+2\pi i} \frac{f(\zeta) \, d\zeta}{\zeta - z}$$

with the integral taken over a path with $\operatorname{Re} \zeta = n + r$; we see that $C = 1/r < \infty$ when $r > 0$.

(3) If X is a compact manifold, then the following gradings on $\mathcal{C}^\infty(X)$ are equivalent:

(a) the supremum norms $\|f\|_n = \|f\|_{C^n(X)}$,

(b) the Hölder norms $\|f\|_n = \|f\|_{C^{n+\alpha}(X)}$ for $0 < \alpha < 1$,

(c) the Sobolev norms $\|f\|_n = \|f\|_{L_p^n(X)}$ for $1 < p < \infty$,

(d) the Besov norms $\|f\|_n = \|f\|_{B_{p,q}^{n+\alpha}(X)}$ for $0 < \alpha < 1$, $1 < p < \infty$, $1 \leqslant q \leqslant \infty$.

For example, by the Sobolev embedding theorem if $r > \dim X/p$ then

$$\|f\|_{C^n(X)} \leqslant C\|f\|_{L_p^{n+r}(X)}.$$

(4) If $\{\|\ \|_n\}$ is a grading and $\|f\|'_n = \|f\|_{2n}$, then $\{\|\ \|'_n\}$ is also a grading but is *not* tamely equivalent to the first.

(5) Let $L_2^\infty(R^d)$ denote the Fréchet space of smooth functions on R^n all of whose derivatives lie in L_2. Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be a positive multi-index and define weighted Sobolev norms

$$\|f\|_n^\alpha = \left\{\int_{R^d} \left(1 + \xi_1^{2\alpha_1} + \cdots + \xi_d^{2\alpha_d}\right) |\hat{f}(\xi)|^2 \, d\xi\right\}^{1/2}.$$

Then $L_2^\infty(R^d)$ is a graded space for any of the gradings $\{\|\ \|_n^\alpha\}$, but they are *not* tamely equivalent for distinct α. Weighted gradings are very useful for parabolic PDE's, where two space derivatives count for one time derivative.

1.2. *Tame linear maps.*

1.2.1. DEFINITION. We say that a linear map $L\colon F \to G$ of one graded space into another satisfies a *tame estimate* of degree r and base b if

$$\|Lf\|_n \leqslant C\|f\|_{n+r}$$

for each $n \geqslant b$ (with a constant C which may depend on n). We say L is *tame* if it satisfies a tame estimate for some r and b. A tame linear map is automatically continuous in the Fréchet space topologies.

1.2.2. EXAMPLES. (1) Define $L\colon \Sigma(B) \to \Sigma(B)$ by $(Lf)_k = e^{rk} f_k$. Then $\|Lf\|_n \leqslant \|f\|_{n+r}$ so L is tame.

(2) Define $L\colon L_1^\infty(X, \mu, w) \to L_1^\infty(X, \mu, w)$ by $Lf = w^r f$. Then $\|Lf\|_n \leqslant \|f\|_{n+r}$ so L is tame.

(3) Let $L\colon \mathcal{C}^\infty(X) \to \mathcal{C}^\infty(X)$ be any linear partial differential operator of degree r. Then $\|Lf\|_n \leqslant C\|f\|_{n+r}$ so L is tame.

(4) Let $\mathcal{P}$ be the space of entire holomorphic functions periodic with period $2\pi i$ and bounded in each left half-plane, with

$$\|f\|_n = \sup\{|f(z)| : \operatorname{Re} z = n\}.$$

Define $L\colon \mathcal{P} \to \mathcal{P}$ by $Lf(z) = f(2z)$. Then $\|Lf\|_n \leqslant \|f\|_{2n}$, so L is *not* tame.

(5) Let $\mathcal{C}^\infty[a, b]$ denote the space of smooth functions on the interval $[a, b]$ with the grading

$$\|f\|_n = \sup_{k \leqslant n} \sup_{a \leqslant x \leqslant b} |D^k f(x)|.$$

Define a linear map $L\colon \mathcal{C}^\infty[0, 1] \to \mathcal{C}^\infty[-1, 1]$ by $Lf(x) = f(x^2)$. Then L is tame and $\|Lf\|_n \leqslant C\|f\|_n$. The image of L is the closed subspace $\mathcal{C}_S^\infty[-1, 1]$ of

symmetric functions with $f(-x) = f(x)$, because the (formal) Taylor series of a symmetric function contains only even powers of x. Since L is one-to-one, the inverse map L^{-1}: $\mathcal{C}_S^\infty[-1, 1] \to \mathcal{C}^\infty[0, 1]$ exists and is continuous by the open mapping theorem. However L^{-1} is *not* tame. For the $2n$th coefficient of the Taylor series for $Lf = g$ becomes the nth coefficient for the Taylor series of $f = L^{-1}g$, so the best estimate possible is

$$\| L^{-1}g \|_n \leqslant C \| g \|_{2n}.$$

(6) Define L: $\Sigma(B) \to \Sigma(B)$ by $(Lf)_k = f_{2k}$. Then $\| Lf \|_n \leqslant \| f \|_{2n}$ so L is *not* tame.

If L satisfies a tame estimate of degree r and base b, then it also satisfies one of degree $\tilde{r}$ and base $\tilde{b}$ for $\tilde{r} \geqslant r$ and $\tilde{b} \geqslant b$. If L is tame with respect to one pair of gradings on F and G then it is also tame with respect to any pair of tamely equivalent gradings.

1.2.3. DEFINITION. We say L is a *tame isomorphism* if L is a linear isomorphism and both L and L^{-1} are tame. Note that two gradings on a space are tamely equivalent if and only if the identity map is a tame isomorphism from the space with one grading to the space with the other.

A composition of tame linear maps is tame. Indeed if L satisfies a tame estimate of degree r and base b, while M satisfies a tame estimate of degree s and base $b + r$, then LM satisfies a tame estimate of degree $r + s$ and base b; for if $n \geqslant b$

$$\| LMf \|_n \leqslant C \| Mf \|_{n+r} \leqslant C \| f \|_{n+r+s}.$$

1.3. *Tame Fréchet spaces.*

1.3.1. DEFINITION. Let F and G be graded spaces. We say that F is a *tame direct summand* of G if we can find tame linear maps L: $F \to G$ and M: $G \to F$ such that the composition ML: $F \to F$ is the identity

$$F \xrightarrow{L} G \xrightarrow{M} F.$$

1.3.2. DEFINITION. We say a graded space is *tame* if it is a tame direct summand of a space $\Sigma(B)$ of exponentially decreasing sequences in some Banach space B.

1.3.3. LEMMA. *A tame direct summand of a tame space is tame.*

PROOF. If F is a tame direct summand of G and G is a tame direct summand of H, then F is a tame direct summand of H.

1.3.4. LEMMA. *A cartesian product of two tame spaces is tame.*

PROOF. If F is a tame direct summand of $\Sigma(B)$ and G is a tame direct summand of $\Sigma(C)$, then $F \times G$ is a tame direct summand of $\Sigma(B) \times \Sigma(C) = \Sigma(B \times C)$.

1.3.5. LEMMA. *The space $L_1^\infty(X, \mu, w)$ is tame.*

PROOF. Recall that this is the space of all measurable functions with

$$\| f \|_n = \int_X e^{nw} |f| \, d\mu < \infty$$

for all n. Let $X_k = \{k \leqslant w < k+1\}$ and let χ_k be the characteristic function of X_k. Define maps

$$L_1^\infty(X,\mu,w) \overset{L}{\to} \Sigma(L_1(X,\mu)) \overset{M}{\to} L_1^\infty(X,\mu,w)$$

by $(Lf)_k = \lambda_k f$ and $M\{f_k\} = \Sigma_k \chi_k f_k$. Since

$$\|\{f_k\}\|_n = \sum e^{nk} \|f_k\|_{L_1(X,\mu)}$$

it is easy to verify that $\|Lf\|_n \leqslant C\|f\|_n$ and $\|M\{f_k\}\|_n \leqslant C\|\{f_k\}\|_n$.

Since $ML = I$, $L_1^\infty(X,\mu,w)$ is a tame direct summand of $\Sigma(L_1(X,\mu))$.

1.3.6. THEOREM. *If X is a compact manifold then $\mathcal{C}^\infty(X)$ is tame.*

PROOF. The strategy is to write $\mathcal{C}^\infty(X)$ as a tame direct summand of a space $L_1^\infty(X,\mu,w)$ where X is a Euclidean space R^d with coordinates $\xi = (\xi_1,\ldots,\xi_d)$, $\mu = d\xi$ is Lebesgue measure, and the weight function $w = \log(1 + |\xi|)$ where $|\xi|^2 = \xi_1^2 + \cdots + \xi_d^2$. This will of course be accomplished by the Fourier transform.

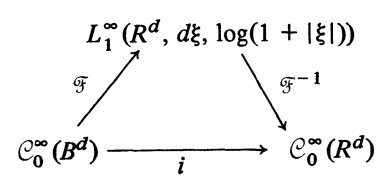
To begin, embed X in some Euclidean space R^d, which we can always do if $d \geqslant 2\dim X + 1$. Since X is compact it will lie in a large ball B^d. Let $\mathcal{C}_0^\infty(R^d)$ denote the graded space of smooth functions on R^d for which all derivatives tend to zero at infinity with the grading

$$\|f\|_n = \sup_{|\alpha| \leqslant n} \sup_X |D^\alpha f(x)|$$

where the sup runs over all partial derivatives D^α of degree $|\alpha|$ at most n. Let $\mathcal{C}_0^\infty(B^d)$ denote the closed subspace of functions which vanish outside the ball B^d, with the induced grading. We can find a continuous linear extension operator ε: $\mathcal{C}^\infty(X) \to \mathcal{C}_0^\infty(B^d)$ satisfying $\|\varepsilon f\|_n \leqslant C\|f\|_n$. It suffices to choose a tubular neighborhood of X, extend the function to be constant along the fibres, and cut off with a smooth bump function. If i denotes the natural inclusion and ρ the restriction map then $\|if\|_n \leqslant C\|f\|_n$ and $\|\rho f\|_n \leqslant C\|f\|_n$ and we have three tame linear maps

$$\mathcal{C}^\infty(X) \overset{\varepsilon}{\to} \mathcal{C}_0^\infty(B^d) \overset{i}{\to} \mathcal{C}_0^\infty(R^d) \overset{\rho}{\to} \mathcal{C}^\infty(X)$$

whose composition is the identity. We proceed to factor i with the Fourier transform $\mathcal{F}$ and its inverse



Functions in $\mathcal{C}_0^\infty(B^d)$ are of Schwartz class $\mathcal{S}$, so their Fourier transforms exists. Since the Fourier transform takes a derivative D^α of degree $|\alpha| \leqslant n$ into multiplication by a polynomial ζ^α of degree $|\alpha| \leqslant n$, we can relate the derivative norms to the power norms. The Fourier transform naturally takes

functions in L_1 into functions in $\mathcal{C}_0$, so the map $\mathcal{F}^{-1}$ is well defined and satisfies

$$\|\mathcal{F}^{-1}g\|_n \leqslant C\|g\|_n.$$

Then $\mathcal{F}^{-1}$ is tame. The same is true for $\mathcal{F}$; however the estimate is not so precise. Since B^d is compact we can convert a sup norm on B^d into an L_1 norm. But when we apply $\mathcal{F}$ we get a sup norm estimate on the transform. To convert this to an L_1 norm we must consume a power $(1 + |\zeta|)^r$ with $r > d$. This gives an estimate $\|\mathcal{F}f\|_n \leqslant C\|f\|_{n+r}$, which shows $\mathcal{F}$ also is tame. This completes the representation of $\mathcal{C}^\infty(X)$ as a tame direct summand of

$$L_1^\infty\big(R^d, d\zeta, \log(1 + |\zeta|)\big),$$

which is itself a tame direct summand of $\Sigma(L_1(R^d, d\zeta))$.

1.3.7. COROLLARY. *If X is a compact manifold with boundary then $\mathcal{C}^\infty(X)$ is also tame.*

PROOF. Let $\tilde{X}$ be the double of X. We can write $\mathcal{C}^\infty(X)$ as a tame direct summand of $\mathcal{C}^\infty(\tilde{X})$ by choosing an extension map ε such that its composition with the natural restriction map is the identity

$$\mathcal{C}^\infty(X) \xrightarrow{\varepsilon} \mathcal{C}^\infty(\tilde{X}) \xrightarrow{\rho} \mathcal{C}^\infty(X).$$

We can define ε first in local coordinates and then patch together with a partition of unity. Let R^n have coordinates (x, y) with $x \in R$ and $y \in R^{n-1}$, and let R^n_+ be the subspace with $x \geqslant 0$. If $f(x, y)$ is a function on R^n_+, we extend it to the other side by putting

$$f(-x, y) = \int_{t=0}^{\infty} \phi(t) f(tx, y)\, dt$$

where $\phi(t)$ is a function which satisfies the condition

$$\int_0^\infty t^n \phi(t)\, dt = (-1)^n$$

for $n = 0, 1, 2, \ldots$, which makes all the derivatives match up along the boundary $x = 0$. An example of such a function is

$$\phi(t) = \frac{e^{2\sqrt{2}}}{\pi(1 + t)} e^{-(t^{1/4} + t^{-1/4})} \sin(t^{1/4} - t^{-1/4})$$

as the reader may "easily" verify by writing $\phi(t)$ as the real part of a holomorphic function and using contour integration.

Since $\phi(t)$ goes to zero very rapidly as t approaches zero or infinity, there is no problem with the convergence of the integral, and we have an estimate $\|\varepsilon f\|_n \leqslant C\|f\|_n$ for the sup norms $\|f\|_n = \|f\|_{C^n}$ (or the Hölder, Sobolev and Besov norms too). Since $\|\rho f\|_n \leqslant C\|f\|_n$ also, we have written $\mathcal{C}^\infty(X)$ as a tame direct summand of $\mathcal{C}^\infty(\tilde{X})$. The above construction generalizes easily to manifolds with interior corners, which are locally diffeomorphic to products of lines R and half-lines R^+. Observe that if we extend in two different directions by the above operator the extensions commute.

1.3.8. COROLLARY. *If X is a compact manifold with boundary and $\mathcal{C}_0^\infty(X)$ is the closed subspace of functions in $\mathcal{C}^\infty(X)$ which vanish on the boundary together with all their derivatives, then $\mathcal{C}_0^\infty(X)$ is tame.*

PROOF. Beware that $\mathcal{C}_0^\infty(X)$ is not even a continuous direct summand of $\mathcal{C}^\infty(X)$. However $\mathcal{C}_0^\infty(X)$ is a tame direct summand of $\mathcal{C}^\infty(\tilde{X})$ for the double $\tilde{X}$. Let X' denote the "mirror image" of X in $\tilde{X}$, and let ε' and ρ' denote the extension and restriction operators defined before for X'. Then $\rho'\varepsilon' = I$ so $\varepsilon'\rho'$ is a projection operator on $\mathcal{C}^\infty(\tilde{X})$. The complementary projection $1 - \varepsilon'\rho'$ has the subspace $\mathcal{C}_0^\infty(X)$ as its image, so it is a direct summand. Since ε' and ρ' are tame, so is $1 - \varepsilon'\rho'$.

1.3.9. COROLLARY. *If X is a compact manifold and V is a vector bundle over X, then the space $\mathcal{C}^\infty(X, V)$ of sections of V over X is tame.*

PROOF. Let $I = X \times R$ be the trivial bundle. Then $\mathcal{C}^\infty(X, I) = \mathcal{C}^\infty(X)$. If I^d denotes the d-fold Whitney sum of I with itself, then $\mathcal{C}^\infty(X, I^d)$ is the cartesian product of d copies of $\mathcal{C}^\infty(X, I)$, and hence is tame. We can write any vector bundle V as a direct summand of some trivial bundle I^d. Since

$$\mathcal{C}^\infty(X, V \oplus W) = \mathcal{C}^\infty(X, V) \times \mathcal{C}^\infty(X, W)$$

it follows that $\mathcal{C}^\infty(X, V)$ is a tame direct summand of $\mathcal{C}^\infty(X, I^d)$ and hence is tame.

1.3.10. THEOREM. *The space $\mathcal{P}$ of entire holomorphic functions periodic of period $2\pi i$ and bounded in each left half-plane is tame.*

PROOF. We claim that $\mathcal{P}$ is tamely isomorphic to the space $\Sigma(R)$ of exponentially decreasing sequences of real numbers. Any function $f \in \mathcal{P}$ may be expanded in a series

$$f(z) = \sum_k c_k e^{kz}.$$

We define the isomorphism $\mathcal{P} \to \Sigma(R)$ by identifying the function f in $\mathcal{P}$ with the series $\{c_k\}$ in $\Sigma(R)$. We can recover the constants c_k by the integral formula

$$c_k = \frac{1}{2\pi i} \int_n^{n+2\pi i} e^{-kz} f(z)\, dz$$

for any value of n, since the integral of a periodic function over a period is independent of the path.

$\mathcal{P}$ has two equivalent gradings

$$\| f \|_{L_1^n} = \frac{1}{2\pi} \int_n^{n+2\pi i} |f(z)|\, dz, \qquad \| f \|_{L_\infty^n} = \sup\{|f(z)| : \operatorname{Re} z = n\}$$

and $\Sigma(R)$ has two equivalent gradings

$$\| \{c_k\} \|_{l_1^n} = \sum_k e^{nk} |c_k|, \qquad \| \{c_k\} \|_{l_1^n} = \sup_k e^{nk} |c_k|.$$

We have the obvious estimates

$$\| \{c_k\} \|_{l_\infty^n} \leq \| f \|_{L_1^n} \leq \| f \|_{L_\infty^n} \leq \| \{c_k\} \|_{l_1^n}$$

which shows that $\mathcal{P}$ is tamely isomorphic to $\Sigma(R)$.

1.3.11. EXAMPLE. Let $E[0,\infty)$ be the space of continuous functions $f(x)$ defined on $0 \leqslant x < \infty$ which go to zero faster than any exponential in x, with grading given by the norms

$$\| f \|_n = \sup_x e^{nx} |f(x)| .$$

Then $E[0,\infty)$ is tame. To see this, let $C[0,2]$ be the Banach space of continuous functions on $0 \leqslant x \leqslant 2$. Define a linear map

$$L: E[0,\infty) \to \Sigma(C[0,2])$$

by letting $Lf = \{g_k\}$ where $g_k(x) = f(x+k)$ for $0 \leqslant x \leqslant 2$. Next choose continuous functions $\phi_k(x)$ with $\phi_0(x) = 1$ for $0 \leqslant x \leqslant 1$ and vanishing for $x \geqslant 2$, while $\phi_k(x)$ vanishes for $x \leqslant k$ or $x \geqslant k+2$ and $\Sigma\phi_k = 1$. Define a linear map

$$M: \Sigma(C[0,2]) \to E[0,\infty)$$

by letting $M\{g_k\} = f$ where

$$f(x) = \sum_{k=0}^{\infty} \phi_k(x) g_k(x-k).$$

Then L and M are easily seen to be tame. Thus $E[0,\infty)$ is a tame direct summand of $\Sigma(C[0,2])$, and hence is tame.

Let $E_0[1,\infty)$ be the closed subspace of functions in $E[0,\infty)$ vanishing for $0 \leqslant x \leqslant 1$. It is easy to write $E_0[1,\infty)$ as a tame direct summand of $E[0,\infty)$, so it is also tame. Note that $E_0[1,\infty)$ is the space of E^∞ of Counterexample 5.3.4. To define the splitting map $E[0,\infty) \to E_0[1,\infty)$, we restrict the function to $0 \leqslant x \leqslant 1$, extend the restriction to $1 \leqslant x \leqslant 2$ by reflection, cut off by a continuous bump function equal to 1 on $0 \leqslant x \leqslant 1$ and 0 for $x \geqslant 2$, and subtract the result from the original function.

The linear map $Lf(x) = f(x/2)$ defines an isomorphism of $E[0,\infty)$ onto itself. Since $\| Lf \|_n = \| f \|_{2n}$ the map L is *not* tame.

II.2. Tame maps.

2.1. *Definition of a tame map.*

2.1.1. DEFINITION. Let F and G be graded spaces and P: $U \subseteq F \to G$ a nonlinear map of a subset U of F into G. We say that P satisfies a *tame estimate* of degree r and base b if

$$\| P(f) \|_n \leqslant C(1 + \| f \|_{n+r})$$

for all $f \in U$ and all $n \geqslant b$ (with a constant C which may depend on n). We say that P is a *tame* map if P is defined on an open set and is continuous, and satisfies a tame estimate in a neighborhood of each point. (We allow the degree r, base b, and constants C to vary from neighborhood to neighborhood.)

2.1.2. EXAMPLE. Let $E[0,\infty)$ be the Fréchet space of continuous functions $f(x)$ on $0 \leqslant x < \infty$ decreasing faster than any exponential, graded by

$$\| f \|_n = \sup_x e^{nx} |f(x)| ,$$

introduced in Example 1.3.11. We define a map

$$P\colon E[0,\infty) \to E[0,\infty)$$

by letting $P(f) = f^2$. Then P is tame. To see this we observe that

$$\|fg\|_{p+q} \leqslant \|f\|_p \|g\|_q$$

for all p and q. Hence if $\|f\|_0 \leqslant C$ then $\|P(f)\|_n \leqslant C\|f\|_n$. This shows that P satisfies a tame estimate of degree 0 and base 0 on the balls $\{\|f\|_0 < C\}$, which are open and fill out the space.

2.1.3. COUNTEREXAMPLE. Let $\mathcal{P}$ be the space of entire holomorphic functions periodic with period $2\pi i$ and bounded in each left half-plane, with norms

$$\|f\|_n = \sup\{|f(z)| : \operatorname{Re} z = n\}.$$

Define an operator $E\colon \mathcal{P} \to \mathcal{P}$ by $Ef = e^f$. Then E is *not* tame. For any neighborhood of zero contains a set $\{f\colon \|f\|_b \leqslant \delta\}$ for some $b < \infty$ and $\delta > 0$. Let $f_k(z) = \delta e^{k(z-b)}$. Then $f_k \in \mathcal{P}$ and $\|f_k\|_b = \delta$, while $\|f_k\|_n = \delta e^{k(n-b)}$ and $\|Pf_k\|_n = e^{\delta e^{k(n-k)}}$. If E satisfies an estimate $\|Ef\|_n \leqslant C(1 + \|f\|_{n+r})$ for any $n \geqslant b$ on the set $\{f\colon \|f\|_b \leqslant \delta\}$, then we would have an estimate

$$e^{\delta e^{k(n-b)}} \leqslant C\delta e^{k(n+r-b)}$$

for a fixed n, b, δ, and C and all $k \to \infty$. This is impossible, so E is not tame. The best estimate is the obvious one $\|e^f\|_n \leqslant e^{\|f\|_n}$. Notice how differently the estimates for e^f behave in $\mathcal{P}$ and $C^\infty[a, b]$. (See Example 2.2.4.)

2.1.4. EXAMPLE. Any continuous map of a graded Fréchet space into a Banach space is tame. Any continuous map of a finite dimensional space into a graded Fréchet space is tame.

PROOF. Let $P\colon U \subseteq F \to B$ be a continuous map into a Banach space B. Given $f_0 \in U$, pick a constant $C > \|P(f_0)\|$ and let $\tilde{U} = \{f\colon \|P(f)\| < C\}$. Since P is continuous, $\tilde{U}$ is open, and $\tilde{U}$ is a neighborhood of f_0. For all n we have $\|P(f)\|_n = \|P(f)\| \leqslant C \leqslant C(1 + \|f\|_n)$. Thus P is tame.

For the other case, let $P\colon U \subseteq R^n \to F$ be continuous. Given $f_0 \in U$, pick a compact neighborhood K of f_0. Then since P is continuous, $\|P(f)\|_n$ is continuous on K for all n, so we can find a constant C_n with

$$\|P(f)\|_n \leqslant C_n \leqslant C_n(1 + \|f\|_n)$$

on K. Thus P is tame.

If a map P satisfies a tame estimate for one pair of gradings on F and G, then it also satisfies a tame estimate for any tamely equivalent pair of gradings. If P satisfies a tame estimate of degree r and base b, then it also satisfies a tame estimate of degree $\tilde{r}$ and base $\tilde{b}$ when $\tilde{r} \geqslant r$ and $\tilde{b} \geqslant b$. Moreover, our definition of tame for general (possibly) nonlinear maps coincides with our previous definition when the map is linear.

2.1.5. THEOREM. *A map is a tame linear map if and only if it is linear and tame.*

PROOF. Suppose L is linear and tame. Then L satisfies a tame estimate

$$\|Lf\|_n \leqslant C(1 + \|f\|_{n+r})$$

for all $n \geqslant b$ and all f in some neighborhood of the origin. Increasing b if necessary, we may assume that the neighborhood is of the form $\{f\colon \| f \|_{b+r} \leqslant \varepsilon\}$ for some $\varepsilon > 0$. For any $g \neq 0$ let $f = \varepsilon g / \| g \|_b$. Then $\| f \|_b = \varepsilon$ so $\| Lf \|_n \leqslant C(1 + \| f \|_{n+r})$. Since L is linear we have

$$Lf = \varepsilon Lg / \| g \|_{b+r}$$

and if we multiply the estimate for f by $\| g \|_{b+r}/\varepsilon$ we get

$$\| Lg \|_n \leqslant C(\| g \|_{b+r}/\varepsilon + \| g \|_{n+r}).$$

Since $\| g \|_{b+r} \leqslant C \| g \|_{n+r}$ for $n \geqslant b$ and $1/\varepsilon \leqslant C$, we have

$$\| Lg \|_n \leqslant C \| g \|_{n+r}$$

when $n \geqslant b$. Thus L is a tame linear map.

2.1.6. THEOREM. *A composition of tame maps is tame.*

PROOF. Let F, G and H be graded spaces with open sets U, V and W. Let P and Q be two tame maps

$$U \subseteq F \overset{P}{\to} V \subseteq G \overset{G}{\to} W \subseteq H$$

such that the composition QP is defined. Given any $f_0 \in U \subseteq F$ we can find a neighborhood $\tilde{U}$ of f_0 on which P satisfies a tame estimate of some degree r and base b

$$\| P(f) \|_m \leqslant C(1 + \| f \|_{m+r}) \quad \text{for } m \geqslant b.$$

Likewise we can find a neighborhood $\tilde{V}$ of $g_0 = P(f_0)$ on which Q satisfies a tame estimate of some degree s and base c

$$\| Q(g) \|_n \leqslant C(1 + \| g \|_{n+s}) \quad \text{for } n \geqslant c.$$

By increasing b or c if necessary we may assume $b = c + r$. Since a tame map is assumed to be continuous, $P^{-1}(\tilde{V})$ is a neighborhood of f_0, so by shrinking $\tilde{U}$ if necessary we may assume $P(\tilde{U}) \subseteq \tilde{V}$. Then for all $f \in \tilde{U}$ and all $n \geqslant c$ we have

$$\| Q(P(f)) \|_n \leqslant C(1 + \| P(f) \|_{n+s}) \leqslant C(1 + \| f \|_{n+r+s})$$

which shows that the composition $Q \circ P$ satisfies a tame estimate of degree $r + s$ and base c in a neighborhood $\tilde{U}$ of f_0.

For a function $P(f, h) = g$ of two variables, we may wish to assign different degrees to f and h. We say that P satisfies a tame estimate of degree r in f and s in h and base b if

$$\| P(f, h) \|_n \leqslant C(\| f \|_{n+r} + \| h \|_{n+s} + 1)$$

for all $n \geqslant b$, with constants $C = C_n$ independent of f and h. On the product space P satisfies a joint tame estimate of degree $\max(r, s)$.

More generally, consider a map $L\colon F \times H \to G$ linear only in the second factor separately, which we write as $L(f)h = g$, grouping the parentheses to reflect the linearity.

2.1.7. LEMMA. *If $L(f)h$ is linear in h and satisfies a tame estimate of degree r in f and s in h and base b for all f in a $\| \ \|_{b+r}$ neighborhood of f_0 and all h in a $\| \ \|_{b+s}$ neighborhood of 0, then it satisfies the estimate*

$$\|L(f)h\|_n \leqslant C(\|f\|_{n+r}\|h\|_{b+s} + \|h\|_{n+s})$$

for all $n \geqslant b$, with constants $C = C_n$ independent of f and h, for all f in a $\| \ \|_{b+s}$ neighborhood of f_0 and all h without restriction.

PROOF. Suppose the given estimate holds for all $\tilde{h}$ in the zero-neighborhood $\|\tilde{h}\|_{b+s} \leqslant \varepsilon$. Given any h, pick $\tilde{h} = \lambda h$ with $\lambda = \varepsilon/\|h\|_{b+s}$, and apply the given estimate to $L(f)\tilde{h}$, so that

$$\|L(f)\tilde{h}\|_n \leqslant C\big(\|f\|_{n+r} + \|\tilde{h}\|_{n+s} + 1\big).$$

By linearity $L(f)\tilde{h} = \lambda L(f)h$. Clearly we have the estimate

$$\|L(f)h\|_n \leqslant \frac{C}{\varepsilon}(\|f\|_{n+r} + 1)\|h\|_{b+s} + C\|h\|_{n+s}.$$

Since ε is constant and $\|h\|_{b+s} \leqslant C\|h\|_{n+s}$, the lemma follows. In the same way we can prove the following result.

2.1.8. LEMMA. *Let $B(f)\{g, h\}$ be bilinear in g and h and satisfy a tame estimate of degree r in f, s in g, t in h and base b, for f in a $\| \ \|_{b+r}$ neighborhood of f_0, g in a $\| \ \|_{b+s}$ neighborhood of 0, and h in a $\| \ \|_{b+t}$ neighborhood of 0. Then B satisfies the estimate*

$$\|B(f)\{g, h\}\|_n \leqslant C(\|f\|_{n+r}\|g\|_{b+s}\|h\|_{b+t} + \|g\|_{n+s}\|h\|_{b+t} + \|g\|_{b+s}\|h\|_{n+t})$$

for all $n \geqslant b$, with constants $C = C_n$ independent of f, g, and h, for all f in a $\| \ \|_{b+r}$ neighborhood of f_0, and all g and h without restriction.

Suppose now that F and G are graded spaces and P: $U \subseteq F \to G$ is a nonlinear map. We say P is a *smooth tame* map if P is smooth and all its derivatives D^kP are tame. The category $\mathcal{T}$ is defined to be that whose objects are tame spaces and whose maps are smooth tame maps. We shall state and prove the Nash-Moser theorem in terms of this category.

2.2. *Partial differential operators.* We prove first a standard result in interpolation theory which will be useful in estimating differential operators.

2.2.1. THEOREM. *If X is a compact manifold and $l \leqslant m \leqslant n$ then for all $f \in \mathcal{C}^\infty(X)$*

$$\|f\|_m^{n-l} \leqslant C\|f\|_n^{m-l}\|f\|_l^{n-m}.$$

PROOF. It suffices to prove the result for $f \in \mathcal{C}_0^\infty(R^d)$ a smooth function on Euclidean space with compact support, since we can then use a partition of unity argument on X. Moreover it suffices to prove the case

$$\|f\|_1^2 \leqslant C\|f\|_2\|f\|_0$$

since we can derive the other cases by replacing f by a derivative $D^\alpha f$. Since the maximum of the first derivative occurs at a given point in a given direction, it

suffices to prove the result on the line. We claim then that for $f \in \mathcal{C}_0^\infty(R)$ we have

$$\sup |f'| \leqslant 2 \sup |f''| \sup |f| .$$

If we replace f by af and x by bx, we multiply f' by a/b and f'' by a/b^2. If $f \not\equiv 0$ we may thus assume $\sup |f'| = 1$ and $\sup |f''| = 1$. Moreover by translation we may assume $\sup |f'|$ occurs at the origin. Replacing f by $-f$ and x by $-x$ if necessary, we may assume $f(0) \geqslant 0$ and $f'(0) = 1$. By the second mean value theorem

$$f(1) = f(0) + f'(0) + \tfrac{1}{2} f''(\xi)$$

for some ξ in $[0, 1]$. Then $f''(\xi) \geqslant -1$ so

$$\sup |f| \geqslant f(1) \geqslant \tfrac{1}{2}.$$

This completes the proof.

The above estimate also holds for the Hölder norms $C^{n+\alpha}$, the Sobolev norms L_p^n, and the Besov norms $B_{p,q}^{n+\alpha}$.

2.2.2. COROLLARY. *If (i, j) lies on the line segment joining (k, l) and (m, n) then for any functions f and g in $\mathcal{C}^\infty(X)$*

$$\| f \|_i \| g \|_j \leqslant C(\| f \|_k \| g \|_l + \| f \|_m \| g \|_n).$$

PROOF. Let $(i, j) = t(k, l) + (1 - t)(m, n)$ with $0 \leqslant t \leqslant 1$. By the previous result

$$\| f \|_i \leqslant C \| f \|_k^t \| f \|_m^{1-t}, \qquad \| g \|_j \leqslant C \| g \|_l^t \| g \|_n^{1-t}.$$

Since $x^t y^{1-t} \leqslant C(x + y)$ for $x \geqslant 0$, $y \geqslant 0$, the result follows.

2.2.3. COROLLARY. *If f and g are functions in $\mathcal{C}^\infty(X)$ then*

$$\| fg \|_n \leqslant C(\| f \|_n \| g \|_0 + \| f \|_0 \| g \|_n).$$

PROOF. Again it suffices to prove the result on $\mathcal{C}_0^\infty(R^d)$ since we can patch together with a partition of unity. If $D^\alpha = (\partial/\partial x^1)^{\alpha_1} \cdots (\partial/\partial x^d)^{\alpha_d}$ denotes a partial derivative with multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ of length $|\alpha| = \alpha_1 + \cdots + \alpha_d$ then by the product rule

$$D^\alpha(fg) = \sum_{\beta+\gamma=\alpha} D^\beta f \, D^\gamma g.$$

2.2.4. EXAMPLE. Define P: $\mathcal{C}^\infty[a, b] \to \mathcal{C}^\infty[a, b]$ by $P(f) = e^f$. Any point has a neighborhood on which $\| f \|_0 \leqslant C$. Then on that neighborhood we also have $\| e^f \|_0 \leqslant C$. A typical derivative of e^f looks like

$$\left(\frac{d}{dx}\right)^2 e^f = e^f \left[\frac{d^2 f}{dx^2} + \left(\frac{df}{dx}\right)^2\right].$$

This gives us an estimate

$$\left\| \left(\frac{d}{dx}\right)^2 e^f \right\|_0 \leqslant \| e^f \|_0 \left\{ \left\| \frac{d^2 f}{dx^2} \right\|_0 + \left\| \frac{df}{dx} \right\|_0^2 \right\}.$$

Since $\|f\|_1^2 \leqslant C\|f\|_2\|f\|_0$ by interpolation, we see that when $\|f\|_0 \leqslant C$ we have $\|Pf\|_2 \leqslant C(1 + \|f\|_2)$. The same argument works for higher derivatives, and P is tame.

Let U be an open subset of $[a, b] \times R$ and let $p(x, y)$ be a smooth function on U. Let $\tilde{U}$ be the open subset of $\mathcal{C}^\infty[a, b]$ consisting of functions $y = f(x)$ whose graph lies in U.

2.2.5. THEOREM. *The operator* P: $\tilde{U} \subseteq \mathcal{C}^\infty[a, b] \to \mathcal{C}^\infty[a, b]$ *defined by*

$$Pf(x) = p(x, f(x))$$

is tame.

PROOF. We shall show that P satisfies a tame estimate of degree 0 and base 0 in a neighborhood of each function $f_0 \in \tilde{U} \subseteq \mathcal{C}^\infty[a, b]$. Given f_0, let N be a compact neighborhood of the graph of f_0 in $U \subseteq [a, b] \times R$. Then $p(x, y)$ and all its derivatives $D_x^j D_y^k p(x, y)$ are bounded on N. The set $\tilde{N} \subseteq \tilde{U} \subseteq \mathcal{C}^\infty[a, b]$ of all functions whose graphs lie in N is a neighborhood of f_0 in $\tilde{U} \subseteq \mathcal{C}^\infty[a, b]$. Using the chain rule we can write

$$\left(\frac{d}{dx}\right)^n p(x, f(x)) = \sum \frac{\partial^j}{\partial x^j}\frac{\partial^k}{\partial y^k} p(x, f(x)) \frac{d^{i_1} f}{dx^{i_1}} \cdots \frac{d^{i_k} f}{dx^{i_k}}$$

where the sum ranges over terms with $j + i_1 + \cdots + i_k = n$ (and some terms may occur more than once). For $f \in \tilde{N}$ we have $\|f\|_0 \leqslant C$ and

$$\left\| \frac{\partial^j}{\partial x^j}\frac{\partial^k}{\partial y^k} p(x, f(x)) \right\|_0 \leqslant C$$

so we get an estimate

$$\left\| \left(\frac{d}{dx}\right)^n p(x, f(x)) \right\|_0 \leqslant C \sum \|f\|_{i_1} \cdots \|f\|_{i_k}$$

where the sum ranges over terms with $i_1 + \cdots + i_k \leqslant n$. By interpolation we can bound $\|f\|_i \leqslant C\|f\|_n^{i/n}$ when $\|f\|_0 \leqslant C$. Therefore

$$\|P(f)\|_n \leqslant C(1 + \|f\|_n)$$

when $f \in \tilde{N}$.

Let X be a compact manifold and let V and W be vector bundles over X. Let U be an open subset of V and p: $U \subseteq V \to W$ a smooth map of U into W which takes fibres into fibres. If $\mathcal{C}^\infty(X, V)$ is the space of sections of the bundle V over X, then the set $\tilde{U}$ of sections whose image lies in U is an open subset. Define a map P: $\tilde{U} \subseteq \mathcal{C}^\infty(X, V) \to \mathcal{C}^\infty(X, W)$ by composition with p, so that $Pf(x) = p(f(x))$. We call p a *nonlinear vector bundle map* and P a *nonlinear vector bundle operator.*

2.2.6. THEOREM. *A nonlinear vector bundle operator P is tame.*

PROOF. If $f_0 \in \mathcal{C}^\infty(X, V)$ we claim that P satisfies a tame estimate of degree 0 and base 0 on a neighborhood of f_0. Indeed let N be a compact neighborhood of the image of f_0 in V, and let $\tilde{N}$ be the set of sections whose images lie in N.

Then $\tilde{N}$ is a neighborhood of f_0 in $\mathcal{C}^\infty(X, V)$, and we claim that $\|Pf\|_n \leq C(1 + \|f\|_n)$ for all $f \in \tilde{N}$. By using a partition of unity we can reduce the problem to proving such an estimate in local coordinates. If X has dimension k and V has fibre dimension l, then V has local vector bundle charts in R^{k+l}, and p is given locally by a function $p(x, y)$ on an open set in R^{k+l} with $x \in R^k$ and $y \in R^l$, while a section f is given locally by a function $y = f(x)$. Then we must prove an estimate

$$\|p(x, f(x))\|_n \leq C(1 + \|f(x)\|_n).$$

The proof is now exactly the same as before, except for having more dimensions.

Again let X be a compact manifold and let V and W be vector bundles over X. A *nonlinear partial differential operator* of degree r from V to W is a map P: $U \subseteq \mathcal{C}^\infty(X, V) \to \mathcal{C}^\infty(X, W)$ such that $Pf(x)$ is a smooth function of f and its partial derivatives of degree at most r at x in any local charts.

2.2.7. COROLLARY. *Any nonlinear partial differential operator* P: $\mathcal{C}^\infty(X, V) \to \mathcal{C}^\infty(X, W)$ *is tame.*

PROOF. Using jet bundles, we can write $P = QL$ where L is a linear differential operator of degree r and Q is a nonlinear vector bundle operator as before (involving no derivatives). Then $\|Lf\|_n \leq C\|f\|_{n+r}$ and $\|Qg\|_n \leq C(1 + \|g\|_n)$, so $\|Pf\|_n \leq C(1 + \|f\|_{n+r})$ for all f in a neighborhood of any point.

If $P(f)$ is a nonlinear differential operator of degree r in f, then its derivative $DP(f)h$ is also a differential operator of degree r in f and h. From this it follows that P is a smooth tame map, and all its derivatives also are tame with degree r and base b.

2.3. *Tame Fréchet manifolds.* We can define a tame manifold as one with coordinate charts in tame spaces whose coordinate transition functions are smooth tame maps.

2.3.1. THEOREM. *Let X be a compact manifold and B a fibre bundle over X. Then the space $\mathcal{C}^\infty(B)$ of smooth sections of B is a tame manifold.*

PROOF. Let f be a section of B. We can find a vector bundle and a diffeomorphism of a neighborhood of the zero section onto a neighborhood of the image of f in B which takes fibres into fibres. We take the collection of all the maps on sections induced by such diffeomorphisms as our atlas of coordinate charts. The coordinate transition functions are vector bundle maps, which we proved are smooth tame maps in the last section.

2.3.2. COROLLARY. *If X is a compact manifold and Y is another manifold (of finite dimension) then the space $\mathfrak{M}(X, Y)$ of smooth maps of X into Y is a tame manifold.*

PROOF. $\mathfrak{M}(X, Y)$ is the space of sections of the product bundle $X \times Y$ over X.

2.3.3. THEOREM. *Let X, Y, and Z be finite dimensional manifolds with X and Y compact. Then the composition map*

$$C\colon \mathfrak{M}(Y, Z) \times \mathfrak{M}(X, Y) \to \mathfrak{M}(X, Z)$$

is a smooth tame map.

PROOF. First we show C is tame. Let $f_0 \in \mathfrak{M}(Y, Z)$ and $g_0 \in \mathfrak{M}(X, Y)$ be two reference maps. We can cover X, Y, and Z by coordinate charts with ranges being closed balls in Euclidean spaces R^l, R^m, and R^n, such that each map takes a closed ball into the interior of another ball (by a positive amount). Then any maps f and g close to f_0 and g_0 in the $\| \ \|_0$ sense will take the same closed balls into the same closed balls. In order to estimate the maps in our coordinate charts on the Fréchet spaces $\mathfrak{M}(X, Y)$ (etc.), it suffices to estimate their local representatives in the balls in R^l (etc.). We may as well take them to be all unit balls. Let B^l be the unit ball in R^l (etc.). Then we have the following estimate.

2.3.4. LEMMA. *Let $f\colon B^m \to B^n$ and $g\colon B^l \to B^m$ be smooth maps. Fix a constant K and assume that $\|f\|_1 \leqslant K$ and $\|g\|_1 \leqslant K$. Then we can find constants C_n depending on K such that for $n \geqslant 1$*

$$\|f \circ g\|_n \leqslant C_n(\|f\|_n + \|g\|_n + 1).$$

PROOF. By a repeated application of the chain rule we can find constants $c_{k,j_1,\ldots,j_k}$ such that for $n \geqslant 1$

$$D^n(f \circ g)(x) = \sum_{k=1}^{n} \sum_{j_1 + \cdots + j_k = n} C_{k,j_1,\ldots,j_k} D^k f(g(x))\{D^{j_1}g(x), \ldots, D^{j_k}g(x)\}$$

with $j_1, \ldots, j_k \geqslant 1$. This gives us the estimate

$$\|D^n(f \circ g)\|_0 \leqslant C \sum_{k=1}^{n} \sum_{j_1 + \cdots + j_k = n} \|f\|_k \|g\|_{j_1} \cdots \|g\|_{j_k}.$$

Now by interpolation

$$\|f\|_k \leqslant C\|f\|_1^{(n-k)/(n-1)} \|f\|_n^{(k-1)/(n-1)},$$
$$\|g\|_j \leqslant C\|g\|_1^{(n-j)/(n-1)} \|g\|_n^{(j-1)/(n-1)}$$

and inserting this above we get

$$\|D^n(f \circ g)\| \leqslant C\|g\|_1^{k-1}(\|g\|_1\|f\|_n + \|f\|_1\|g\|_n).$$

When $\|f\|_1 \leqslant K$ and $\|g\|_1 \leqslant K$ and C may depend on K

$$\|D^n(f \circ g)\|_0 \leqslant C(\|f\|_n + \|g\|_n)$$

for $n \geqslant 1$. When $n = 0$, $\|f \circ g\|_0 \leqslant C$. Since $\|f \circ g\|_n = \sum_{k=0}^{n} \|D^k(f \circ g)\|$ the lemma follows.

Now the lemma shows that the composition map $C(f, g) = f \circ g$ satisfies a tame estimate of degree 0 and base 1 on neighborhoods of f_0 and g_0 open in $\| \ \|_1$ (so that $\|f\|_1$ and $\|g\|_1$ are bounded). Hence C is tame.

We saw earlier in 4.4.5 that C is a smooth map, and its higher tangents T^nC themselves are formed by compositions and derivatives, so the T^nC are also tame. Thus C is a smooth tame map of manifolds.

A smooth *tame Lie group* is a smooth tame Fréchet manifold $\mathcal{G}$ which has a group structure such that the multiplication map $C\colon \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ and the inverse map

$$V\colon \mathcal{G} \to \mathcal{G}$$

are smooth tame maps. The following gives an example.

2.3.5. THEOREM. *Let X be a compact manifold and let $\mathcal{D}(X)$ be the diffeomorphism group. Then $\mathcal{D}(X)$ is a tame Fréchet manifold. The composition maps*

$$C\colon \mathcal{D}(X) \times \mathcal{D}(X) \to \mathcal{D}(X)$$

and the inverse map

$$V\colon \mathcal{D}(X) \to \mathcal{D}(X)$$

introduced in Example 4.4.6 *are all smooth tame maps. Hence $\mathcal{D}(X)$ is a smooth tame Lie group.*

PROOF. $\mathcal{D}(X)$ is an open subset of $\mathcal{M}(X, X)$ and hence a smooth tame Fréchet manifold. The composition map C is a smooth tame map by the preceding example. We shall show V is a smooth tame map. Fix $f_0 \in \mathcal{D}(X)$. If f is near f_0, then $f_0^{-1}f = g$ is near the identity map 1, and since $f = f_0 g$ we have $f^{-1} = g^{-1}f_0^{-1} = (f_0^{-1}f)^{-1}f_0^{-1}$. Thus

$$V(f) = C\big(V(C(V(f_0), f)), V(f_0)\big).$$

If we can show V is a smooth tame map in a neighborhood of 1, then it will be also in a neighborhood of f_0.

Let B_r denote the closed ball of radius r in R^n. We can find coordinate charts on X with values in the ball B_3 such that the inverse images of the ball B_1 still cover X. If f is a diffeomorphism of X near the identity, the local representative of f in each coordinate chart will map B_2 into B_3 and cover B_1. Then f^{-1} will map B_1 into B_2. We prove the following estimate.

2.3.6. LEMMA. *Let f be a smooth map of B_2 into B_3 in R^n. If $\varepsilon > 0$ is sufficiently small and if $\| f(x) - x \|_1 \leqslant \varepsilon$ then f^{-1} is a smooth map of B_1 into B_2 and*

$$\| f^{-1} \|_n \leqslant C_n(\| f \|_n + 1)$$

for all $n \geqslant 1$.

PROOF. It follows from the inverse function theorem that f is 1-1 and locally has a smooth inverse. That the image of B_2 covers B_1 is a topological result from degree theory. Let $g = f^{-1}$ on B_1. We know that

$$Dg(x) = Df\big(f^{-1}(x)\big)^{-1}$$

and if $\varepsilon > 0$ is small we have an estimate $\|Dg\|_0 \leqslant C$. Since $f \circ g = 1$ we have $D^n(f \circ g) = 0$ for $n \geqslant 2$. By our previous formula

$$Df(g(x))D^n g(x) + \sum_{k=2}^{n} \sum_{j_1 + \cdots + j_k = n} C_{k,j_1,\ldots j_k} D^k f(g(x))\{D^{j_1}g(x),\ldots,D^{j_k}g(x)\} = 0$$

which we can solve for $D^n g(x)$ to get

$$D^n g(x) = -Dg(x) \sum_{k=2}^{n} \sum_{j_1 + \cdots + j_k = n} C_{k,j_1,\ldots j_k} D^k f(g(x))\{D^{j_1}g(x),\ldots,D^{j_k}g(x)\}.$$

Since $j_1,\ldots,j_k \geqslant 1$ and $k \geqslant 2$ we have $j_1,\ldots,j_k \leqslant n-1$ also. We have the estimate

$$\|D^n g\|_0 \leqslant C \sum_{k=2}^{n} \sum_{j_1 + \cdots + j_k = n} \|f\|_k \|g\|_{j_1} \cdots \|g\|_{j_k}.$$

By interpolation

$$\|f\|_k \leqslant C\|f\|_1^{(n-k)/(n-1)} \|f\|_n^{(k-1)/(n-1)},$$
$$\|g\|_j \leqslant C\|g\|_1^{(n-j-1)/(n-2)} \|g\|_{n-1}^{(j-1)/(n-2)}$$

and $\|f\|_1 \leqslant C$ and $\|g\|_1 \leqslant C$. This gives

$$\|D^n g\|_0 \leqslant C \sum_{k=2}^{n} \|f\|_n^{(k-1)/(n-1)} \|g\|_{n-1}^{(n-k)/(n-2)}.$$

We now proceed by induction on n. When $n = 1$ the theorem holds. Suppose we know that

$$\|g\|_{n-1} \leqslant C(\|f\|_{n-1} + 1).$$

By interpolation

$$(\|f\|_{n-1} + 1) \leqslant C(\|f\|_n + 1)^{(n-2)/(n-1)}$$

when $\|f\|_1 \leqslant C$. Inserting this above gives

$$\|D^n g\|_0 \leqslant C\|f\|_n.$$

Adding in $\|g\|_{n-1}$ proves the lemma. Since the higher tangents $T^n V$ also involve only derivatives, compositions and V, we see that $T^n V$ is also tame, and V is a smooth tame map.

2.3.7. COROLLARY. *The space $\mathcal{S}(X)$ of all compact smooth submanifolds of a finite dimensional manifold X is a smooth tame Fréchet manifold. So is the space $\mathcal{R}(X)$ of all compact regions in X with a smooth boundary.*

PROOF. We refer the reader to Example I.4.4.7, where we show that the coordinate transition functions are composed of compositions and inverses of the type already discussed.

We can also define a smooth tame vector bundle as a Fréchet vector bundle whose coordinate charts are smooth tame maps. Likewise we can define a smooth tame connection on such a vector bundle as one for which the local representatives Γ are smooth tame maps.

2.3.8. EXAMPLE. The bundle $\mathcal{C}^\infty\mathcal{R}(X)$ of all smooth functions on compact regions with smooth boundary in a finite dimensional manifold X is a smooth tame vector bundle. The intrinsic flat connection is a smooth tame connection.

PROOF. We refer the reader to the discussion in Examples I.4.3.6 and I.4.5.4. The previous estimates should make it clear that all the maps involved are smooth tame maps.

II.3. Inverses of families of linear maps.

3.1. *Smooth tame inverses.* The following result is very useful.

3.1.1. THEOREM. *Let* $L\colon (U \subseteq F) \times H \to K$ *be a smooth tame family of linear maps. Suppose that the equation* $L(f)h = k$ *has a unique solution* h *for all* f *and* k, *and the family of inverses* $V(f)k = h$ *is continuous and tame as a map*

Then V *is also a smooth tame map.*

$$V\colon (U \subseteq F) \times K \to H.$$

PROOF. By Theorem 5.3.1, V is smooth and

$$DV(f)\{k, g\} = -V(f)DL(f)\{V(f)k, g\}.$$

Since V and DL are tame, so is DV. That D^nV is also tame follows by induction on n.

3.2. *Ordinary differential equations.* It is an important observation that not only do differential equations provide us with examples of tame maps, but so do their solutions. This is particularly easy to verify for ordinary differential equations. We begin with the solution in Example I.5.3.5.

3.2.1. THEOREM. *Let* $C^\infty_{2\pi}$ *be the space of smooth functions periodic with period* 2π, *and let* U *be the open subset of functions* $f(t)$ *with*

$$\int_0^{2\pi} f(t)\,dt \neq 0.$$

Let $L(f)h = k$ *be the linear differential equation*

$$dh/dt + fh = k$$

and let $k = V(f)h$ *be its solution. Then* V *is a smooth tame map*

$$V\colon (U \subseteq \mathcal{C}^\infty_{2\pi}) \times \mathcal{C}^\infty_{2\pi} \to \mathcal{C}^\infty_{2\pi}.$$

PROOF. We already saw that the solution exists and is unique when $f \in U$ and that the solution V is smooth. We shall show that V is tame. We saw that

V is given by the following formulas:

$$F(t) = \int_0^t f(\theta)\,d\theta, \qquad c(f) = \int_0^{2\pi} f(\theta)\,d\theta,$$

$$C(f,h) = \frac{1}{e^{c(f)} - 1} \int_0^{2\pi} e^{F(\theta)} h(\theta)\,d\theta,$$

$$k(t) = e^{-F(t)} \left\{ \int_0^t e^{F(\theta)} h(\theta)\,d\theta + C(f,h) \right\}.$$

Now $\mathcal{C}^\infty_{2\pi}$ is a closed subspace of $\mathcal{C}^\infty[0,2\pi]$, but not a topological direct summand (see Example 1.2.2). Even if $f(t) \in \mathcal{C}^\infty_{2\pi}$ the integral $F(t)$ will not, unless $c(f) = 0$, which is exactly the case we must avoid. Therefore we regard the formulas above as giving a map

$$\tilde{V}\colon (\tilde{U} \subseteq \mathcal{C}^\infty[0,2\pi]) \times \mathcal{C}^\infty[0,2\pi] \to \mathcal{C}^\infty[0,2\pi]$$

where $\tilde{U}$ is all functions not necessarily periodic with integral not zero. Then V will take $\mathcal{C}^\infty_{2\pi}$ into itself, and $V = \tilde{V} \mid \mathcal{C}^\infty_{2\pi}$. We claim $\tilde{V}$ is tame. It then follows that V is also, for the same estimates will hold on the subspace.

Now all of the following maps are smooth tame maps:

$$\begin{aligned}
\mathcal{C}^\infty[0,2\pi] &\to \mathcal{C}^\infty[0,2\pi], & f(t) &\to F(t) = \int_0^t f(\theta)\,d\theta,\\
\mathcal{C}^\infty[0,2\pi] &\to R, & c(f) &= \int_0^{2\pi} f(\theta)\,d\theta,\\
\mathcal{C}^\infty[0,2\pi] &\to \mathcal{C}^\infty[0,2\pi], & F(t) &\to e^{F(t)},\\
\mathcal{C}^\infty[0,2\pi] &\to \mathcal{C}^\infty[0,2\pi], & F(t) &\to -F(t),\\
\mathcal{C}^\infty[0,2\pi] \times \mathcal{C}^\infty[0,2\pi] &\to \mathcal{C}^\infty[0,2\pi], & (f(t), g(t)) &\to f(t)g(t),\\
R - \{0\} &\to R, & x &\to (e^x - 1)^{-1},\\
R &\to \mathcal{C}^\infty[0,\pi], & c &\to f(t) \equiv c
\end{aligned}$$

and the solution $\tilde{V}$ is a rather complicated composition of these maps. Since a composition of smooth tame maps is a smooth tame map, $\tilde{V}$ and V are smooth tame maps.

We can also make tame estimates on the solution even when we lack an explicit formula, by the usual procedure for a priori estimates. As an illustration, we consider linear first order systems of ODE's. Let H be a finite dimensional vector space and let $h = h(t)$ be a smooth path in H. Let $L(H,H)$ be the space of linear maps of H into itself and let $f = f(t)$ be a smooth path in $L(H,H)$. We write $f \cdot h$ for the linear map applied to the vector. Then we can consider the linear first order system

$$dh/dt - f \cdot h = k$$

on $\alpha \leqslant t \leqslant \omega$ with initial conditions $h(\alpha) = \bar{h} \in H$. This defines a family of linear maps

$$L\colon \mathcal{C}^\infty([\alpha,\omega], L(H,H)) \times \mathcal{C}^\infty([\alpha,\omega], H) \to \mathcal{C}^\infty([\alpha,\omega], H) \times H$$

with $L(f)h = (k, \bar{h})$. For a given f, k and $\bar{h}$ there will exist a unique solution h by the usual existence theorem for ODE's, and we write $h = V(f)(k, \bar{h})$ for the solution, which defines a map

$$V\colon \mathcal{C}^\infty([\alpha, \omega], L(H, H)) \times \mathcal{C}^\infty([\alpha, \omega], H) \times H \to \mathcal{C}^\infty([\alpha, \omega], H).$$

The Fréchet space $\mathcal{C}^\infty([\alpha, \omega], H)$ is graded by the norms

$$\|h\|_n = \sup_{j \leq n} \sup_t |D^j h(t)|$$

where $|h|$ is any convenient norm on H. If $H = R^n$ we can take the Euclidean length. In this case $L(H, H)$ is the space of $n \times n$ matrices, and the system looks like

$$\frac{dh^i}{dt}(t) - \sum_{j=1}^{n} f_j^i(t) h^j(t) = k^i(t) \qquad (1 \leq i \leq n).$$

3.2.2. THEOREM. *The solution map V for the first order linear system is a smooth tame map. On any set $\|f\|_0 \leq K$ it satisfies estimates*

$$\|h\|_{n+1} \leq C_n\{\|k\|_n + \|f\|_n(\|k\|_0 + |\bar{h}|)\}$$

with constants C_n depending on K.

PROOF. We begin with the following estimate.

LEMMA.

$$\|h\|_0 \leq C(\|k\|_0 + |\bar{h}|).$$

PROOF. We introduce the norm

$$|[h]|_\lambda = \sup_t e^{-\lambda t} |h(t)|.$$

Since

$$h(t) = \int_\alpha^t [k(\theta) + f(\theta) \cdot h(\theta)]\, d\theta + \bar{h}$$

we have

$$e^{-\lambda t} h(t) = \int_\alpha^t e^{\lambda(\theta - t)} [e^{-\lambda\theta} k(\theta) + f(\theta) \cdot e^{-\lambda\theta} h(\theta)]\, d\theta + e^{-\lambda t} \bar{h}.$$

Since

$$\int_\alpha^t e^{\lambda(\theta - t)}\, d\theta \leq \frac{1}{\lambda}$$

we have the estimate for $\lambda > 0$

$$|[h]|_\lambda \leq \tfrac{1}{\lambda}\{|[k]|_\lambda + \|f\|_0 |[h]|_\lambda + |\bar{h}|\}.$$

If $\|f\|_0 \leq K$ and we choose $\lambda > K$, we can subtract the middle term on the right from the left to get the estimate

$$|[h]|_\lambda \leq C\{|[k]|_\lambda + |\bar{h}|\}$$

with a constant C depending on K and λ. Since $|[h]|_\lambda$ is equivalent to $\|h\|_0$, we have

$$\|h\|_0 \leqslant C\big(\|k\|_0 + |\bar{h}|\big).$$

Having a low norm estimate, we can use the equation to get estimates on the high norms. Since

$$dh/dt = f \cdot h + k$$

we get the estimate

$$\|dh/dt\|_n \leqslant C\big(\|f\|_n \|h\|_0 + \|f\|_0 \|h\|_n + \|k\|_n\big).$$

When $n = 0$ we have

$$\|h\|_1 = \|h\|_0 + \|dh/dt\|_0 \leqslant C\big(\|k\|_0 + |\bar{h}|\big).$$

We proceed by induction on n. Suppose

$$\|h\|_n \leqslant C\big\{\|k\|_{n-1} + \|f\|_{n-1}\big(\|k\|_0 + |\bar{h}|\big)\big\}.$$

Then remembering $\|f\|_0 \leqslant C$

$$\|h\|_{n+1} = \|h\|_n + \|dh/dt\|_n \leqslant C\big\{\|k\|_n + \|f\|_n\big(\|k\|_0 + |\bar{h}|\big)\big\}$$

which proves the estimate. Thus V is tame. It is a smooth tame map by Theorem 3.1.1, since L is a smooth tame map, being a differential operator.

We can apply this result to get a theorem for periodic solutions of periodic systems of ODE's, analogous to our previous result for a single equation. Let $\mathcal{C}_p^\infty(H)$ be the space of smooth functions $h(t)$ with values in the finite dimensional vector space H which are periodic with period p, so that $h(t + p) = h(t)$. Let $f(t) \in \mathcal{C}_p^\infty(L(H, H))$ and consider the first order linear system

$$df/dt - f \cdot h = k$$

as defining a smooth tame map

$$M\colon \mathcal{C}_p^\infty(L(H, H)) \times \mathcal{C}_p^\infty(H) \to \mathcal{C}_p^\infty(H)$$

by $M(f)h = k$. There may not always be a solution, as we see when $p = 2\pi$, $f = 0$, and $k = \sin t$. We let U be the set of all $f \in \mathcal{C}_p^\infty(L(H, H))$ for which the homogeneous equation

$$dh/dt - f \cdot h = 0$$

has only the trivial solution $h = 0$. If H has an inner product and $f(t)$ is symmetric and positive-definite or negative-definite with respect to this inner product for all t, then $f \in U$; for

$$\int_0^p \langle f \cdot h, h \rangle \, dt = \int_0^p \left\langle \frac{dh}{dt}, h \right\rangle dt = 0.$$

3.2.3. THEOREM. *The set U is open. For all $f \in U$ and k the equation $M(f)h = k$ has a unique solution $h = W(f)k$. The map*

$$W\colon \big(U \subseteq \mathcal{C}_p^\infty(L(H, H))\big) \times \mathcal{C}_p^\infty(H) \to \mathcal{C}_p^\infty(H)$$

is a smooth tame map.

PROOF. We can regard $\mathcal{C}_p^\infty(H)$ as a closed subspace of $\mathcal{C}^\infty([0, p], H)$. Given $f \in \mathcal{C}^\infty([0, p], L(H, H))$ and $\bar{h} \in H$ we can solve the homogeneous equation

$$dh/dt - f \circ h = 0$$

with $h(0) = \bar{h}$ and let $h(p) = \bar{\bar{h}}$. We write $\bar{\bar{h}} = N(f)h$. This defines a family of linear maps

$$N: \mathcal{C}^\infty([0, p], L(H, H)) \times H \to H$$

with $\bar{\bar{h}} = V(f)(0, \bar{h}) \mid \{t = p\}$. Since V is a smooth tame map by Theorem 3.2.2 and evaluation at p is a continuous linear map to a finite dimensional space, we see N is a smooth tame map. Suppose f is periodic. Then we get a nonzero solution of the homogeneous equation with coefficients f if we can find a nonzero $\bar{h} \in H$ with $N(f)\bar{h} = \bar{h}$. Thus $f \in U$ if and only if $I - N(f)$ is one-to-one; and hence also invertible, since H is finite dimensional. Since the set of invertible linear maps in $L(H, H)$ is open and $N(f)$ depends continuously on f, we see that U is open.

When $f \in U$, we can try to find a solution h of the periodic equation for a given periodic k by solving the initial value problem on $[0, p]$, and hope that a fortuitous choice of $\bar{h}$ will make h periodic. This clearly happens when $\bar{\bar{h}} = \bar{h}$. To be systematic, we can start with $\bar{h} = 0$ and let $h_1 = V(f)(k, 0)$ be the solution of the inhomogeneous equation

$$dh_1/dt - f \cdot h_1 = k$$

with $h_1(0) = \bar{h}_1 = 0$ at $t = 0$. We then let $\bar{\bar{h}}_1 = h_1(p)$. Clearly $\bar{\bar{h}}_1$ is a smooth tame function of f and k. Now we wish to change h_1 by a solution h_2 of the homogeneous equation so that $h = h_1 + h_2$ will be periodic. We need $\bar{h} = \bar{\bar{h}}$ or $\bar{h}_1 + \bar{h}_2 = \bar{\bar{h}}_1 + \bar{\bar{h}}_2$. Now $\bar{h}_1 = 0$ and $\bar{\bar{h}}_2 = N(f)\bar{h}_2$. Thus we must solve

$$\bar{h}_2 - N(f)\bar{h}_2 = \bar{\bar{h}}_1.$$

Since H is finite dimensional, it is easy to see that N defines a smooth tame map

$$N: U \subseteq \mathcal{C}_p^\infty(L(H, H)) \to L(H, H).$$

It suffices to choose a basis in H and represent $N(f)$ by a matrix, each element of which depends smoothly on f. It is tame because $L(H, H)$ is finite dimensional. Since taking the inverse of a matrix is a smooth map, we see that the map $f \to (I - N(f))^{-1}$ is also a smooth tame map. Then we choose

$$\bar{h}_2 = [I - N(f)]^{-1}\bar{\bar{h}}_1.$$

As a composition, $\bar{h}_2$ is a smooth tame function of f and k. Thus so is

$$h = h_1 + h_2 = V(f)(k, 0) + V(f)(0, \bar{h}_2).$$

Letting $h = \tilde{W}(f)k$ defines a smooth tame map which gives a commutative diagram

$$\begin{array}{ccc} \left(U \subseteq \mathcal{C}_p^\infty(L(H,H))\right) \times \mathcal{C}_p^\infty(H) & \overset{W}{\to} & \mathcal{C}_p^\infty(H) \\ \downarrow & & \downarrow \\ \left(\tilde{U} \subseteq \mathcal{C}^\infty([0,p], L(H,H))\right) \times \mathcal{C}^\infty([0,p], H) & \overset{\tilde{W}}{\to} & \mathcal{C}^\infty([0,p], H) \end{array}$$

on the set $\tilde{U}$ for which $I - N(f)$ is invertible. It is then clear that W is a smooth tame map also, since the estimates for $\tilde{W}$ will hold for W.

3.3. *Elliptic equations.* Let X be a compact manifold without boundary, and let V and W be vector bundles over X. A linear partial differential operator of degree r from V to W assigns to each section $h \in \mathcal{C}^\infty(X, V)$ a section $k \in \mathcal{C}^\infty(X, W)$ in a manner which depends linearly on the derivatives of degree r or less, so that locally

$$(*) \qquad \sum_{|\alpha| \leqslant r} f_\alpha \cdot D^\alpha h = k$$

where $D = (\partial/\partial x^1)^{\alpha_1} \cdots (\partial/\partial x^n)^{\alpha_n}$ is a partial derivative of degree $|\alpha| = \alpha_1 + \cdots + \alpha_n$. The coefficients $f = \{f_\alpha\}$ themselves form a section of a bundle $D^r(V, W)$ of all operators of degree r, as may be seen by considering what happens to the f_α under a coordinate change. We can identify $D^r(V, W)$ with the bundle $L(J^r V, W)$ where J^r is the r-jet bundle of V, or with $L(V, J_r W)$ where $J_r W$ is the r-cojet bundle of W. The formula $(*)$ defines a map $L(f)h = k$ where

$$L\colon \mathcal{C}^\infty(X, D^r(V, W)) \times \mathcal{C}^\infty(X, V) \to \mathcal{C}^\infty(X, W).$$

It is the generic partial differential operator of degree r with coefficients f. Of course L is a smooth tame family of linear maps.

To each linear differential operator $L(f)$ of degree r we associate a principal symbol $\sigma(f)$ which is a homogeneous polynomial of degree r on the cotangent bundle T^*X with values in $L(V, W)$, given in local coordinates by

$$\sigma(f)\zeta = \sum_{|\alpha| = r} f_\alpha \xi^\alpha$$

where $\xi^\alpha = \xi_1^{\alpha_1} \xi_2^{\alpha_2} \cdots \xi_n^{\alpha_n}$. The operator $L(f)$ is called *elliptic* if $\sigma(f)\xi \in L(V, W)$ is always invertible when $\xi \neq 0$. If $L(f)$ is elliptic, then it is well known that the null space of $L(f)$ is finite dimensional, and its range is closed with finite codimension. We let $U \subseteq \mathcal{C}^\infty(X, D^r(V, W))$ be the open set of all f for which the linear operator $L(f)$ with coefficients f is elliptic and invertible. Then for each $f \in U$ and each k there exists a unique h solving the equation $L(f)h = k$.

3.3.1. THEOREM. *The solution $S(f)k = h$ defines a smooth tame map*

$$S\colon \left(U \subseteq \mathcal{C}^\infty(X, D^r(V, W))\right) \times \mathcal{C}^\infty(X, W) \to \mathcal{C}^\infty(X, V).$$

PROOF. We need to exercise some care in the choice of the gradings, since Gårding's inequality fails for C^n.

To make the following argument work, we need only the following simple properties:

(1) The norm $\| \ \|_n$ is the norm $\| \ \|_0$ of the section and its derivatives of degree less than or equal to n.

(2) The interpolation inequalities hold.

(3) There is a multiplicative estimate for the f and h norms

$$\| fh \|_0 \leqslant C \| f \|_0 \| h \|_0 .$$

(4) Gårding's inequality holds for any elliptic operator L and the h-norm

$$\| h \|_r \leqslant C(\| Lh \|_0 + \| h \|_0).$$

There are many possible choices. We can take norms $\| \ \|_n$ in $C^{n+\gamma}$ measuring the nth derivative with a Hölder exponent $0 < \gamma < 1$. Or we can take $\| f \|_n$ to be the C^n norm and $\| h \|_n$ or $\| k \|_n$ to be the L_p^n norm measuring n derivatives in L_p for $1 < p < \infty$. All these gradings will be tamely equivalent.

LEMMA. *If $f_0 \in U$ then we can find an $\varepsilon > 0$ and a constant C such that if $\| f - f_0 \|_0 < \varepsilon$ then $f \in U$ and for all h*

$$\| h \|_r \leqslant C \| L(f)h \|_0 .$$

PROOF. If $f_0 \in U$ then $L(f_0)$ is invertible, so we do not need the slush term in Gårding's inequality, and

$$\| h \|_r \leqslant C \| L(f_0)h \|_0 .$$

Now $\| L(f)h \|_0 \leqslant C \| f \|_0 \| h \|_r$ and $L(f)h - L(f_0)h = L(f - f_0)h$, so

$$\| h \|_r \leqslant C \| L(f)h \|_r + C \| f - f_0 \|_0 \| h \|_r .$$

When $\| f - f_0 \|_0 < \varepsilon$ and $C\varepsilon < 1$, we can subtract the last term from the other side. This proves the lemma.

Now we proceed to estimate the higher derivatives in the usual way, by differentiating through the equation, paying careful attention to how the estimates depend on the coefficients. To avoid local coordinates, we have recourse to the following strategem. Pick a vector field v on X and pick first order linear differential operators ∇_v from $\mathcal{C}^\infty(X, V)$ and $\mathcal{C}^\infty(X, W)$ to themselves whose symbols are

$$\sigma \nabla_v(\xi) = \langle \xi, v \rangle I$$

where I is the identity matrix. Then the operator $\nabla_v L(f) - L(f)\nabla_v$ has its principal symbol zero, so it is itself of degree r. Hence there is a uniquely determined $\nabla_v f$ with

$$\nabla_v L(f) - L(f)\nabla_v = L(\nabla_v f)$$

and ∇_v is also a first order linear differential operator on $\mathcal{C}^\infty(X, D'(V, W))$ with symbol as above, as we can see from the formula (*). If we pick a finite number of vector fields v which span the tangent space at each point, then

$$\| h \|_{n+1} \leqslant C\left(\| h \|_n + \sum_v \| \nabla_v h \|_n \right).$$

We shall prove the following estimate.

3.3.2. LEMMA. *If* $\|f - f_0\| < \varepsilon$ *and* $L(f)h = k$ *then*

$$\|h\|_{n+r} \leqslant C(\|k\|_n + \|f\|_n \|k\|_0).$$

Thus $V(f)k = h$ *satisfies a tame estimate of degree* $-r$ *and base* r.

PROOF. This holds if $n = 0$. We assume it is true for some n and proceed by induction. First note that by the product rule and interpolation

$$\|L(f)h\|_n \leqslant C(\|h\|_{n+r} + \|f\|_{n+r}\|h\|_r)$$

which shows L is tame of degree r and base 0. The induction hypothesis says

$$\|h\|_{n+r} \leqslant C(\|L(f)h\|_n + \|f\|_n \|L(f)h\|_0).$$

Applying this to $\nabla_v h$ instead of h

$$\|\nabla_v h\|_{n+r} \leqslant C(\|L(f)\nabla_v h\|_n + \|f\|_n \|L(f)\nabla_v h\|_0).$$

Now we use the fact that

$$L(f)\nabla_v h = \nabla_v L(f)h - L(\nabla_v f)h.$$

Since $L(f)h = k$ we get

$$\|L(f)\nabla_v h\|_n \leqslant C(\|k\|_{n+1} + \|h\|_{n+r} + \|f\|_{n+1}\|h\|_r).$$

Using our estimates on $\|h\|_r$ and $\|h\|_{n+r}$ from the previous lemma and the induction hypothesis and simplifying

$$\|L(f)\nabla_v h\|_n \leqslant C(\|k\|_{n+1} + \|f\|_{n+1}\|k\|_0).$$

We can use this estimate and the corresponding one for $n = 0$ in our previous estimate for $\nabla_v h$ to get

$$\|\nabla_v h\|_{n+r} \leqslant C\{(\|k\|_{n+1} + \|f\|_{n+1}\|k\|_0) + \|f\|_n(\|k\|_1 + \|f\|_1\|k\|_0)\}.$$

Now by interpolation

$$\|f\|_n\|k\|_1 \leqslant C(\|f\|_{n+1}\|k\|_0 + \|f\|_0\|k\|_{n+1}),$$
$$\|f\|_n\|f\|_1 \leqslant C\|f\|_{n+1}\|f\|_0.$$

On our set $\|f\|_0 \leqslant C$ so

$$\|\nabla_v h\|_{n+r} \leqslant C(\|k\|_{n+1} + \|f\|_{n+1}\|k\|_0).$$

Summing over a finite number of v

$$\|h\|_{n+r+1} \leqslant C\left(\|h\|_{n+r} + \sum_v \|\nabla_v h\|_{n+r}\right)$$

which shows that

$$\|h\|_{n+r+1} \leqslant C(\|k\|_{n+1} + \|f\|_{n+1}\|k\|_0).$$

This completes the induction.

Even in the case where $L(f)$ is not invertible we can still define an inverse of sorts, up to something of finite dimension. Choose finite dimensional vector spaces N and M and continuous linear maps

$$j: \mathcal{C}^\infty(X, V) \to N, \qquad i: M \to \mathcal{C}^\infty(X, W).$$

We assume that in some norms $|\ |$ on M and N we have

$$|jh| \leqslant C\|h\|_{r-1}$$

while for all n we will surely have

$$|ix|_n \leqslant C|x|$$

for all $x \in M$. We then define a map

$$\tilde{L}\colon \mathcal{C}^\infty(X, D^r(V,W)) \times \mathcal{C}^\infty(X,V) \times M \to \mathcal{C}^\infty(X,W) \times N$$

by letting $\tilde{L}(f)(h,x) = (k,y)$ where $f \in \mathcal{C}^\infty(X, D^r(V,W))$, $h \in \mathcal{C}^\infty(X,V)$, $k \in \mathcal{C}^\infty(X,W)$, $x \in M$, $y \in N$, and

$$L(f)h + ix = k, \qquad jh = y.$$

We let $\tilde{U} \subseteq \mathcal{C}^\infty(X, D^r(V,W))$ be the open set (by Fredholm theory) of all coefficients f of elliptic operators $L(f)$ for which $\tilde{L}(f)$ is an isomorphism. We can do this in a neighborhood of any f_0. It follows that for all $f \in \tilde{U}$ and all k and y there exist unique h and x solving the equation.

3.3.3. THEOREM. *The solution* $\tilde{S}(f)(k,y) = (h,x)$ *defines a smooth tame family of linear maps*

$$\tilde{S}\colon \big(\tilde{U} \subseteq \mathcal{C}^\infty(X, D^r(V,W))\big) \times \mathcal{C}^\infty(X,W) \times N \to \mathcal{C}^\infty(X,V) \times M.$$

When $L(f)$ *is injective we can forget* N, *and when it is surjective we can forget* M. *Always put* $y = 0$ *and forget* x. *Then for each* $f \in \tilde{U}$ *and each* k *there exists a unique* h *in the null space of* j (*a subspace of finite codimension*) *such that* $L(f)h - k$ *lies in the image of* i (*a subspace of finite dimension*). *Moreover if we define the Green's operator* $h = G(f)k$ *as above, then* G *is a smooth tame family of linear maps.*

PROOF. The argument proceeds much as before, except for the annoying presence of i and j. For a given f_0, Gårding's inequality can be modified to

$$\|h\|_r + |x| \leqslant C\big(\|L(f_0)h + ix\|_0 + |jh|\big)$$

since j is injective on the null space and i is transversal to the image of $L(f)$. Then as before if $\|f - f_0\|_0 < \varepsilon$ we have

$$\|h\|_r + |x| \leqslant C\big(\|L(f)h + ix\|_0 + |jh|\big)$$

when $\varepsilon > 0$ is small enough to subtract off the extra error term on the right from the left. For the induction step we prove the following estimate.

3.3.4. LEMMA. *If* $\|f - f_0\|_0 < \varepsilon$ *and* $L(f)h + ix = k$ *and* $jh = y$ *then*

$$\|h\|_{n+r} + |x| \leqslant C(\|k\|_n + \|f\|_n\|k\|_0) + C(1 + \|f\|_n)|y|\ .$$

PROOF. The result holds for $n = 0$. We assume it is true for some n and proceed by induction. We can put $x = 0$ in the estimate. Then for all h

$$\|h\|_{n+r} \leqslant C(\|L(f)h\|_n + \|f\|_n\|L(f)h\|_0) + C(1 + \|f\|_n)|jh|\ .$$

Applying this to $\nabla_v h$ instead

$$\|\nabla_v h\|_{n+r} \leqslant C(\|L(f)\nabla_v h\|_n + \|f\|_n \|L(f)\nabla_v h\|_0) + C(1 + \|f\|_n)\,|j\nabla_v h|\,.$$

If $|jh| \leqslant C\|h\|_{r-1}$ then $|j\nabla_v h| \leqslant C\|h\|_r$. By the first estimate $\|h\|_r \leqslant C(\|k\|_0 + |y|)$. Hence this term will cause no trouble. For the others, we have

$$\nabla_v L(f)h = \nabla_v k - \nabla_v ix$$

and we always have

$$\|\nabla_v ix\|_n \leqslant C\|ix\|_{n+1} \leqslant C|x|$$

since i is continuous. But by the basic estimate $|x| \leqslant C(\|k\|_0 + |y|)$ also. Thus

$$\|\nabla_v L(f)h\|_n \leqslant C(\|k\|_{n+1} + |y|).$$

As before

$$L(f)\nabla_v h = \nabla_v L(f)h - L(\nabla_v f)h$$

and

$$\|L(\nabla_v f)h\|_n \leqslant C(\|h\|_{n+r} + \|f\|_{n+1}\|h\|_r).$$

By the induction hypothesis

$$\|L(f)\nabla_v h\|_n \leqslant C\{\|k\|_{n+1} + \|f\|_{n+1}\|k\|_0 + (1 + \|f\|_{n+1})\,|y|\}.$$

Using this estimate also for $n = 0$ and interpolating as before

$$\|\nabla_v h\|_{n+r} \leqslant C\{\|k\|_{n+1} + \|f\|_{n+1}\|k\|_0 + (1 + \|f\|_{n+1})\,|y|\}.$$

Summing over a finite number of v gives the desired estimate for $\|h\|_{n+r+1}$. As $|x|$ is already estimated, we are done.

We can also prove tame estimates for solutions of elliptic boundary value problems. We shall only discuss the simplest cases, but the results extend even to weighted systems (see Martin Lo [**11**]). Let X be a compact manifold with a smooth boundary ∂X, and let V and W be vector bundles over X and Z a vector bundle over ∂X. We shall consider a single equation $L(f)h = k$ of degree r taking a section h of V to a section k of W, with coefficients f forming a section of the bundle $D^r(V, W)$. For boundary conditions we will also have a single equation $B(g)h = l$ of degree $s \leqslant r$ taking the section h of V into a section l of W with coefficients g in a bundle $D^s(V, Z)$, so that locally

$$B(g)h = \sum_{|\beta| \leqslant s} g_\beta D^\beta h \,|\, \partial X$$

and we put $M(f, g)h = (k, l)$. We let U be the open set of all $f \in \mathcal{C}^\infty(X, D^r(V, W))$ and $g \in \mathcal{C}^\infty(X, D^s(V, Z))$ such that $L(f)$ is elliptic, the boundary condition $B(g)$ is coercive, and the boundary value problem $L(f)h = k$ on X and $B(g)h = l$ on ∂X has a unique solution $h \in \mathcal{C}^\infty(X, V)$ for all $k \in \mathcal{C}^\infty(X, W)$ and $l \in \mathcal{C}^\infty(\partial X, Z)$, i.e., $M(f, g)$ is invertible. There are well-known algebraic conditions for the boundary conditions to be coercive, but they are not so easy to state. They include the following special cases. We say ν

is a normal cotangent vector if $\nu \perp T\partial X$, and we say $\nu > 0$ if ν points inward. (With local coordinates $x^1, \ldots, x^{d-1}$ and $x^d \geqslant 0$, then $\nu = \nu(x)dx^n$ is normal and points inward if $\nu(x) > 0$.)

Case 1. Let V, W and Z be trivial bundles so that h, k and l are functions, and let $r = 2$. Then in local coordinates

$$L(f)h = \sum_{ij} f^{ij} \frac{\partial^2 h}{\partial x^i \partial x^j} + f^i \frac{\partial h}{\partial x^i} + f^0 h.$$

We say L is strongly elliptic if $\sigma L(f)\xi > 0$ for all $\xi \neq 0$, which happens if f^{ij} is positive definite. Let 1 be the constant function. Then the coefficient $f^0 = L(f)1$. If $L(f)1 < 0$ and $L(f)h = 0$ then the maximum of h occurs on the boundary.

Case 1a. Suppose $s = 0$. Then g is a function on ∂X and $B(g)h = gh$. If $g > 0$ then this is a Dirichlet boundary condition, which is known to be coercive. Moreover it has index zero, so the operator $M(f, g)$ is invertible if and only if the only solutions of the homogeneous equations $L(f)h = 0$ and $B(g)h = 0$ are $h = 0$. If $L(f)1 < 0$ this happens by the maximum principle.

Case 1b. Let $s = 1$. Then in local coordinates

$$B(g)h = \sum_i g^i \frac{\partial h}{\partial x^i} + g^0 h.$$

Note that $g = (\Sigma_i g^i \partial/\partial x^i, g^0)$ is given by a vector field on ∂X pointing along X, and a function on ∂X. The vector field points inward if and only if $\sigma B(g)\nu > 0$ for any normal $\nu > 0$. In this we have Neumann boundary conditions, which are known to be coercive. If $L(f)1 < 0$ and $B(g)1 < 0$ then the only solution of the homogeneous equations is trivial by the maximum principle, and $M(f, g)$ is invertible since the index is also zero.

Case 2. Let V, W and Z be general bundles, and let $r = 1$ and $s = 0$. Then in local coordinates

$$L(f)h^\beta = \sum_{\alpha, i} f_\alpha^{\beta i} \frac{\partial h^\alpha}{\partial x^i} + \sum_\alpha f_\alpha^\beta h^\alpha$$

and

$$B(g)h^\gamma = \sum_\alpha g_\alpha^\gamma h^\alpha.$$

Since $B(g)$ has degree zero, it is a vector bundle map of $V \mid \partial X$ into Z. We shall require that $B(g)$ always be surjective. Its null space Null $B(g)$ is a vector subbundle of $V \mid \partial X$. The homogeneous boundary condition $B(g)h = 0$ says $h \mid \partial X \in$ Null $B(g)$. We distinguish between real and complex equations.

Case 2a. For real equations with real coefficients, any boundary conditions are coercive provided

$$\dim Z = \tfrac{1}{2} \dim V.$$

Case 2b. For complex equations with complex coefficients the condition is more complicated. Let ξ be any real cotangent vector at a point in the boundary which is not normal to the boundary, and let ν be any cotangent

vector at the same point which is normal to the boundary at the same point and points inward, so $\nu > 0$. If $L(f)$ is elliptic then $\sigma L(f)\nu$ is invertible, and the linear map

$$[\sigma L(f)\nu]^{-1}[\sigma L(f)\xi]$$

is an automorphism of the fibre of V at that point which has no real eigenvalues. For $\sigma L(f)\xi$ is linear in ξ since $r = 1$, and if

$$[\sigma L(f)\nu]^{-1}[\sigma L(f)\xi]v = \lambda v$$

for some $v \in V$ then

$$\sigma L(f)(\xi - \lambda\nu)v = 0.$$

If λ is real then $\xi - \lambda\nu$ is real, and $\xi \neq \lambda\nu$ since ν is normal and ξ is not. Since $L(f)$ is elliptic we have $v = 0$.

We can decompose the fibre of V at the point as a direct sum

$$V^+(f, \xi, \nu) \oplus V^-(f, \xi, \nu)$$

of eigenspaces whose eigenvalues have imaginary part positive or negative. Then $B(g)$ is a coercive boundary condition for the elliptic operator $L(f)$ if and only if Null $B(g)$ is transversal to $V^-(f, \xi, \nu)$ for all such ξ and ν. Using the homogeneity in ξ and ν, it suffices to check the condition for $|\xi| = 1$, $|\nu| = 1$, and $\xi \perp \nu$ in some inner product on V. Hence the set of pairs (f, g) such that $L(f)$ is elliptic and $B(g)$ is coercive is open. When $\dim X \geqslant 3$ we can continuously rotate ξ into $-\xi$ without passing through the normal covectors. Since $\sigma L(f)\xi$ is linear in ξ

$$V^+(f, -\xi, \nu) = V^-(f, \xi, \nu).$$

Hence $\dim V^+ = \dim V^- = \dim Z$ when $\dim X \geqslant 3$.

For an elliptic equation $L(f)$ with a coercive boundary condition $B(g)$, the map $M(f, g)$ always has finite dimensional null space and closed range of finite codimension, and its index

$$I(f, g) = \dim \text{Null } M(f, g) - \text{codim Im } M(f, g)$$

is locally constant. Hence one way to evaluate the index is to deform the given problem into one for which it can be computed. If the index is zero and the only solution of the homogeneous equation is trivial, then $M(f, g)$ is invertible.

We now give the following result.

3.3.5. THEOREM. *Let U be the open set of all pairs (f, g) in*

$$\mathcal{C}^\infty(X, D^r(V, W)) \times \mathcal{C}^\infty(\partial M, D^s(V, Z))$$

such that $L(f)$ is an elliptic linear operator of degree r and $B(g)$ is a linear coercive boundary condition of degree s, and the boundary value problem

$$L(f)h = k \quad \text{on } X, \qquad B(g)h = l \quad \text{on } \partial X$$

has a unique solution $h \in \mathcal{C}^\infty(X, V)$ *for all* $k \in \mathcal{C}^\infty(X, W)$ *and* $l \in \mathcal{C}^\infty(\partial X, Z)$. *Then the solution* $h = S(f, g)(k, l)$ *defines a smooth tame family of linear maps*

$$S\colon (U \subseteq \mathcal{C}^\infty(X, D^r(V, W))) \times \mathcal{C}^\infty(\partial X, D^s(V, Z)) \times \mathcal{C}^\infty(X, W) \times \mathcal{C}^\infty(\partial X, Z) \to \mathcal{C}^\infty(X, V).$$

PROOF. Again we must choose gradings on our spaces with the properties listed before, and we also need the condition that for $l \in \mathcal{C}^\infty(\partial X, Z)$ and $\tilde{l} \in \mathcal{C}^\infty(X, Z)$

$$\|l\|_0 = \inf\{\|\tilde{l}\|_0 \colon l \mid \partial X = l\}.$$

This condition works for the grading $C^{n+\gamma}$ but not for L^n_p. Another possibility that works is to take $\|\ \|_n$ as the norm in the Besov space $B^{(d/p)+n}_{p,1}$ where d is the dimension (with $\dim \partial X = \dim X - 1$). Then we have $C^\gamma \subseteq B^{d/p}_{p,1} \subseteq C^0$ for $\gamma > d/p$.

Choose $f_0 \in U$. Since the equation $L(f_0)$ is elliptic, the boundary condition $B(g_0)$ is coercive, and the solutions are unique, we have a strong Gårding's inequality

$$\|h\|_r \leqslant C(\|L(f_0)h\|_0 + \|B(g_0)h\|_{r-s}).$$

As before, if $\|f - f_0\|_0 < \varepsilon$ and $\|g - g_0\|_{r-s} < \delta$ we have

$$\|h\|_r \leqslant C(\|L(f)h\|_0 + \|B(g)h\|_{r-s}) + C(\|f - f_0\|_0 + \|g - g_0\|_{r-s})\|h\|_r.$$

When $\varepsilon > 0$ and $\delta > 0$ are small enough then

$$\|h\|_r \leqslant C(\|L(f)h\|_0 + \|B(g)h\|_{r-s}).$$

3.3.6. LEMMA. *If* $\|f - f_0\|_0 < \varepsilon$, $\|g - g_0\|_{r-s} < \delta$, *and* $L(f)h = k$ *and* $B(g)h = l$ *then for all* $n \geqslant r$

$$\|h\|_n \leqslant C(\|k\|_{n-r} + \|l\|_{n-s}) + C(\|f\|_{n-r} + \|g\|_{n-s})(\|k\|_0 + \|l\|_{r-s}).$$

PROOF. This holds for $n = r$. We assume it is true for some n and proceed by induction. Let v be a vector field on X which is tangent to the boundary, so that $v \mid \partial X$ is also a vector field on ∂X. We choose first order linear differential operators ∇_v on $\mathcal{C}^\infty(X, V)$, $\mathcal{C}^\infty(X, W)$ and $\mathcal{C}^\infty(\partial X, Z)$ whose symbols are $\sigma\nabla_v(\xi) = \langle \xi, v \rangle$. Then we also get first order linear operators on $\mathcal{C}^\infty(X, D^r(V, W))$ and $\mathcal{C}^\infty(\partial X, D^s(V, Z))$ so that

$$\nabla_v L(f) = L(f)\nabla_v + L(\nabla_v f), \qquad \nabla_v B(g) = B(g)\nabla_v + B(\nabla_v g).$$

By the induction hypothesis, for all h

$$\|h\|_n \leqslant C(\|L(f)h\|_{n-r} + \|B(g)h\|_{n-s}) + C(\|f\|_{n-r} + \|g\|_{n-s})(\|L(f)h\|_0 + \|B(f)g\|_{r-s}).$$

Applying this to $\nabla_v h$ instead

$$\|\nabla_v h\|_n \leqslant C(\|L(f)\nabla_v h\|_{n-r} + \|B(g)\nabla_v h\|_{n-s}) + C(\|f\|_{n-r} + \|g\|_{n-s})(\|L(f)\nabla_v h\|_0 + \|B(g)\nabla_v h\|_{r-s}).$$

Now we have

$$\|L(f)\nabla_v h\|_{n-r} \leqslant \|\nabla_v k\|_{n-r} + \|L(\nabla_v f)h\|_{n-r},$$
$$\|B(g)\nabla_v h\|_{n-s} \leqslant \|\nabla_v g\|_{n-s} + \|B(\nabla_v g)h\|_{n-s}$$

and

$$\|L(\nabla_v f)h\|_{n-r} \leqslant C(\|h\|_n + \|f\|_{n-r+1}\|h\|_r),$$
$$\|B(\nabla_v g)h\|_{n-s} \leqslant C(\|h\|_n + \|g\|_{n-s+1}\|h\|_r)$$

and from the induction hypothesis

$$\|h\|_n \leqslant C(\|k\|_{n-r} + \|l\|_{n-s}) + C(\|f\|_{n-r} + \|g\|_{n-s})(\|k\|_0 + \|l\|_{r-s}),$$
$$\|h\|_r \leqslant C(\|k\|_0 + \|l\|_{r-s}).$$

Combining these estimates and interpolating where necessary

$$\|L(f)\nabla_v h\|_{n-r} + \|B(g)\nabla_v h\|_{n-s} \leqslant C(\|k\|_{n-r+1} + \|l\|_{n-s+1}) + C(\|f\|_{n-r+1} + \|g\|_{n-s+1})(\|k\|_0 + \|l\|_{r-s}).$$

Then using this estimate for n and for 0

$$\|\nabla_v h\|_n \leqslant C(\|k\|_{n-r+1} + \|l\|_{n-s+1}) + C(\|f\|_{n-r+1} + \|g\|_{n-s+1})(\|k\|_0 + \|l\|_{r-s}).$$

This estimates all the tangential derivatives, but we must still estimate the normal derivative at the boundary from the equation $L(f)h = k$.

Choose local coordinates near the boundary $x^1, \ldots, x^{d-1}$ and $x^d \geqslant 0$. Let K be a compact neighborhood of a given point and let $\|h\|_{n,K}$ denote the n-norm over K. We can cover the boundary with a finite number of patches of this sort. Then for $n \geqslant r$ we will have

$$\|h\|_{n+1} \leqslant c\left(\sum_v \|\nabla_v h\|_n + \sum_K \sum_\alpha \left\| \frac{\partial^r h^\alpha}{(\partial x^n)^r} \right\|_{n-r+1,K}\right)$$

summing over a finite number of vector fields v tangent to the boundary and a finite number of patches K at the boundary. On the patch K we can write

$$L(f)h = L^\sharp(f)h + \sum_{j=1}^{d-1} L_j(f)\frac{\partial h}{\partial x^j} + L_0(f)h$$

where $L_j(f)$ and $L_0(f)$ are linear differential operators of degree $r-1$ whose coefficients are just some of the coefficients of $L(f)$, and where

$$L^\sharp(f)h^\beta = \sum_\alpha f_\alpha^\beta \frac{\partial^r h^\alpha}{(\partial x^d)^r}$$

and $f_\alpha^\beta = \sigma L(f)(dx^d)$ is the symbol of $L(f)$ in the normal direction. Then f_α^β is always invertible, and if $\|f - f_0\|_0 < \varepsilon$ with ε sufficiently small we will have the estimate

$$\|(f_\alpha^\beta)^{-1}\|_{n,k} \leqslant C(1 + \|f_\alpha^\beta\|_{m,k}).$$

We can solve for

$$\frac{\partial^r h^\alpha}{(\partial x^d)^r} = \sum_\beta \left(f_\alpha^\beta\right)^{-1} L^\sharp(f)h^\beta$$

and we see that

$$\left\| \frac{\partial^r h^\alpha}{(\partial x^d)^r} \right\|_{m,K} \leqslant C\left(\|L^\sharp(f)h\|_{m,K} + \|f\|_{m,K}\|L^\sharp(f)h\|_{0,K}\right).$$

Now $L^\sharp(f)h = k - \sum_{j=1}^{d-1} L_j(f)\partial h/\partial x^j - L_0(f)h$ and

$$\left\| L_j(f)\frac{\partial h}{\partial x^j} \right\|_{m,K} \leqslant C\left(\left\| \frac{\partial h}{\partial x^j} \right\|_{m+r-1,K} + \|f\|_{n,K} \left\| \frac{\partial h}{\partial x^j} \right\|_{r-1,K} \right),$$

$$\|L_0(f)h\|_{m,K} \leqslant C(\|h\|_{m+r-1,K} + \|f\|_{m,K}\|h\|_{r-1,K}).$$

We can extend $\partial/\partial x^j$ outside of K to be a smooth vector field tangent to the boundary for $1 \leqslant j \leqslant d-1$. This gives us the estimate

$$\|L^\sharp(f)h\|_m \leqslant C\left(\|k\|_m + \sum_v \|\nabla_v h\|_{m+r-1} + \|f\|_m\|h\|_r\right)$$

summing over a finite number of v tangent to the boundary. Since $\|L^\sharp(f)h\|_{0,k} \leqslant C\|h\|_r$ we have

$$\left\| \frac{\partial^r h}{(\partial x^d)^r} \right\|_{m,K} \leqslant C\left(\|k\|_m + \sum_v \|\nabla_v h\|_{m+r-1} + \|f\|_m\|h\|_r\right).$$

We now let $m = n - r + 1$ and get

$$\left\| \frac{\partial^r h}{(\partial x^d)^r} \right\|_{n-r+1,K} \leqslant C\left(\|k\|_{n-r+1} + \sum_v \|\nabla_v h\|_n + \|f\|_{n-r+1}\|h\|_r\right).$$

Combined with our previous estimate on $\|\nabla_v h\|_n$ we get

$$\|h\|_{n+1} \leqslant C(\|k\|_{n-r+1} + \|l\|_{n-s+1}) + C(\|f\|_{n-r+1} + \|g\|_{n-s+1})(\|k\|_0 + \|l\|_{r-s}).$$

3.4. *Symmetric systems.* Let X be a compact manifold with a smooth boundary ∂X. Let V be a vector bundle over X, and let $V^* = L(V, R)$ be the dual bundle. We write $\langle\ ,\ \rangle$ for the pairing of V and V^*. Let $L(f)h = k$ be a linear differential operator of degree 1 mapping sections h in V to sections k in V^* with coefficients f in the bundle $D^1(V, V^*)$ of all such operators. We say $L(f)$ is a symmetric operator if its symbol $\sigma L(f)(\xi)$ is symmetric, so that for all h_1 and h_2 in V

$$\langle \sigma L(\xi)h_1, h_2 \rangle = \langle \sigma L(\xi)h_2, h_1 \rangle.$$

We say that the boundary ∂X is nowhere characteristic if $\sigma L(\nu)h \neq 0$ when $h \neq 0$ and $\nu \neq 0$ is a normal covector at the boundary. In this case the boundary ∂X divides into two components $\partial^+ X$ and $\partial^- X$, such that for an inward pointing normal covector $\nu > 0$ we have $\sigma L(\nu)$ positive-definite on $\partial^+ X$ and negative-definite on $\partial^- X$.

We wish to solve the symmetric system $L(f)h = k$ with given boundary values $h = h^+$ on $\partial^+ X$ and no condition on $\partial^- X$. This will be possible provided there exists a function t on X such that $\sigma L(f)(dt)$ is always positive-definite for the covector $dt \in T^*X$. This result is due to Friedrichs. We call t a positive weight function.

LEMMA. *Let $L(f)$ be a symmetric system nowhere characteristic at the boundary. If there exists a positive weight function, then both $\partial^+ X$ and $\partial^- X$ are not empty.*

PROOF. Let $\check{x}$ and $\hat{x}$ be the points where t has its minimum and maximum. If $\sigma L(f)(dt)$ is positive-definite, surely $dt \neq 0$. Hence $\check{x}$ and $\hat{x}$ lie in ∂X. At $\check{x}$ the covector dt is an inward pointing normal covector, and at $\hat{x}$ it is an outward pointing normal covector. Then $\check{x} \in \partial^+ X$ and $\hat{x} \in \partial^- X$.

It is also possible to consider the case where X is a manifold with corners. At the corner we allow X to be modeled on $R^{m-2} \times \{R \geqslant 0\} \times \{R \geqslant 0\}$. We must assume that at each corner there are two boundary components intersecting, one of which belongs to $\partial^+ X$ and the other to $\partial^- X$. The following argument goes through unchanged, because in our estimates we take derivatives parallel to $\partial^+ X$ only, and they need not be parallel to $\partial^- X$ also. This is important for finding local solutions of the equation, or for showing that compact support is maintained in a certain set.

Let $D_S^1(V, V^*)$ denote the subbundle of the vector bundle $D^1(V, V^*)$ of first order operators which are symmetric. Let $h \in \mathcal{C}^\infty(X, V)$ and $k \in \mathcal{C}^\infty(X, V^*)$. In local coordinates $h = \{h^\alpha\}$ and $k = \{k_\alpha\}$, and the pairing is given by

$$\langle h, k \rangle = \sum_\alpha h^\alpha k_\alpha.$$

Then locally $f = \{f^i_{\alpha\beta}, f_{\alpha\beta}\}$ and the operator is given by

$$L(f)h_\beta = \sum_{\alpha,i} f^i_{\alpha\beta} \frac{\partial h^\alpha}{\partial x^i} + \sum_\alpha f_{\alpha\beta} h^\alpha = k_\beta.$$

We see that $L(f)$ is symmetric if and only if $f^i_{\alpha\beta} = f^i_{\beta\alpha}$. The zero order terms of $f_{\alpha\beta}$ need not be symmetric. We let U denote the open subset of $\mathcal{C}^\infty(X, D_S^1(V, V^*))$ of coefficients f of symmetric operators which are nowhere characteristic at the boundary and have at least one positive weight function.

3.4.1. THEOREM. *For each $f \in U \subseteq \mathcal{C}^\infty(X, D_S^1(V, V^*))$, each $k \in \mathcal{C}^\infty(X, V^*)$ and each $h^+ \in \mathcal{C}^\infty(\partial^+ X, V)$ there exists a unique $h \in \mathcal{C}^\infty(X, V)$ with $L(f)h = k$ and $h \mid \partial^+ X = h^+$. The solution $h = S(f)(k, h^+)$ defines a smooth tame family of linear maps*

$$S: \big(U \subseteq \mathcal{C}^\infty\big(X, D_S^1(V, V^*)\big)\big) \times \mathcal{C}^\infty(X, V^*) \times \mathcal{C}^\infty(\partial^+ X, V) \to \mathcal{C}^\infty(X, V).$$

PROOF. We shall show how to derive the a priori estimates that prove S is tame. If we replaced functions by distributions and derivatives by differences, we could also prove the necessary regularity for the solution, but we leave this to the reader.

To begin, we choose an inner product $\langle\ ,\ \rangle$ on the bundle V giving an isomorphism between V and V^*. We also choose a smooth positive measure on X given locally by $d\mu = \mu(x)dx$ where $\mu(x)$ is a positive function and dx is Lebesgue measure. Finally we choose a smooth normal cotangent vector field ν on ∂X pointing inward. Then there is a uniquely determined smooth positive measure $d\tilde{\mu}$ on ∂X such that $d\mu = \nu d\tilde{\mu}$. In local coordinates if $x = (x^1,\dots,x^m)$ with $x^m \geqslant 0$ and $\tilde{x} = (x^1,\dots,x^{m-1})$ then $d\mu = \mu(x^1,\dots,x^m)dx^1\dots dx^m$ and $d\tilde{\mu} = \tilde{\mu}(x^1,\dots,x^{m-1})dx^1\dots dx^{m-1}$ and $\nu = \nu(x^1,\dots,x^{m-1})dx^m$. Then we have

$$\mu(x^1,\dots,x^{m-1},0) = \nu(x^1,\dots,x^{m-1})\tilde{\mu}(x^1,\dots,x^{m-1}).$$

We define inner products on $\mathcal{C}^\infty(X,V)$, $\mathcal{C}^\infty(\partial X,V)$ and $\mathcal{C}^\infty(\partial^+ X,V)$ by letting

$$\langle\langle h,k\rangle\rangle = \int_X \langle h,k\rangle\, d\mu,$$

$$\langle\langle h,k\rangle\rangle_\partial = \int_{\partial X}\langle h,k\rangle\, d\tilde{\mu},\quad \langle\langle h,k\rangle\rangle_+ = \int_{\partial^+ X}\langle h,k\rangle\, d\tilde{\mu}$$

and we let

$$\|h\|_0^2 = \langle\langle h,h\rangle\rangle,\quad \|h\|_{0,\partial}^2 = \langle\langle h,h\rangle\rangle_\partial,\quad \|h\|_{0,+}^2 = \langle\langle h,h\rangle\rangle_+ .$$

We shall measure $h \in \mathcal{C}^\infty(X,V)$ in a norm $\|h\|_n$ which measures the size of the first n derivatives in the L_2 norm $\|h\|_0$. Similarly for k, and for $h^+ = h \mid \partial^+ X$ we have $\|h^+\|_{n,+}$ analogously on $\partial^+ X$. But we shall always measure $f \in \mathcal{C}^\infty(X, D_1'(V,V^*))$ in a norm $\|f\|_n$ which measures the size of the first n derivatives in the supremum norm $\|f\|_0 = \sup_x |f(x)|$. It will be clear from the context which is meant. Both gradings satisfy the interpolation inequalities, and $\|fh\|_0 \leqslant C\|f\|_0\|h\|_0$.

We define the divergence of a section $f = \{f_{\alpha\beta}^i, f_{\alpha\beta}\}$ of the bundle $D_S^1(V,V^*)$ to be the section $\operatorname{div}(f) = \tilde{f} = \{\tilde{f}_{\alpha\beta}\}$ of the bundle $L(V,V^*)$ given locally by

$$\tilde{f}_{\alpha\beta} = \sum_i \frac{\partial}{\partial x^i} f_{\alpha\beta}^i + \sum_i f_{\alpha\beta}^i \frac{\partial}{\partial x^i}\log\mu(x) - 2f_{\alpha\beta}.$$

Thus div is itself a linear first order operator. We get the following formula.

3.4.2. LEMMA. *For all* $h \in \mathcal{C}^\infty(X,V)$

$$2\langle\langle L(f)h,h\rangle\rangle = \langle\langle \operatorname{div}(f)h,h\rangle\rangle - \langle\langle \sigma L(f)(\nu)h,h\rangle\rangle_\partial.$$

PROOF. This follows at once when we integrate by parts.

Pick an $f_0 \in U$. Then for all f in a neighborhood $\|f - f_0\|_0 < \varepsilon$ of f_0 the symbols $\sigma L(f)(dt)$ on X and $\sigma L(f)(\nu)$ on a neighborhood of $\partial^+ X$ will be uniformly positive-definite and $\sigma L(f)(\nu)$ on $\partial^- X$ will be negative-definite, for $\varepsilon > 0$ sufficiently small. If in addition $\|f - f_0\|_1 < \varepsilon$ then $\|f\|_1 \leqslant C$. We let $\tilde{U}$ denote any open neighborhood of f_0 such that for all $f \in U$ the symbols $\sigma L(f)(dt)$ and $\sigma L(f)(\nu)$ are uniformly positive- or negative-definite and $\|f\|_1$ is uniformly bounded.

3.4.3. LEMMA. *For all $f \in \tilde{U}$ and all h*

$$-\langle\langle L(f)h, h\rangle\rangle \leqslant C\big(\|h\|_0^2 + \|h \mid \partial^+ X\|_{0,+}^2\big).$$

PROOF. This follows from Lemma 3.4.2. Since $\|f\|_1 \leqslant C$ on $\tilde{U}$, $\|\mathrm{div}(f)\|_0 \leqslant C$, and

$$|\langle\langle \mathrm{div}(f)h, h\rangle\rangle| \leqslant C\|h\|_0^2.$$

On the other hand, $\sigma L(f)(\nu)$ is positive-definite on $\partial^+ X$ and negative-definite on $\partial^- X$, while $\|f\|_0 \leqslant C$ on $\tilde{U}$, so

$$\langle\langle \sigma L(f)(\nu)h, h\rangle\rangle_\partial \leqslant C\|h \mid \partial^+ X\|_{0,+}^2.$$

This proves the lemma.

Next we introduce the parameter p, and we measure $e^{-pt}h$ using the weight function t. The following estimates will hold for all p sufficiently large with various constants C all independent of p. We always take $f \in \tilde{U}$ and let $k = L(f)h$ and $h^+ = h \mid \partial^+ X$.

3.4.4. LEMMA. *We have the estimate*

$$p\|e^{-pt}h\|_0 \leqslant C\big(\|e^{-pt}k\|_0 + \sqrt{p}\,\|e^{-pt}h^+\|_{0,+}\big).$$

PROOF. For any function ϕ on X

$$L(f)\phi h = \phi L(f)h + \sigma L(f)(d\phi)h$$

and taking $\phi = e^{-pt}$

$$L(f)e^{-pt}h = e^{-pt}L(f)h - p\sigma L(f)(dt)e^{-pt}h.$$

Then

$$p\langle\langle \sigma L(f)(dt)e^{-pt}h, e^{-pt}h\rangle\rangle = \langle\langle e^{-pt}L(f)h, e^{-pt}h\rangle\rangle - \langle\langle L(f)e^{-pt}h, e^{-pt}h\rangle\rangle.$$

In the first term $\sigma L(f)(dt)$ is uniformly positive definite for $f \in \tilde{U}$, while the last term is bounded by Lemma 3.4.3. Thus

$$p\|e^{-pt}h\|_0^2 \leqslant C\|e^{-pt}h\|_0\|e^{-pt}k\|_0 + C\|e^{-pt}h\|_0^2 + C\|e^{-pt}h^+\|_{0,+}^2.$$

When $p \geqslant 2C$ we may eliminate the term $C\|e^{-pt}h\|_0^2$ on the right. Then completing the square gives

$$\left(\sqrt{p}\,\|e^{-pt}h\|_0 - \frac{C}{\sqrt{p}}\|e^{-pt}k\|_0\right)^2 \leqslant \frac{C}{p}\|e^{-pt}k\|_0^2 + C\|e^{-pt}h^+\|_{0,+}^2$$

which gives

$$\sqrt{p}\,\|e^{-pt}h\|_0 \leqslant \frac{C}{\sqrt{p}}\|e^{-pt}k\|_0 + C\|e^{-pt}h^+\|_{0,+}$$

which proves the lemma.

Now we prove the main estimate

3.4.5. LEMMA. *For all $n \geqslant 0$ we have*

$$p\|e^{-pt}h\|_n \leqslant C\big(\|e^{-pt}k\|_n + \sqrt{p}\,\|e^{-pt}h^+\|_{n,+}\big) + C(\|f\|_{n+1} + p^n)\big(\|e^{-pt}k\|_0 + \sqrt{p}\,\|e^{-pt}h^+\|_{0,+}\big).$$

PROOF. The result holds when $n = 0$ by the previous lemma. We assume it is true for some n and proceed by induction. To simplify the discussion we let

$$E_n = \|e^{-pt}k\|_n + \sqrt{p}\,\|e^{-pt}h^+\|_{n,+} + (\|f\|_{n+1} + p^n)\big(\|e^{-pt}k\|_0 + \sqrt{p}\,\|e^{-pt}h^+\|_{0,+}\big)$$

so that we have

$$p\|e^{-pt}h\|_n \leqslant E_n$$

for the induction hypothesis. We point out first that by interpolation (since $\|f\|_1 \leqslant C$)

$$\big(\|f\|_{l+1} + p^l\big)\|e^{-pt}k\|_m \leqslant CE_n,$$
$$\big(\|f\|_{l+1} + p^l\big)\sqrt{p}\,\|e^{-pt}h^+\|_{m,+} \leqslant CE_n$$

where $l, m \geqslant 0$ and $l + m \leqslant n$. As a consequence we have

$$pE_n \leqslant CE_{n+1}.$$

We begin by estimating tangential derivatives. Let v be any smooth vector field tangent to the boundary. Choose a first order linear differential operator ∇_v on sections of $V \approx V^*$ with symbol

$$\sigma\nabla_v(\xi) = \langle \xi, v \rangle.$$

Then there is a uniquely determined operator ∇_v on sections of $D_S^1(V, V^*)$ with

$$\nabla_v L(f) = L(f)\nabla_v + L(\nabla_v f).$$

In general if $\|f\|_0 \leqslant C$ we have

$$\|L(f)h\|_n \leqslant C(\|h\|_{n+1} + \|f\|_n\|h\|_1)$$

by interpolation. Now since

$$\nabla_v(e^{-pt}k) = e^{-pt}\nabla_v k - pe^{-pt}(\nabla_v t)k$$

we have

$$\|e^{-pt}\nabla_v k\|_n \leqslant C\|e^{-pt}k\|_{n+1} + Cp\|e^{-pt}k\|_n \leqslant CE_{n+1}.$$

Also since $\|\nabla_v f\|_0 \leqslant C\|f\|_1 \leqslant C$ we have

$$\|L(\nabla_v f)e^{-pt}h\|_n \leqslant C\big(\|e^{-pt}h\|_{n+1} + \|f\|_{n+1}\|e^{-pt}h\|_1\big)$$

and

$$p\|\sigma L(\nabla_v f)e^{-pt}h\|_n \leqslant C\big(p\|e^{-pt}h\|_n + p\|f\|_{n+1}\|e^{-pt}h\|_0\big).$$

Since $e^{-pt}L(\nabla_v f) = L(\nabla_v f)e^{-pt} + p\sigma L(\nabla_v f)e^{-pt}$, by interpolation we get

$$\|e^{-pt}L(\nabla_v f)h\|_n \leqslant C\|e^{-pt}h\|_{n+1} + C\big(\|f\|_{n+2} + p^{n+1}\big)\|e^{-pt}h\|_0.$$

But we know that for $p \geqslant 1$

$$\|e^{-pt}h\|_0 \leqslant p\|e^{-pt}h\|_0 \leqslant C\Big(\|e^{-pt}k\|_0 + \sqrt{p}\,\|e^{-pt}h^+\|_{0,+}\Big)$$

so

$$\|e^{-pt}L(\nabla_v f)h\|_n \leqslant C\|e^{-pt}h\|_{n+1} + CE_{n+1}.$$

Now $e^{-pt}L(f)\nabla_v h = e^{-pt}\nabla_v k - e^{-pt}L(\nabla_v f)h$, so that

$$\|e^{-pt}L(f)\nabla_v h\|_n \leqslant C\|e^{-pt}h\|_{n+1} + CE_{n+1}.$$

We now apply the induction hypothesis to $\nabla_v h$ instead of h. This gives

$$p\|e^{-pt}\nabla_v h\|_n \leqslant C\Big(\|e^{-pt}L(f)\nabla_v h\|_n + \sqrt{p}\,\|e^{-pt}\nabla_v h \mid \partial^+ X\|_{n,+}\Big)$$
$$+ C\big(\|f\|_{n+1} + p^n\big)\Big(\|e^{-pt}L(f)\nabla_v h\|_0 + \sqrt{p}\,\|e^{-pt}\nabla_v h \mid \partial^+ X\|_{0,+}\Big).$$

Since v is parallel to the boundary, and

$$e^{-pt}\nabla_v h = \nabla_v(e^{-pt}h) + p(\nabla_v t)e^{-pt}h$$

we have

$$\|e^{-pt}\nabla_v h \mid \partial^+ X\|_{n,+} \leqslant C\|e^{-pt}h^+\|_{n+1,+} + Cp\|e^{-pt}h^+\|_{n,+}.$$

This gives us

$$p\|e^{-pt}\nabla_v h\|_n \leqslant C\|e^{-pt}h\|_{n+1} + CE_{n+1}.$$

Now $\nabla_v(e^{-pt}h) = e^{-pt}\nabla_v h + p(\nabla_v t)e^{-pt}h$ so we get

$$p\|\nabla_v(e^{-pt}h)\|_n \leqslant C\|e^{-pt}h\|_{n+1} + Cp\|e^{-pt}h\|_n + CE_{n+1}.$$

By the induction hypothesis

$$p\|e^{-pt}h\|_n \leqslant CpE_n \leqslant CE_{n+1}$$

so we have the following estimate.

3.4.6. LEMMA. *If v is parallel to the boundary*

$$p\|\nabla_v(e^{-pt}h)\|_n \leqslant C\|e^{-pt}h\|_{n+1} + CE_{n+1}.$$

Next we must estimate the normal derivatives. To do this, we pick a finite number of vector fields $v_1, \ldots, v_{m-1}, v_m$ which span the tangent space at each point, such that $v_1, \ldots, v_{m-1}$ are all parallel to the boundary, while $\langle \nu, v_m \rangle = 1$. We choose operators ∇_j as before with symbols

$$\sigma\nabla_j(\xi) = \langle v_j, \xi \rangle I.$$

Let $g = (g_0, g_1, \ldots, g_m)$ be a section of the bundle $L_S(V, V^*)^{m+1}$, that is, $(m+1)$ sections of $L_S(V, V^*)$. We can form the operator

$$L(f) = g_0 + g_1\nabla_1 + \cdots + g_m\nabla_m$$

with $f \in D_S^1(V, V^*)$. This defines a linear vector bundle map

$$A\colon L_S(V, V^*)^{m+1} \to D_S^1(V, V^*)$$

with $Ag = f$. Since the vector fields span TX, A is surjective. Then we can find a smooth right inverse (which is not unique)

$$B\colon D_S^1(V, V^*) \to L_S(V, V^*)^{m+1}$$

with $AB = I$. Let

$$B(f) = (B_0(f), B_1(f), \ldots, B_m(f)) = (g_0, g_1, \ldots, g_m).$$

Then each B_j is a linear vector bundle map

$$B_j\colon D_S^1(V, V^*) \to L_S(V, V^*)$$

and for any f

$$L(f) = B_0(f) + B_1(f)\nabla_1 + \cdots + B_m(f)\nabla_m.$$

The symbol of $L(f)$ is

$$\sigma L(f)(\xi) = \sum_{j=1}^{m} \langle \xi, v_j \rangle B_j(f)$$

and since $v_1, \ldots, v_{m-1}$ are parallel to ∂X,

$$\sigma L(f)(\nu) = B_m(f).$$

For all $f \in \tilde{U}$ the automorphism $B_m(f)$ is uniformly invertible on a neighborhood of the boundary. Let ϕ be a smooth function with compact support in this neighborhood which is identically 1 on a smaller neighborhood of the boundary, and let

$$C_j(f) = \phi B_m(f)^{-1} B_j(f).$$

Then $\phi B_m(f)^{-1}$ and each $C_j(f)$ is a smooth nonlinear vector bundle operator, and for all $f \in \tilde{U}$ and $n \geqslant 0$

$$\|\phi B_m(f)^{-1}\|_n \leqslant C(1 + \|f\|_n), \qquad \|C_j(f)\|_n \leqslant C(1 + \|f\|_n).$$

We can write

$$\phi\nabla_m = \phi B_m(f)^{-1} L(f) - C_0(f) - \sum_{j<m} C_j(f)\nabla_j$$

and apply this formula to $e^{-pt}h$. Thus

$$\phi\nabla_m(e^{-pt}h) = \phi B_m(f)^{-1} L(f) e^{-pt}h - C_0(f)e^{-pt}h - \sum_{j<m} C_j(f)\nabla_j(e^{-pt}h).$$

This gives us

$$\begin{aligned}\|\phi\nabla_m(e^{-pt}h)\|_n \leqslant{}& C\big(\|L(f)e^{-pt}h\|_n + \|f\|_n \|L(f)e^{-pt}h\|_0\big)\\ &+ C\big(\|e^{-pt}h\|_n + \|f\|_n \|e^{-pt}h\|_0\big)\\ &+ C\sum_{j<m}\big(\|\nabla_j(e^{-pt}h)\|_n + \|f\|_n \|\nabla_j(e^{-pt}h)\|_0\big).\end{aligned}$$

Since $L(f)e^{-pt}h = e^{-pt}k - p\sigma L(f)(dt)e^{-pt}h$

$$\|L(f)e^{-pt}h\|_n \leqslant C\|e^{-pt}k\|_n + Cp\big(\|e^{-pt}h\|_n + \|f\|_n\|e^{-pt}h\|_0\big).$$

By interpolation we get

$$\begin{aligned}\|\phi\nabla_m(e^{-pt}h)\|_n \leqslant\ & C\big(\|e^{-pt}k\|_n + \|f\|_n\|e^{-pt}k\|_0\big)\\ & + Cp\big(\|e^{-pt}h\|_n + \|f\|_{n+1}\|e^{-pt}h\|_0\big)\\ & + C\sum_{j<m}\|\nabla_j(e^{-pt}h)\|_n.\end{aligned}$$

Now $p\|e^{-pt}h\|_n \leqslant CE_n$ by the induction hypothesis, while

$$p\|f\|_{n+1}\|e^{-pt}h\|_0 \leqslant C\|f\|_{n+1}\big(\|e^{-pt}k\|_0 + \sqrt{p}\,\|e^{-pt}h^+\|_{0,+}\big) \leqslant CE_n$$

also. Thus we get the following result.

3.4.7. LEMMA. *We can find a differential operator ∇_m with symbol $\sigma\nabla_m(\nu) = I$ on $\partial^+ X$, a finite number of operators ∇_j acting parallel to the boundary, and a function ϕ equal to* 1 *in a neighborhood of the boundary, with*

$$\|\phi\nabla_m(e^{-pt}h)\|_n \leqslant C\sum_{j<m}\|\nabla_j(e^{-pt}h)\|_n + CE_n.$$

Now clearly we have an estimate

$$\|h\|_{n+1} \leqslant C\Big(\|h\|_n + \|\phi\nabla_m h\|_n + \sum_v\|\nabla_v h\|_n\Big)$$

where v runs over a finite number of vector fields parallel to the boundary. Applying this to $e^{-pt}h$ and using Lemmas 3.4.6 and 3.4.7 we get

$$p\|e^{-pt}h\|_{n+1} \leqslant C\|e^{-pt}h\|_{n+1} + CE_{n+1}$$

remembering that $pE_n \leqslant CE_{n+1}$. Now if p is large enough compared to C (which may depend on n) we get

$$p\|e^{-pt}h\|_{n+1} \leqslant CE_{n+1}$$

which completes the induction. This proves Lemma 3.4.5.

3.4.8. COROLLARY. *For all n*

$$\|h\|_n \leqslant C(\|k\|_n + \|h^+\|_{n,+}) + C\|f\|_{n+1}(\|k\|_0 + \|h^+\|_{0,+}).$$

PROOF. Take any p large enough in Lemma 3.4.5. This shows that $h = S(f)(k, h^+)$ satisfies a tame estimate of base zero and degrees 0 in k and h^+ and degree 1 in f on a $\|\ \|_1$ neighborhood of any $f_0 \in U$. Thus S is tame. Since it is a family of inverses of a smooth tame linear family, S is even a smooth tame map.

Part III. THE NASH-MOSER THEOREM

III.1. The proof.

1.1. *Statement of the theorem.*

1.1.1. THE NASH-MOSER THEOREM. *Let F and G be tame spaces and P: $U \subseteq F \to G$ a smooth tame map. Suppose that the equation for the derivative $DP(f)h = k$ has a unique solution $h = VP(f)k$ for all f in U and all k, and that*

the family of inverses VP: $U \times G \to F$ is a smooth tame map. Then P is locally invertible, and each local inverse P^{-1} is a smooth tame map.

Observe by contrast with the inverse function theorem in Banach spaces that we need to invert the derivative $DP(f)$ for all f, not just one, and the whole collection of inverses must be "nice".

Of course we also have partial results if DP is merely injective or surjective. Let P be a smooth tame map.

1.1.2. THEOREM. *Suppose DP is injective with a smooth tame family of left inverses VP. Then P is locally injective.*

1.1.3. THEOREM. *Suppose DP is surjective with a smooth tame family of right inverses VP. Then P is locally surjective. Moreover in a neighborhood of any point P has a smooth tame right inverse.*

1.2. *Normalizations.* To begin the proof of the Nash-Moser theorem we make several normalizations which simplify the proof of the theorem. Let F and G be tame graded spaces and let P: $U \subseteq F \to G$ be a smooth tame map of an open set U in F into G. We suppose that $DP(f)h = k$ has a solution $h = VP(f)k$ where VP: $U \times G \to F$ is also a smooth tame map.

1.2.1. LEMMA. *We may assume $F = \Sigma(B)$ and $G = \Sigma(C)$ for two Banach spaces B and C.*

PROOF. If VP is invertible, then F and G are tamely isomorphic, and we may assume $F = G$, identifying them by $DP(0)$. Since F is tame, we can find an $\tilde{F}$ with $F \times \tilde{F} \approx \Sigma(B)$ for some Banach space B. Define a new map $\bar{P}$: $\bar{U} \subseteq \Sigma(B) \to \Sigma(B)$ by letting $\bar{U} = U \times \tilde{F}$ and

$$\bar{P}(f, \tilde{f}) = (P(f), \tilde{f}).$$

Then we have

$$D\bar{P}(f, \tilde{f})(h, \tilde{h}) = (DP(f)h, \tilde{h}) = (k, \tilde{k})$$

which has the solution

$$V\bar{P}(f, \tilde{f})(k, \tilde{k}) = (VP(f)k, \tilde{k}) = (h, \tilde{h}).$$

Then $\bar{P}$ satisfies all the same hypotheses as P.

If DP is only injective with a left inverse VP, we can let $F \times \tilde{F} \approx \Sigma(B)$ and $G \times \tilde{G} \approx \Sigma(C)$ and define a new map $\bar{P}$: $\bar{U} \subseteq \Sigma(B) \to \Sigma(B \times C)$ by letting $\bar{U} = U \times \tilde{F}$ and

$$\bar{P}(f, \tilde{f}) = (0, \tilde{f}, P(f), 0)$$

since $\Sigma(B \times C) = \Sigma(B) \times \Sigma(C) = F \times \tilde{F} \times G \times \tilde{G}$. Then

$$D\bar{P}(f, \tilde{f})(h, \tilde{h}) = (0, \tilde{h}, DP(f)h, 0)$$

is also injective with left inverse

$$V\bar{P}(f, \tilde{f})(l, \tilde{l}, k, \tilde{k}) = (VP(f)k, \tilde{l})$$

and $\bar{P}$ satisfies the same hypotheses as P. If DP is only surjective with a right inverse VP, we let $\bar{U} = U \times \tilde{F} \times G \times \tilde{G}$ and define a map $\bar{P}$: $\bar{U} \subseteq \Sigma(B \times C) \to \Sigma(C)$ by letting

$$\bar{P}(f, \tilde{f}, g, \tilde{g}) = (P(f), \tilde{g}).$$

Then

$$D\bar{P}(f, \tilde{f}, g, \tilde{g})(h, \tilde{h}, k, \tilde{k}) = (DP(f)h, \tilde{k})$$

is also surjective with right inverse

$$V\bar{P}(f, \tilde{f}, g, \tilde{g})(k, \tilde{k}) = (VP(f)k, \tilde{k})$$

and $\bar{P}$ satisfies the same hypotheses as P.

We wish to find a solution of the equation $P(f) = g$. Given any $f_0 \in U \subseteq F$ with $P(f_0) = g_0 \in G$, we shall find a solution f near f_0 for any g near g_0.

1.2.2. LEMMA. *We may assume* $f_0 = 0$ *and* $g_0 = 0$, *and hence that* $P(0) = 0$.

PROOF. We may replace $P(f) = g$ by $\bar{P}(\tilde{f}) = \tilde{g}$ where $\tilde{f} = f - f_0$ and $\tilde{g} = g - g_0$. Then $\bar{P}$ satisfies the same hypotheses as P, since it differs only by composition with translations, which are invertible.

The maps P, DP, D^2P and VP will all satisfy tame estimates on a neighborhood of zero. It is convenient to normalize these estimates.

1.2.3. LEMMA. *We may assume that* $P(f)$, $DP(f)h$, $D^2P(f)h^2$ *and* $VP(f)k$ *satisfy tame estimates of degree* $2r$ *in* f, r *in* h, *and* 0 *in* k (*for some* $r > 0$) *and base* 0 *on the set* $\{\| f \|_{2r} \leqslant 1\}$.

PROOF. On the space $F = \Sigma(B)$ of sequences $f = \{f_k\}$ of elements in B we have the operators ∇^p: $F \to F$ given by $(\nabla^p f)_k = e^{pk} f_k$, which satisfy $\| \nabla^p f \|_n = \| f \|_{n+p}$. Similar operators exist on $G = \Sigma(C)$. We may replace P by a new map of the form

$$\bar{P}(f) = \nabla^q P(\nabla^{-p} f).$$

If P satisfies an estimate

$$\| P(f) \|_n \leqslant C(1 + \| f \|_{n+s})$$

on $\| f \|_a \leqslant \delta$ for $n \geqslant b$, then $\bar{P}$ will satisfy an estimate

$$\| \bar{P}(f) \|_m \leqslant C(1 + \| f \|_{m+q-p+s})$$

on $\| f \|_{a-q} \leqslant \delta$ for all $m \geqslant b - q$. By choosing p and q large enough we can make P, DP, D^2P and VP satisfy tame estimates of base 0 on a set $\{\| f \|_0 \leqslant \delta\}$. Of course we cannot make them all have degree zero, since the degree of DP plus the degree of VP will be an invariant. But we do have enough freedom to make $VP(f)k$ have degree 0 in k. Then the degrees of $P(f)$, $DP(f)h$, $D^2P(f)h^2$ and $VP(f)k$ in f and h are finite, so they will be at most $2r$ in f and r in h for some r. For convenience we assume $r > 0$. Finally if we replace $P(f)$ by $\bar{P}(\tilde{f})$ where $\tilde{f} = cf$ for some constant c, this dilation will make $\delta = 1$. Since the dilation and the operators ∇^p are all invertible linear maps, the other hypotheses of the theorem all remain valid.

1.2.4. LEMMA. *On* $\|f\|_{2r} \leqslant 1$ *we have estimates*

$$\|P(f)\|_n \leqslant C\|f\|_{n+2r},$$
$$\|DP(f)h\|_n \leqslant C(\|h\|_{n+r} + \|f\|_{n+2r}\|h\|_r),$$
$$\|D^2P(f)\{h_1, h_2\}\|_n \leqslant C(\|h_1\|_{n+r}\|h_2\|_r + \|h_1\|_r\|h_2\|_{n+r} + \|f\|_{n+2r}\|h_1\|_r\|h_2\|_r),$$
$$\|VP(f)k\|_n \leqslant C(\|k\|_n + \|f\|_{n+2r}\|k\|_0)$$

for every $n \geqslant 0$ *with a constant* C *which may depend on* n.

PROOF. The second, third and fourth estimates are immediate consequences of the fact that the maps satisfy tame estimates of degree $2r$ in f, r in h (or h_1 and h_2) and 0 in k and are linear in h (or h_1 and h_2 separately) and k. For the first estimate we would expect $\|P(f)\|_n \leqslant C(1 + \|f\|_{n+2r})$, but $P(0) = 0$, so we can do without the 1 in the estimate. This may be seen by integrating the second estimate using the formula

$$P(f) = P(0) + \int_0^1 DP(tf)f\,dt$$

which gives us the estimate

$$\|P(f)\|_n \leqslant C(\|f\|_{n+r} + \|f\|_{n+2r}\|f\|_r).$$

Since $\|f\|_{n+r} \leqslant C\|f\|_{n+2r}$ and $\|f\|_r \leqslant C\|f\|_{2r} \leqslant C$ the result follows.

1.3. *Injectivity.* In this section we only need that DP is injective with a left inverse VP.

1.3.1. THEOREM. *There exists a* $\delta > 0$ *such that if* $\|f_0\|_{2r} \leqslant \delta$ *and* $\|f_1\|_{2r} \leqslant \delta$ *then*

$$\|f_1 - f_0\|_0 \leqslant C\|P(f_1) - P(f_0)\|_0.$$

Hence P *is injective on* $\|f\|_{2r} \leqslant \delta$.

PROOF. We use Taylor's theorem with integral remainder

$$P(f_1) = P(f_0) + DP(f_0)(f_1 - f_0) + \int_0^1 (1-t)D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\,dt.$$

Since $VP(f_0)DP(f_0)h = h$ we have

$$f_1 - f_0 = VP(f_0)\Big\{P(f_1) - P(f_0) - \int_0^1 (1-t)D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\,dt\Big\}.$$

Using the tame estimates on $\|f\|_{2r} \leqslant 1$

$$\|f_1 - f_0\|_0 \leqslant C\|P(f_1) - P(f_0)\|_0 + C\|f_1 - f_0\|_r^2.$$

Now by interpolation

$$\|f_1 - f_0\|_r^2 \leqslant C\|f_1 - f_0\|_{2r}\|f_1 - f_0\|_0$$

and $\|f_1 - f_0\|_{2r} \leq \|f_1\|_{2r} + \|f_0\|_{2r} \leq 2\delta$. Hence

$$\|f_1 - f_0\|_0 \leq C\|P(f_1) - P(f_0)\|_0 + 2C\delta\|f_1 - f_0\|_0.$$

If we take $\delta < 1/2C$ the result follows.

1.3.2. COROLLARY. *If* $\|f_0\|_{2r} \leq \delta$ *and* $\|f_1\|_{2r} \leq \delta$ *for the same* δ *as before, then for all* $n \geq 0$

$$\|f_1 - f_0\|_n \leq C\|P(f_1) - P(f_0)\|_n + C(\|f_1\|_{n+2r} + \|f_0\|_{n+2r})\|P(f_1) - P(f_0)\|_0.$$

PROOF. Using the higher norm estimates

$$\|D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\|_n \leq C\|f_1 - f_0\|_{n+r}\|f_1 - f_0\|_r + C(\|f_1\|_{n+2r} + \|f_0\|_{n+2r})\|f_1 - f_0\|_r^2.$$

By interpolation

$$\|f_1 - f_0\|_{n+r}\|f_1 - f_0\|_r \leq C\|f_1 - f_0\|_{n+2r}\|f_1 - f_0\|_0,$$
$$\|f_1 - f_0\|_r^2 \leq C\|f_1 - f_0\|_{2r}\|f_1 - f_0\|_0.$$

But

$$\|f_1 - f_0\|_{n+2r} \leq \|f_1\|_{n+2r} + \|f_0\|_{n+2r}, \qquad \|f_1 - f_0\|_{2r} \leq C$$

so

$$\|D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\|_n \leq C(\|f_1\|_{n+2r} + \|f_0\|_{n+2r})\|f_1 - f_0\|_0.$$

Then using the previous formula and the higher norm estimates on VP

$$\begin{aligned}\|f_1 - f_0\|_n \leq{}& C\|P(f_1) - P(f_0)\|_n \\ &+ C(\|f_1\|_{n+2r} + \|f_0\|_{n+2r})\|P(f_1) - P(f_0)\|_0 \\ &+ C\|D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\|_n \\ &+ C(\|f_1\|_{n+2r} + \|f_0\|_{n+2r})\|D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\|_0.\end{aligned}$$

Plugging in our earlier estimate on D^2P and using $\|f_1 - f_0\|_0 \leq C\|P(f_1) - P(f_0)\|_0$ from the previous theorem completes the proof.

1.3.3. COROLLARY. *We also have*

$$\|f_1 - f_0 - VP(f_0)[P(f_1) - P(f_0)]\|_n \leq C\|f_1 - f_0\|_{n+2r}\|f_1 - f_0\|_0 + C\|f_0\|_{n+2r}\|f_1 - f_0\|_0^2.$$

PROOF. By our previous estimate

$$\|D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\|_n \leq C\|f_1 - f_0\|_{n+2r}\|f_1 - f_0\| + C\|f_0\|_{n+2r}\|f_1 - f_0\|_0^2$$

since $\| f_1 \|_{n+2r} \leq \| f_0 \|_{n+2r} + \| f_1 - f_0 \|_{n+2r}$. We can rewrite our formula as

$$f_1 - f_0 - VP(f_0)[P(f_1) - P(f_0)] = -VP(f_0)\int_0^1 (1-t)D^2P((1-t)f_0 + tf_1)(f_1 - f_0)^2\, dt.$$

Then using the estimates on VP the result follows.

1.3.4. *Problem.* How much more is true? Is $\| f \|_n \leq C \| P(f) \|_{n+s}$ for some s? Is the image of P relatively closed in a neighborhood of zero? Is P a homeomorphism onto its image? Is the image a submanifold? Does there exist a smooth tame left inverse for P locally?

1.4. *Smoothing operators.* The proof of the Nash-Moser theorem involves the use of smoothing operators, which are particularly easy to construct on the model spaces $\Sigma(B)$. For each t we construct a linear map S_t: $\Sigma(B) \to \Sigma(B)$ such that $S_t = 0$ for $t \leq 0$ and $S_t \to I$ as $t \to \infty$. Choose a smooth function $s(u)$ with $s(u) = 0$ for $u \leq 0$ and $s(u) = 1$ for $u \geq 1$, while $0 \leq s(u) \leq 1$ in between. If $f = \{f_k\}$ is a sequence in $\Sigma(B)$ we let

$$(S_t f)_k = s(t-k) f_k.$$

The smoothing operators S_t satisfy the following estimates.

1.4.1. LEMMA. *For all $m \leq n$*

$$\| S_t f \|_n \leq Ce^{(n-m)t} \| f \|_m, \qquad \| (I - S_t) f \|_m \leq Ce^{(m-n)t} \| f \|_n$$

with a constant C which may depend on m and n.

PROOF. Recall that

$$\| f \|_n = \sum_k e^{nk} \| f_k \|_B.$$

Now $\| (S_t f)_k \|_B \leq \| f_k \|_B$ for all t and k, while $(S_t f)_k = 0$ for $t \leq k$ and $([I - S_t] f)_k = 0$ for $t \geq k + 1$. Thus

$$\| S_t f \|_n \leq \sum_{k \leq t} e^{nk} \| f_k \|_B \leq e^{(n-m)t} \| f \|_m$$

and

$$\| (I - S_t) f \|_m \leq \sum_{k \geq t-1} e^{mk} \| f_k \|_B \leq Ce^{(m-n)t} \| f \|_n$$

with $C = e^{n-m}$.

1.4.2. COROLLARY. *If $l \leq m \leq n$ then we have the interpolation estimate*

$$\| f \|_m^{n-l} \leq C \| f \|_l^{n-m} \| f \|_n^{m-l}$$

with a constant C depending on l, m and n.

PROOF. For all t

$$\| f \|_m \leq \| S_t f \|_m + \| (I - S_t) f \|_m \leq Ce^{(m-l)t} \| f \|_l + Ce^{(m-n)t} \| f \|_n.$$

If $f \neq 0$, we can choose t with the last two terms equal. This happens when

$$e^{(n-l)t} = \| f \|_n / \| f \|_l.$$

Then

$$\| f \|_m^{n-l} \leqslant C \| f \|_l^{n-m} \| f \|_n^{m-l}$$

as desired.

We shall need to estimate the solution of the equation in $\Sigma(B)$,

$$k_t' + cS_t k_t = l_t$$

where c is a constant. Since k_t is a smooth path in $\Sigma(B)$, it is represented by a sequence of smooth paths $k_{t,j}$ in B for $j = 0, 1, 2, \ldots$. Likewise l_t is represented by a sequence $l_{t,j}$ of paths in B. We define a smooth path a_t in $\Sigma(R)$, represented by a sequence $a_{t,j}$ of smooth functions with values in R, by setting

$$a_{t,j} = \exp\left(c \int_0^t s(\tau - j)\, d\tau \right).$$

Letting a prime denote differentiation with respect to t, we have

$$a_{t,j}' = cs(t-j) a_{t,j}$$

and

$$(a_t k_t)' = a_t (k_t' + cS_t k_t) = a_t l_t.$$

Therefore

$$a_t k_t = a_0 k_0 + \int_{\theta=0}^t a_\theta l_\theta \, d\theta.$$

Let $a_{\theta,t} = a_\theta / a_t$, so that

$$a_{\theta,t,j} = \exp\left(-c \int_\theta^t s(\tau - j)\, d\tau \right).$$

Dividing by a_t we get the following result.

1.4.3. THEOREM. *If* $k_t' + cS_t k_t = l_t$ *then*

$$k_t = a_{0,t} k_0 + \int_{\theta=0}^t a_{\theta,t} l_\theta \, d\theta.$$

This is of course the classical formula for the solution of a linear first order ordinary differential equation, adapted to our space of sequences in a Banach space. In order to use the formula we must estimate the kernel $a_{\theta,t}$. Clearly we always have $0 \leqslant a_{\theta,t,j} \leqslant 1$.

1.4.4. LEMMA. *For all* θ, t, *and* j *we have*

$$e^{ct} a_{\theta,t,j} \leqslant C(e^{c\theta} + e^{cj}).$$

PROOF. If $t \leqslant \theta$ or $t \leqslant j + 1$ the result follows from $a_{\theta,t,j} \leqslant 1$. If $\theta \geqslant j + 1$ then $s(\tau - j) = 1$ for all $\tau \geqslant \theta$, so $a_{\theta,t,j} = e^{-c(t-\theta)}$ and $e^{ct} a_{\theta,t,j} = e^{c\theta}$. If $\theta \leqslant j + 1$ then

$$a_{\theta,t,j} \leqslant \exp\left(-c \int_{j+1}^t s(\tau - j)\, d\tau \right) = e^{-c(t-j-1)}$$

so $e^{ct} a_{\theta,t,j} \leqslant e^{c(j+1)} \leqslant e^c e^{cj}$. Therefore the lemma holds with $C = e^c$.

1.4.5. LEMMA. *If* $0 < q < c$ *then*

$$\int_{t=\theta}^{\infty} e^{qt} a_{\theta,t,j}\, dt \leqslant C(e^{qj} + e^{q\theta}).$$

PROOF. If $j \leqslant \theta$ we have $a_{\theta,t,j} \leqslant Ce^{c(\theta - t)}$ and

$$\int_{t=\theta}^{\infty} e^{qt} a_{\theta,t,j}\, dt \leqslant Ce^{c\theta} \int_{t=\theta}^{\infty} e^{(q-c)t}\, dt \leqslant Ce^{q\theta}$$

provided $q < c$. If $\theta < j$ we break up the integral into two parts. Since $a_{\theta,t,j} \leqslant 1$

$$\int_{t=\theta}^{j} e^{qt} a_{\theta,t,j}\, dt \leqslant \int_{t=\theta}^{j} e^{qt}\, dt \leqslant Ce^{qj}$$

while since $a_{\theta,t,j} \leqslant Ce^{c(j-t)}$

$$\int_{t=j}^{\infty} e^{qt} a_{\theta,t,j}\, dt \leqslant Ce^{cj} \int_{t=j}^{\infty} e^{(q-c)t}\, dt \leqslant Ce^{qj}.$$

This proves the lemma.

1.4.6. THEOREM. *If* $k_t' + cS_t k_t = l_t$ *for* $0 \leqslant t \leqslant T$ *then for all* $p \geqslant 0$ *and for* $0 < q < c$

$$\int_0^T e^{qt} \| k_t \|_p\, dt \leqslant C \| k_0 \|_{p+q} + C \int_0^T e^{qt} \| l_t \|_p + \| l_t \|_{p+q}\, dt.$$

PROOF. We have

$$k_t = a_{0,t} k_0 + \int_{\theta=0}^{t} a_{\theta,t} l_\theta\, d\theta.$$

Since

$$\| k_t \|_p = \sum_j e^{pj} \| k_{t,j} \|_B$$

we have

$$\int_0^T e^{qt} \| k_t \|_p\, dt = \int_0^T \sum_j e^{qt} e^{pj} \| k_{t,j} \|_B\, dt$$

$$\leqslant \int_{t=0}^{T} \sum_j e^{qt} e^{pj} \left\{ a_{0,t,j} \| k_{0,j} \|_B + \int_{\theta=0}^{t} a_{\theta,t,j} \| l_{\theta,j} \|_B\, d\theta \right\} dt.$$

The first part is bounded by

$$\sum_j e^{pj} \left\{ \int_{t=0}^{T} e^{qt} a_{0,t,j}\, dt \right\} \| k_{0,j} \|_B \leqslant C \sum_j e^{(p+q)j} \| k_{0,j} \|_B \leqslant C \| k_0 \|_{p+q}.$$

The second part is bounded by

$$\int_{\theta=0}^{T} \sum_j e^{pj} \left\{ \int_{t=\theta}^{T} e^{qt} a_{\theta,t,j}\, dt \right\} \| l_{\theta,j} \|_B\, d\theta \leqslant \int_{\theta=0}^{T} \sum_j e^{pj} (e^{qj} + e^{q\theta}) \| l_{\theta,j} \|_B\, d\theta$$

$$\leqslant C \int_{\theta=0}^{T} \| l_\theta \|_{p+q} + e^{q\theta} \| l_\theta \|_p\, d\theta$$

which completes the proof of the theorem.

1.5. *Surjectivity.* We now describe the plan of attack for solving $P(f) = g$. If we were working with Banach spaces, we could solve an ordinary differential equation for a path f_t starting at $f_0 = 0$ such that $P(f_t)$ goes in a straight line from $P(0) = 0$ to g. The equation is

$$f_t' = cVP(f_t)[g - P(f_t)]$$

where a prime denotes differentiation of a path with respect to t, and c is a constant. Since $DP \circ VP = I$ we have

$$DP(f_t)f_t' = c[g - P(f_t)].$$

If $k_t = g - P(f_t)$ is the error then $k_0 = g$ and

$$k_t' + ck_t = 0$$

which shows that $k_t = e^{-ct}g$. Thus the error goes to zero as fast as any exponential we desire, and f_t converges to a solution f of $P(f) = g$ as $t \to \infty$.

In Fréchet spaces there is no reason to expect that the differential equation will have a solution. However the spaces $\Sigma(B)$ and $\Sigma(C)$ possess the smoothing operators S_t described before, which may be used to mollify the differential equation so that it will have a solution. Since $S_t \to I$ as $t \to \infty$, the behavior in the limit is much the same as before.

1.5.1. *Algorithm.* Let f_t be the path in $\Sigma(B)$ with $f_0 = 0$ which solves the differential equation

$$f_t' = cVP(S_t f_t)S_t[g - P(f_t)].$$

We shall show that f_t converges to a solution f of $P(f) = g$ as $t \to \infty$, while the error $k_t = g - P(f_t)$ goes to zero exponentially in any norm. To obtain this result we must choose $c > 2r + 1$. To see that the error does go to zero, we observe that k_t itself satisfies a differential equation.

1.5.2. LEMMA. *Let f_t be the solution of the algorithm and let*

$$k_t = g - P(f_t), \qquad l_t = [DP(S_t f_t) - DP(f_t)]f_t'.$$

Then

$$k_t' + cS_t k_t = l_t.$$

PROOF. The algorithm says that

$$f_t' = cVP(S_t f_t)S_t k_t.$$

Since $DP \circ VP = I$ we have

$$DP(S_t f_t)f_t' = cS_t k_t.$$

Now $P(f_t)' = DP(f_t)f_t'$, so

$$k_t' + cS_t k_t = [DP(S_t f_t) - DP(f_t)]f_t' = l_t.$$

The question naturally arises whether the algorithm will have any solution at all. We can write the differential equation as

$$f_t' = V(t, f_t, g)$$

where

$$V(t, f, g) = cVP(S_t f)S_t[g - P(f)].$$

Here V maps an open set in $R \times F \times G$ into F. Observe that V factors through a Banach space. For let $\Sigma_T(B)$ denote the subspace of sequences $\{f_k\}$ in $\Sigma(B)$ with $f_k = 0$ for $k > T$. Then $\Sigma_T(B)$ is a closed subspace and all the norms $\| \ \|_n$ are equivalent on $\Sigma_T(B)$. Hence $\Sigma_T(B)$ is a Banach space. The image of the smoothing operator S_t lies in $\Sigma_T(B)$ for $t \leqslant T$. Hence $S_t f \in \Sigma_T(B)$ and $S_t[g - P(f)] \in \Sigma_T(C)$ for $t \leqslant T$. This shows that V factors through the Banach space $\Sigma_T(B) \times \Sigma_T(C)$ for $t \leqslant T$. Indeed if we let

$$Z(t, f, g) = (S_t f, S_t[g - P(f)]), \qquad W(\tilde{f}, k) = cVP(\tilde{f})k$$

then $V = W \circ Z$ and Z maps into $\Sigma_T(B) \times \Sigma_T(C)$.

It follows by Theorem I.5.6.3 that the algorithm has a solution for at least a short time with any initial data in its domain. Moreover there will be a largest half-open interval $[0, \omega)$ with $\omega \leqslant \infty$ on which the solution exists. We shall establish a priori estimates for the solution independent of ω, which show that the solution remains in the domain of definition and is uniformly continuous. If ω were finite, the solution f_t would converge to a value of f in the domain of definition as $t \to \omega$. Since $\omega < \infty$, the equation still factors through a Banach space. We could then solve for a solution on $[\omega, \omega + \varepsilon)$ which extends the original solution to $[0, \omega + \varepsilon)$, contradicting the optimality of ω. It then follows that $\omega = \infty$, the solution exists for all time, and converges to a solution of $P(f) = g$ as $t \to \infty$. This is the plan. To make it work, we turn now to the derivation of the a priori estimates.

1.6. *A priori estimates.* We now estimate the various terms in the algorithm using the tame estimates on P, DP, D^2P and VP.

1.6.1. LEMMA. *For all $n \geqslant 0$ and $q \geqslant 0$*

$$\| f_t' \|_{n+q} \leqslant Ce^{qt}(\| k_t \|_n + \| f_t \|_{n+2r} \| k_t \|_0).$$

PROOF. Since $f_t' = cVP(S_t f_t) S_t k_t$ we have

$$\| f_t' \|_{n+q} \leqslant C \| S_t k_t \|_{n+q} + C \| S_t f_t \|_{n+q+2r} \| S_t k_t \|_0.$$

Using the estimates on the smoothing operators the result is immediate.

1.6.2. LEMMA. *For all $n \geqslant 0$*

$$\| l_t \|_n \leqslant C \| f_t \|_{n+2r} \| k_t \|_0.$$

PROOF. We have

$$l_t = [DP(S_t f_t) - DP(f_t)] f_t'.$$

Then

$$l_t = -B(f_t, S_t f_t)\{(I - S_t) f_t, f_t'\}$$

where

$$B(f, \tilde{f})\{h, \tilde{h}\} = \int_{u=0}^{1} D^2P((1-u)f + u\tilde{f})\{h, \tilde{h}\} \, du.$$

It is immediate that $B(f, \tilde{f})\{h, \tilde{h}\}$ is smooth and bilinear in h and $\tilde{h}$ and satisfies a tame estimate of degree $2r$ in f and $\tilde{f}$ and degree r in h and $\tilde{h}$ and base 0. Thus

$$\|B(f, \tilde{f})\{h, \tilde{h}\}\|_n \leq C\|h\|_{n+r}\|\tilde{h}\|_r + C\|h\|_r\|\tilde{h}\|_{n+r} + C(\|f\|_{n+2r} + \|\tilde{f}\|_{n+2r})\|h\|_r\|\tilde{h}\|_r.$$

Applying this estimate to l_t,

$$\|l_t\|_n \leq C\|(I - S_t)f_t\|_{n+r}\|f_t'\|_r + \|(I - S_t)f_t\|_r\|f_t'\|_{n+r} + C(\|f_t\|_{n+2r} + \|S_t f_t\|_{n+2r})\|(I - S_t)f_t\|_r\|f_t'\|_r.$$

We estimate these three terms using the properties of the smoothing operators and our previous estimate for f_t'. For the first use

$$\|(I - S_t)f_t\|_{n+r} \leq Ce^{-rt}\|f_t\|_{n+2r} \quad \text{and} \quad \|f_t'\|_r \leq Ce^{rt}\|k_t\|_0.$$

For the second use

$$\|(I - S_t)f_t\|_r \leq Ce^{-(n+r)t}\|f_t\|_{n+2r} \quad \text{and} \quad \|f_t'\|_{n+r} \leq Ce^{(n+r)t}\|k_t\|_0.$$

For the third use $\|S_t f_t\|_{n+2r} \leq C\|f_t\|_{n+2r}$ and

$$\|(I - S_t)f_t\|_r \leq Ce^{-rt}\|f_t\|_{2r} \leq Ce^{-rt} \quad \text{and} \quad \|f_t'\|_r \leq Ce^{rt}\|k_t\|_0$$

since $\|f_t\|_{2r} \leq 1$. Then the lemma follows.

The following barrier estimate will show that the solution remains bounded in the lowest norm.

1.6.3. LEMMA. *If the algorithm has a solution f_t on $0 \leq t \leq T$ then with a constant C independent of T*

$$\int_0^T e^{2rt}\|k_t\|_0\,dt \leq C\|g\|_{2r} + C\left(\int_0^T e^{2rt}\|k_t\|_0\,dt\right)^2.$$

PROOF. By our estimate for the solution of $k_t' + cS_t k_t = l_t$ we have (for $c > 2r$)

$$\int_0^T e^{2rt}\|k_t\|_0\,dt \leq C\|g\|_{2r} + C\int_0^T e^{2rt}\|l_t\|_0 + \|l_t\|_{2r}\,dt.$$

Our estimates for f_t' say that

$$\|f_t'\|_{2r} \leq Ce^{2rt}\|k_t\|_0 \quad \text{and} \quad \|f_t'\|_{4r} \leq Ce^{4rt}\|k_t\|_0$$

from which we see that

$$\|f_t\|_{2r} \leq C\int_0^t e^{2r\theta}\|k_\theta\|_0\,d\theta,$$

$$\|f_t\|_{4r} \leq C\int_0^t e^{4r\theta}\|k_\theta\|_0\,d\theta \leq Ce^{2rt}\int_0^t e^{2r\theta}\|k_\theta\|_0\,d\theta.$$

By our estimates for l_t we have

$$\|l_t\|_0 \leq C\|f_t\|_{2r}\|k_t\|_0 \quad \text{and} \quad \|l_t\|_{2r} \leq C\|f_t\|_{4r}\|k_t\|_0$$

which shows that

$$e^{2rt}\|l_t\|_0 + \|l_t\|_{2r} \leqslant Ce^{2rt}\|k_t\|_0 \int_0^t e^{2r\theta}\|k_\theta\|_0\, d\theta.$$

For $t \leqslant T$ the integral from 0 to t is of course bounded by the integral from 0 to T. Then if we integrate this estimate the result follows, since

$$\int_0^T e^{2rt}\|l_t\|_0 + \|l_t\|_{2r}\, dt \leqslant C\left(\int_0^T e^{2rt}\|k_t\|_0\, dt\right)^2.$$

To apply the barrier estimate, let

$$K_T = \int_0^T e^{2rt}\|k_t\|_0\, dt.$$

Then we have $K_T \leqslant C\|g\|_{2r} + CK_T^2$.

1.6.4. LEMMA. *Either* $K_T \leqslant 2C\|g\|_{2r}$ *or* $K_T \geqslant 1/2C$.

PROOF. We have $K_T(1 - CK_T) \leqslant C\|g\|_{2r}$. If $K_T \leqslant 1/2C$ then $1 - CK_T \geqslant 1/2$ and $K_T \leqslant 2C\|g\|_{2r}$.

Now K_T clearly depends continuously on T. If we choose $\|g\|_{2r} < 1/4C^2$ with the constant C above, then $2C\|g\|_{2r} < 1/2C$. It follows from the estimate above that K_T cannot cross the forbidden region between $2C\|g\|_{2r}$ and $1/2C$. Since $K_0 = 0$, K_T must remain on the small side, and we get the following result.

1.6.5. LEMMA. *If* $\|g\|_{2r} \leqslant \delta$ *and* δ *is sufficiently small then*

$$\int_0^T e^{2rt}\|k_t\|_0\, dt \leqslant C\|g\|_{2r}.$$

1.6.6. COROLLARY. *We also have* $\|f_t\|_{2r} \leqslant C\|g\|_{2r}$.

PROOF. We have $\|f_t'\|_{2r} \leqslant Ce^{2rt}\|k_t\|_0$, and the result follows by integrating. If $\|g\|_{2r} \leqslant \delta$ and $C\delta \leqslant 1$ then $\|f_t\|_{2r} \leqslant 1$, and the solution of the algorithm remains in the domain of definition.

We still have to estimate the higher order norms. We begin with a preliminary step to start an induction.

1.6.7. LEMMA. *If* $\|g\|_{2r} \leqslant \delta$ *and* δ *is sufficiently small then*

$$\int_0^T e^{(2r+1)t}\|k_t\|_0\, dt \leqslant C\|g\|_{2r+1}.$$

PROOF. By our estimate of the solution of $k_t' + cS_tk_t = l_t$ we have (when $c > 2r + 1$)

$$\int_0^T e^{(2r+1)t}\|k_t\|_0\, dt \leqslant C\|g\|_{2r+1} + C\int_0^T e^{(2r+1)t}\|l_t\|_0 + \|l_t\|_{2r+1}\, dt.$$

Since $\| f_t' \|_{2r} \leqslant Ce^{2rt} \| k_t \|_0$ and $\| f_t' \|_{4r+1} \leqslant Ce^{(4r+1)t} \| k_t \|_0$ from Lemma 1.6.1, we have

$$\| f_t \|_{2r} \leqslant C \int_0^t e^{2r\theta} \| k_\theta \|_0 \, d\theta \leqslant C \| g \|_{2r},$$

$$\| f_t \|_{4r+1} \leqslant C \int_0^t e^{(4r+1)\theta} \| k_\theta \|_0 \, d\theta \leqslant Ce^{(2r+1)t} \int_0^t e^{2r\theta} \| k_\theta \|_0 \, d\theta$$

$$\leqslant Ce^{(2r+1)t} \| g \|_{2r}.$$

By the estimate for l_t in Lemma 1.6.2 we have

$$\| l_t \|_0 \leqslant C \| f_t \|_{2r} \| k_t \|_0 \quad \text{and} \quad \| l_t \|_{2r+1} \leqslant C \| f_t \|_{4r+1} \| k_t \|_0$$

so that

$$e^{(2r+1)t} \| l_t \|_0 + \| l_t \|_{2r+1} \leqslant Ce^{(2r+1)t} \| g \|_{2r} \| k_t \|_0.$$

Integrating this estimate we have

$$\int_0^T e^{(2r+1)t} \| k_t \|_0 \, dt \leqslant C \| g \|_{2r+1} + C \| g \|_{2r} \int_0^T e^{(2r+1)t} \| k_t \|_0 \, dt.$$

If $C \| g \|_{2r} < 1$ we may subtract the last term on the right from the left, and this proves the lemma.

We can now prove the following estimate.

1.6.8. THEOREM. *If f_t is a solution of the algorithm and if $\| g \|_{2r} \leqslant \delta$ with δ sufficiently small, then for all $n \geqslant 2r$ and all $q \geqslant 0$ we have*

$$\int_0^T \| f_t' \|_{n+q} \, dt \leqslant Ce^{qT} \| g \|_n$$

with a constant C which may depend on n and q but is independent of T. Note that $\delta > 0$ is independent of n and q.

PROOF. We write $n = p + 2r$. We shall do the proof by induction on p. From Lemma 1.6.1 we have

$$\| f_t' \|_{1+q+2r} \leqslant Ce^{(1+q+2r)t} \| k_t \|_0$$

so that by the last lemma

$$\int_0^T \| f_t' \|_{1+q+2r} \, d\theta \leqslant Ce^{qT} \int_0^T e^{(2r+1)\theta} \| k_\theta \|_0 \, d\theta \leqslant Ce^{qT} \| g \|_{2r+1}.$$

This starts the induction for $p = 1$.

Assume that for some $p \geqslant 1$ and all $q \geqslant 0$ we have

$$\int_0^T \| f_t' \|_{p+q+2r} \, d\theta \leqslant Ce^{qT} \| g \|_{p+2r}.$$

Then $\| f_t \|_{p+2r} \leqslant C \| g \|_{p+2r}$ and $\| f_t \|_{p+4r+1} \leqslant Ce^{(2r+1)t} \| g \|_{p+2r}$. Consequently by Lemma 1.6.2

$$e^{(2r+1)t} \| l_t \|_p + \| l_t \|_{p+2r+1} \leqslant Ce^{(2r+1)t} \| g \|_{p+2r} \| k_t \|_0.$$

By the estimate for the solution of the equation $k_t' + cS_t k_t = l_t$ we have

$$\int_0^T e^{(2r+1)t} \| k_t \|_p \, dt \leqslant C \| g \|_{p+2r+1} + C \int_0^T e^{(2r+1)t} \| l_t \|_p + \| l_t \|_{p+2r+1} \, dt.$$

Since

$$\int_0^T e^{(2r+1)t} \| k_t \|_0 \, dt \leqslant C \| g \|_{2r+1}$$

we have

$$\int_0^T e^{(2r+1)t} \| k_t \|_p \, dt \leqslant C \| g \|_{p+2r+1} + C \| g \|_{p+2r} \| g \|_{2r+1}.$$

By Lemma 1.6.1 we have

$$\| f_t' \|_{p+q+2r+1} \leqslant C e^{(q+2r+1)t} \big(\| k_t \|_p + \| f_t \|_{p+2r} \| k_t \|_0 \big)$$

and since

$$\| f_t \|_{p+2r} \leqslant C \| g \|_{p+2r} \quad \text{and} \quad \int_0^T e^{(2r+1)t} \| k_t \|_0 \, dt \leqslant C \| g \|_{2r+1}$$

we have

$$\int_0^T \| f_t' \|_{p+1+q+2r} \, dt \leqslant C e^{qT} \big(\| g \|_{p+1+2r} + \| g \|_{p+2r} \| g \|_{2r+1} \big).$$

When $\| g \|_{2r} \leqslant \delta \leqslant 1$ we have by interpolation

$$\| g \|_{p+2r} \| g \|_{2r+1} \leqslant C \| g \|_{p+2r+1} \| g \|_{2r} \leqslant C \| g \|_{p+1+2r}$$

which eliminates the last term and completes the induction.

1.6.9. COROLLARY. *If f_t is a solution of the algorithm on $0 \leqslant t < \omega \leqslant \infty$ and if $\| g \|_{2r} \leqslant \delta$ with δ sufficiently small then*

$$\| f_t \|_n \leqslant C \| g \|_n \quad \text{for } n \geqslant 2r$$

with a constant C depending on n but independent of t, and

$$\lim_{t, \tau \to \omega} \| f_t - f_\tau \|_n = 0.$$

PROOF. The first estimate follows from

$$\| f_t \|_n \leqslant \int_0^t \| f_\theta' \|_n \, d\theta \leqslant C \| g \|_n$$

which is the previous theorem with $q = 0$. Since

$$\int_0^T \| f_t' \|_n \, dt \leqslant C \| g \|_n$$

with a constant independent of T, we have also

$$\int_0^\omega \| f_t' \|_n \, dt \leqslant C \| g \|_n < \infty$$

and therefore

$$\lim_{\alpha \to \omega} \int_\alpha^\omega \| f_t' \|_n \, dt = 0$$

since $\| f_t' \|_n$ is a continuous function of t with a finite improper integral. Then

$$\lim_{t,\tau\to\omega} \| f_t - f_\tau \|_n \leqslant \lim_{t,\tau\to\omega} \int_\tau^t \| f_\theta' \|_n \, d\theta = 0$$

and the corollary holds.

As indicated before, we consider the largest open interval $0 \leqslant t < \omega < \infty$ on which the solution exists. If ω were finite, the previous estimates show that $f_t \to f_\omega$ as $t \to \omega$ for some $f_\omega \in F$ with $\| f_\omega \|_{2r} \leqslant 1$. We could then use our existence theorem to prolong the solution past ω. Therefore $\omega = \infty$, the solution exists for all time, and $f_t \to f_\infty$ as $t \to \infty$ for some $f_\infty \in F$ with $\| f_\infty \|_{2r} \leqslant 1$. Moreover we have $\| f_\infty \|_n \leqslant C \| g \|_n$ for all $n \geqslant 2r$. Since $f_t' = cVP(S_t f_t) S_t [g - P(f_t)]$ and f_t converges to f_∞ as $t \to \infty$, we must also have f_t' converging as $t \to \infty$. Since for all n

$$\int_0^\infty \| f_t' \|_n \, dt \leqslant C \| g \|_n < \infty$$

we must have f_t' converging to zero. Since S_t converges to the identity

$$VP(f_\infty)[g - P(f_\infty)] = 0$$

and if we apply $DP(f_\infty)$, we get that $g - P(f_\infty) = 0$, so $P(f_\infty) = g$. Hence f_∞ is a solution. This shows the following.

1.6.10. THEOREM. *If $\| g \|_{2r} \leqslant \delta$ with δ sufficiently small, then the algorithm f_t converges at $t \to \infty$ to a solution f_∞ of $P(f_\infty) = g$, and for all $n \geqslant 2r$ we have*

$$\| f_\infty \|_n \leqslant C \| g \|_n.$$

We can show the convergence more dramatically.

1.6.11. COROLLARY. *If $\| g \|_{2r} \leqslant \delta$ as above, then for all $n \geqslant 2r$ we have*

$$\int_0^\infty e^t \| f_t' \|_n \, dt \leqslant C \| g \|_{n+1}$$

and consequently

$$\| f_t - f_\infty \|_n \leqslant C e^{-t} \| g \|_{n+1}.$$

PROOF. By the end of the proof of Theorem 1.6.8 we have

$$\int_0^\infty e^{(2r+1)t} \| k_t \|_p \, dt \leqslant C \| g \|_{p+2r+1}$$

for all $p \geqslant 0$. By Lemma 1.6.1

$$\| f_t' \|_n \leqslant C e^{2rt} (\| k_t \|_{n-2r} + \| f_t \|_n \| k_t \|_0).$$

We also have $\| f_t \|_n \leqslant C \| g \|_n$. Then integrating

$$\int_0^\infty e^t \| f_t' \|_n \, dt \leqslant C \| g \|_{n+1} + C \| g \|_n \| g \|_{2r+1}.$$

Since $\| g \|_{2r} \leqslant \delta \leqslant 1$ we have

$$\| g \|_n \| g \|_{2r+1} \leqslant C \| g \|_n$$

by interpolation. This proves the first result, and the second follows immediately.

1.7. *The inverse map.* We finish the proof of Theorem 1.1.1 where DP is invertible. We have seen that in neighborhoods of 0 the map P is both injective and surjective, and hence invertible. From Theorem 1.6.10 we have the estimate

$$\|P^{-1}(g)\|_n \leqslant C\|g\|_n$$

for all $n \geqslant 2r$, which shows that P^{-1} satisfies a tame estimate with degree 0 and base $2r$. Moreover by Theorem 1.3.2 if $P(f_0) = g_0$ and $P(f_1) = g_1$

$$\|P^{-1}(g_1) - P^{-1}(g_0)\|_n \leqslant C\|g_1 - g_0\|_n + C\big(\|P^{-1}(g_1)\|_{n+2r} + \|P^{-1}(g_0)\|_{n+2r}\big)\|g_1 - g_0\|_0.$$

Combining this with the above estimate

$$\|P^{-1}(g_1) - P^{-1}(g_0)\|_n \leqslant C\|g_1 - g_0\|_n + C(\|g_1\|_{n+2r} + \|g_0\|_{n+2r})\|g_1 - g_0\|_0.$$

This shows that P^{-1} is continuous, and provides a tame estimate for the modulus of continuity. By Corollary 1.3.3 we also have

$$\begin{aligned}\|P^{-1}(g_1) - P^{-1}(g_0) - VP\big(P^{-1}(g_0)\big)(g_1 - g_0)\|_n \\ \leqslant C\|P^{-1}(g_1) - P^{-1}(g_0)\|_{n+2r}\|P^{-1}(g_1) - P^{-1}(g_0)\|_0 \\ + C\|P^{-1}(g_0)\|_{n+2r}\|P^{-1}(g_1) - P^{-1}(g_0)\|_0^2.\end{aligned}$$

Using

$$\|P^{-1}(g_0)\|_{n+2r} \leqslant C\|g_0\|_{n+2r},$$

$$\|P^{-1}(g_1) - P^{-1}(g_0)\|_0 \leqslant C\|g_1 - g_0\|_0,$$

$$\begin{aligned}\|P^{-1}(g_1) - P^{-1}(g_0)\|_{n+2r} \\ \leqslant C\|g_1 - g_0\|_{n+2r} + C(\|g_1\|_{n+4r} + \|g_0\|_{n+4r})\|g_1 - g_0\|_0\end{aligned}$$

we get

$$\begin{aligned}\|P^{-1}(g_1) - P^{-1}(g_0) - VP\big(P^{-1}(g_0)\big)(g_1 - g_0)\|_n \\ \leqslant C\|g_1 - g_0\|_{n+2r}\|g_1 - g_0\|_0 + C(\|g_1\|_{n+4r} + \|g_0\|_{n+4r})\|g_1 - g_0\|_0^2.\end{aligned}$$

Replacing g_0 by g and g_1 by $g + tk$ we get

$$\begin{aligned}\left\|\frac{P^{-1}(g + tk) - P^{-1}(g)}{t} - VP\big(P^{-1}(g)\big)k\right\|_n \\ \leqslant Ct\|k\|_{n+2r}\|k\|_0 + Ct(\|g_1\|_{n+2r} + \|g_0\|_{n+2r})\|k\|_0^2.\end{aligned}$$

Letting $t \to 0$, we see that P^{-1} is differentiable, and

$$DP^{-1}(g)k = VP\big(P^{-1}(g)\big)k.$$

Then since VP and P^{-1} are continuous, so is DP^{-1}, and P^{-1} is C^1. It now follows by the chain rule that P^{-1} is C^∞. Since VP and P^{-1} are tame, DP^{-1} is tame. Then so are all the derivatives D^nP^{-1}, and P^{-1} is a smooth tame map.

When we only know that DP is surjective with a right inverse, we proceed in a less direct fashion.

1.7.1. THEOREM. *The solution $f_\infty = S(g)$ defined as the limit of the algorithm*

$$f_t' = cVP(S_t f_t)S_t[g - P(f_t)]$$

defines a smooth tame map in a neighborhood of the origin $\|g\|_{2r} < \delta$ which is a right inverse for P, so that $P \circ S(g) = g$.

PROOF. First we argue that $S(g)$ depends continuously on g. If we let $S_t(g) = f_t$ be the solution at time t, we know by Corollary I.5.6.6 that $S_t(g)$ depends smoothly on g for all t. By Corollary 1.6.11 the continuous functions $S_t(g)$ converge uniformly to $S(g)$, so $S(g)$ is continuous also.

Next we differentiate the algorithm itself. This shows (by the argument at the end of the proof of Theorem I.5.6.3) that $DS_t(g)k = h_t$ is the solution of the differential equation

$$h_t' = VP(S_t f_t)\{cS_t[k - DP(f_t)h_t - D^2P(S_t f_t)\{f_t', S_t h_t\}]\}.$$

Now we recall the definition of the tangent functor

$$TP(f, h) = (P(f), DP(f)h) = (g, k).$$

Its derivative is

$$DTP(f, h)(\tilde{f}, \tilde{h}) = \big(DP(f)\tilde{f},\ DP(f)\tilde{h} + D^2P(f)\{h, \tilde{f}\}\big) = (\tilde{g}, \tilde{k}).$$

It is easy to check that a right inverse for DTP is given by

$$VTP(f, h)(\tilde{g}, \tilde{k}) = \big(VP(f)\tilde{g}, VP(f)[\tilde{k} - D^2P(f)\{h, VP(f)\tilde{g}\}]\big) = (\tilde{f}, \tilde{h}).$$

We can set up the algorithm for TP the same way we did for P. This gives

$$(f_t', f_t') = cVTP(S_t f_t, S_t h_t)[(S_t g, S_t k) - TP(f_t, h_t)]$$

which reduces to the equations

$$f_t' = cVP(S_t f_t)S_t[g - P(f_t)],$$

$$h_t' = cVP(S_t f_t)[S_t k - DP(f_t)h_t - D^2P(S_t f_t)\{S_t h_t, VP(S_t f_t)[S_t g - P(f_t)]\}]$$

which gives the same equation for f_t' as before. Therefore the tangent TS_t to the solution S_t for P is the solution to the algorithm for the tangent TP. It follows that TS_t converge uniformly as $t \to \infty$. Therefore S is C^1 and its tangent $TS = \lim_{t\to\infty} TS_t$. Moreover TS is also tame. Then by induction on n we get T^nS being continuous and tame for all n. Thus S is a smooth tame map.

III.2. Applications.

2.1. *Embedding surfaces of positive curvature.*

2.1.1. EXAMPLE. Let M be a compact strictly convex smooth surface in R^3. Then M acquires a Riemannian metric of strictly positive curvature. The converse is a classical result due to Nirenberg [**15**].

2.1.2. THEOREM. *Let M be a compact oriented surface with a Riemannian metric of strictly positive curvature. Then M can be isometrically embedded as a smooth convex surface in R^3.*

PROOF. Since the surface is oriented and the metric has positive curvature, the surface has positive Euler class by the Gauss-Bonnet theorem, and hence M is the sphere S^2. Let $\mathcal{F}$ denote the Fréchet manifold of all smooth strictly convex embeddings of M into R^3, and let $\mathcal{G}$ denote the Fréchet manifold of all smooth Riemannian metrics on M of strictly positive curvature. There is an operator P: $\mathcal{F} \to \mathcal{G}$ which assigns to each embedding the induced metric. We shall show that P is surjective. In fact P represents $\mathcal{F}$ as a principal fibre bundle over $\mathcal{G}$ with group being the group of Euclidean motions. To see that P is surjective we will show that the image of P is both open and closed and that $\mathcal{G}$ is connected. Our main interest lies in the first of these three assertions, since it is here we use the inverse function theorem.

We let Greek indices α, β, γ denote tensors on R^3 and Latin indices i, j, k denote tensors on M. We let f^α be the functions on M giving the embedding, $g_{\alpha\beta}$ the Euclidean metric on R^3, g_{ij} the induced metric on M, and ∂_i the induced covariant derivative on M. Then we have

$$g_{ij} = g_{\alpha\beta}\partial_i f^\alpha \partial_j f^\beta.$$

We let $\mu_{\alpha\beta\gamma}$ be the volume form on R^3 and μ_{ij} the induced area form on M. Then the unit outward normal n^θ is given by the formula

$$\mu_{\alpha\beta\gamma}\partial_i f^\alpha \partial_j f^\beta = g_{\alpha\theta} n^\theta \mu_{ij}.$$

If b_{ij} is the second fundamental form of the embedding then we have

$$\partial_i n^\alpha = b_{ij} g^{jk} \partial_k f^\alpha, \qquad \partial_i \partial_j f^\alpha = -b_{ij} n^\alpha$$

by the usual calculations.

The operator $P(f^\alpha) = g_{ij}$ is given by the formula

$$g_{ij} = g_{\alpha\beta}\partial_i f^\alpha \partial_j f^\beta$$

and its derivative $DP(f^\alpha)\tilde{f}^\alpha = \tilde{g}_{ij}$ is given by

$$\tilde{g}_{ij} = g_{\alpha\beta}\partial_i \tilde{f}^\alpha \partial_j f^\beta + g_{\alpha\beta}\partial_i f^\alpha \partial_j \tilde{f}^\beta.$$

We can decompose $\tilde{f}^\alpha$ into its tangential and normal components $\tilde{v}^i$ and $\tilde{w}$. Let

$$\tilde{f}^\alpha = \tilde{v}^i \partial_i f^\alpha + \tilde{w} n^\alpha.$$

Then the linearized equation becomes

$$(*) \qquad \tilde{g}_{ij} = g_{ik}\partial_j \tilde{v}^k + g_{jk}\partial_i \tilde{v}^k + 2b_{ij}\tilde{w}.$$

Observe what this equation says; a tangential motion $\tilde{v}^i$ changes the metric g_{ij} by its Lie derivative in the direction $\tilde{v}^i$, while a normal motion $\tilde{w}$ changes the metric proportionally to the second fundamental form. Since the equation $(*)$ does not involve any derivatives of $\tilde{w}$ we may reduce modulo b_{ij} to obtain the equation

$$(**) \qquad \tilde{g}_{ij} \equiv g_{ik}\partial_j \tilde{v}^k + g_{jk}\partial_i \tilde{v}^k \mod b_{ij}.$$

We can first solve $(**)$ for $\tilde{v}^k$ and then solve algebraically for $\tilde{w}$.

2.1.3. LEMMA. *The equation* (**) *for a strictly convex surface is an elliptic first order system mapping sections* $\tilde{v}^k$ *of the two dimensional tangent bundle* TM *into sections of the two dimensional quotient bundle* $Q = L^2_S(TM)/B$, *where* $L^2_S(TM)$ *is the 3 dimensional bundle of symmetric tensors* $\widetilde{g_{ij}}$ *and* B *is the one dimensional subbundle spanned by* b_{ij}.

PROOF. Let

$$L(\tilde{v})_{ij} \equiv g_{ik}\partial_j \tilde{v}^k + g_{jk}\partial_i \tilde{v}^k \mod b_{ij}$$

be the operator. Then its symbol is

$$\sigma L(\zeta)\tilde{v}_{ij} \equiv g_{ik}\zeta_j \tilde{v}^k + g_{jk}\zeta_i \tilde{v}^k \mod b_{ij}$$

in the direction of a cotangent vector ζ_k. If $\sigma L(\zeta)\tilde{v} = 0$ then we have

$$g_{ik}\zeta_j \tilde{v}^k + g_{jk}\zeta_i \tilde{v}^k \equiv 0 \mod b_{ij}.$$

Choosing a special coordinate system at a point we can make

$$\tilde{v} = \begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \end{pmatrix}, \qquad \zeta = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix},$$

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad b_{ij} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Then our equation $\sigma L(\zeta)\tilde{v} = 0$ becomes

$$\begin{pmatrix} 2\zeta_1\tilde{v}^1 & \zeta_2\tilde{v}^1 + \zeta_1\tilde{v}^2 \\ \zeta_2\tilde{v}^1 + \zeta_1\tilde{v}^2 & 2\zeta_2\tilde{v}^2 \end{pmatrix} \equiv 0 \mod \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

This implies the equations

$$\lambda_2\zeta_1\tilde{v}^1 - \lambda_1\zeta_2\tilde{v}^2 = 0, \qquad \zeta_2\tilde{v}^1 + \zeta_1\tilde{v}^2 = 0.$$

There will be a nonzero solution v only if the determinant is zero. Thus

$$\det\begin{pmatrix} \lambda_2\zeta_1 & -\lambda_1\zeta_2 \\ \zeta_2 & \zeta_1 \end{pmatrix} = \lambda_2\zeta_1^2 + \lambda_1\zeta_2^2 = 0.$$

When the surface is convex we have $\lambda_1 > 0$ and $\lambda_2 > 0$. This forces $\zeta_1 = 0$ and $\zeta_2 = 0$. Hence the equation is elliptic.

2.1.4. LEMMA. *For a strictly convex surface* M *the elliptic operator* L *is always surjective and has null space of dimension* 6.

PROOF. Since the rigid motions of R^3 give elements in the null space of L, it is clear why the null space has dimension at least 6. We claim that the index of the elliptic operator is exactly six. Since any two convex embeddings can be joined by a path of convex embeddings, it suffices to check the index for one example. The standard sphere S^2 is particularly simple, since $b_{ij} = g_{ij}$. Then the equation

$$g_{ik}\partial_j \tilde{v}^k + g_{jk}\partial_i \tilde{v}^k \equiv 0 \mod g_{ij}$$

says that the rate of change of g_{ij} in the direction $\tilde{v}^k$ is proportional to g_{ij}, which means that $\tilde{v}^k$ is an infinitesimal conformal transformation. In complex coordinates on the sphere the condition is that $\tilde{v} = \tilde{v}(z)\partial/\partial z$ be an analytic vector field. If $w = 1/z$ then

$$\partial/\partial z = -w^2\partial/\partial w$$

and hence to represent a finite vector field at infinity, $\tilde{v}(z)$ must grow no faster than z^2, so

$$\tilde{v}(z) = az^2 + bz + c.$$

This gives three complex or six real parameters. To complete the lemma, it suffices to show that L is always surjective. Since L is elliptic, this is equivalent to the assertion that the adjoint operator L^* is injective. The dual bundle to the quotient bundle $Q = L^2_S(TM)/B$ is the subbundle of symmetric tensors h^{ij} with $b_{ij}h^{ij} = 0$. The adjoint operator L^* is given by

$$L^*h^j = 2\partial_i h^{ij}.$$

Then the next theorem shows L^* is injective. Its proof is similar to (and slightly easier than) the classical proof of infinitesimal rigidity for convex surfaces (see Spivak [**23**]).

2.1.5. THEOREM. *Let M be a compact strictly convex surface in R^3 and let h^{ij} be a symmetric tensor on M with*

$$\partial_i h^{ij} = 0 \quad \textit{and} \quad b_{ij}h^{ij} = 0.$$

Then $h^{ij} = 0$.

PROOF. We proceed by a series of lemmas.

2.1.6. LEMMA. *Let $\phi^\alpha_i = \mu_{ik}h^{kl}\partial_l f^\alpha$. Then*

$$\mu^{ij}\partial_j\phi^\alpha_i = 0.$$

PROOF. We have $\mu^{ij}\mu_{ik} = \delta^j_k$ and $\partial_j\partial_l f^\alpha = -b_{jl}n^\alpha$. Therefore

$$\mu^{ij}\partial_j\phi^\alpha_i = \partial_j h^{jl}\partial_l f^\alpha - b_{jl}h^{jl}n^\alpha.$$

2.1.7. LEMMA. *Let $v^i_\alpha = h^{ij}\mu_{\alpha\beta\gamma}f^\beta\partial_j f^\gamma$. Then $\partial_i v^i_\alpha = 0$.*

PROOF. We have

$$\mu_{\alpha\beta\gamma}\partial_i f^\beta\partial_j f^\gamma = \mu_{ij}g_{\alpha\theta}n^\theta, \qquad \partial_i\partial_j f^\gamma = -b_{ij}n^\gamma$$

and so

$$\partial_i v^i_\alpha = \partial_i h^{ij}\mu_{\alpha\beta\gamma}f^\beta\partial_j f^\gamma + \mu_{ij}h^{ij}g_{\alpha\theta}n^\theta - b_{ij}h^{ij}\mu_{\alpha\beta\gamma}f^\beta n^\gamma.$$

2.1.8. LEMMA. *Let s be the support function $s = g_{\alpha\beta}f^\alpha n^\beta$. Then*

$$v^i_\alpha\phi^\alpha_i = 2s\det(h^{ij}).$$

PROOF. We have

$$\mu_{\alpha\beta\gamma}\partial_j f^\gamma \partial_l f^\alpha = \mu_{jl} g_{\beta\theta} n^\theta$$

which makes

$$v^i_\alpha \phi^\alpha_i = g_{\beta\theta} f^\beta n^\theta \cdot \mu_{ik}\mu_{jl} h^{ij} h^{kl}$$

and it is easy to see that

$$\mu_{ik}\mu_{jl} h^{ij} h^{kl} = 2 \det(h^{ij}).$$

2.1.9. LEMMA. *If $b_{ij}h^{ij} = 0$ and b_{ij} is strictly positive-definite then* $\det(h^{ij}) \leqslant 0$, *and equality holds only when $h^{ij} = 0$.*

PROOF. We can simultaneously diagonalize b_{ij} and h^{ij}. Since the eigenvalues of b_{ij} are positive, those of h^{ij} must have opposite signs, which makes the determinant negative. If one eigenvalue of h^{ij} is zero, so must be the other.

LEMMA. *If the origin lies inside M, the support function s is strictly positive.*

PROOF. The origin lies on the inside of the tangent plane and the normal on the outside.

2.1.10. LEMMA. *If $\partial_i v^i = 0$ and $\mu^{ij}\partial_j \phi_i = 0$ then*

$$\int_M v^i \phi_i \, d\mu = 0.$$

PROOF. If $\mu^{ij}\partial_j \phi_i = 0$ then $\partial_j \phi_i = \partial_i \phi_j$. Consequently regarding ϕ as a 1-form we have $d\phi = 0$. Since $H^1(M, R) = 0$ for a convex surface, $\phi = d\psi$ for some function ψ, so $\phi_i = \partial_i \psi$. Then by Green's theorem

$$\int_M v^i \phi_i \, d\mu = \int v^i \partial_i \psi \, d\mu = -\int \partial_i v^i \cdot \psi \, d\mu = 0.$$

2.1.11. COROLLARY. *For the v^i_α and ϕ^α_i above we have*

$$\int_M v^i_\alpha \phi^\alpha_i \, d\mu = 0.$$

PROOF. Sum over α. We see now that

$$\int_M s \det(h^{ij}) \, d\mu = 0.$$

Since $s > 0$ and $\det(h^{ij}) \leqslant 0$ we must have $\det(h^{ij}) = 0$ everywhere. Thus $h^{ij} = 0$.

It follows that for each convex embedding f^α the operator $DP(f^\alpha)\tilde{f}^\alpha = \tilde{g}_{ij}$ is surjective with a six dimensional null space. We can find a right inverse $\tilde{f}^\alpha = VP(f^\alpha)\tilde{g}_{ij}$. Since it is obtained by solving the elliptic operator L whose coefficients depend smoothly on f^α, we know that our right inverse VP is a smooth tame operator. Then by the Nash-Moser inverse function theorem the operator P is locally surjective, and the image of P is open. If we wanted, we could normalize six parameters of f^α (for example, the image of a point, the

direction of the tangent plane at the point, and the direction of a curve through the point in the tangent plane) to make P invertible. This shows P is locally a bundle of fibre dimension 6. We shall briefly sketch the rest of the proof. To show that $\mathcal{G}$ is connected we first appeal to the uniformization theorem, which shows that any two metrics on M are conformally equivalent, up to a diffeomorphism. Next we observe by a lengthy calculation that if g and $e^{\phi}g$ are conformally equivalent metrics of strictly positive curvature, then $e^{t\phi}g$ also has strictly positive curvature for $0 \leqslant t \leqslant 1$. Since the orientation-preserving diffeomorphisms of S^2 are connected (and indeed have the homotopy type of the rotation subgroup) this shows how to connect any two metrics of strictly positive curvature by a path of such metrics.

To complete the proof we must show that the image of P is closed. Since translation does not change the induced metric, let $\mathcal{F}_0$ denote the closed subset of convex embeddings in $\mathcal{F}$ whose center of mass lies at the origin. Then $P(\mathcal{F}_0) = P(\mathcal{F})$. We claim that P: $\mathcal{F}_0 \to \mathcal{G}$ is a proper map, which is to say that the inverse image of every compact set is compact. From this it follows easily that $P(\mathcal{F}_0)$ is closed in $\mathcal{G}$. To see that P: $\mathcal{F}_0 \to \mathcal{G}$ is a proper map, we must establish a priori estimates for embeddings f^{α} with center of mass at the origin in terms of the induced metrics. We must estimate the size of the derivatives of the f^{α} using constants which depend only on the size of the derivatives of the induced metric g_{ij} and a lower bound for its curvature. Since $\mathcal{C}^{\infty}(M)$ is a Montel space, every bounded set is compact. In making our estimates on the derivatives of the f^{α} we may use the induced metric g_{ij} and its connection ∂_i, for if we take various metrics in a compact subset of $\mathcal{G}$ they are all equivalent, and the various connections ∂_i differ only by lower order terms involving derivatives of the g_{ij}, which are all bounded in a compact subset of $\mathcal{G}$.

Suppose then that we have a collection of embeddings f^{α} with center of mass at the origin whose induced metrics lie in a compact subset of $\mathcal{G}$, and hence have all their derivatives uniformly bounded and their scalar curvatures bounded uniformly away from zero. A positive lower bound on the scalar curvature gives a positive upper bound on the diameter of a surface. Hence the f^{α} are all uniformly bounded. Since

$$g^{ij}g_{\alpha\beta}\partial_i f^{\alpha}\partial_j f^{\beta} = 2$$

the first derivatives $\partial_i f^{\alpha}$ are also uniformly bounded. The next step involves a very clever estimate due to Weyl. Let $b = \operatorname{trace} b_{ij}$ be the mean curvature of the embedding and $k = \det b_{ij}$ the scalar curvature. Note that by Gauss we know k from g_{ij} without knowing the embedding.

2.1.12. THEOREM (WEYL). *For any compact convex surface we have the estimate*

$$\max_M b^2 \leqslant \max_M 4k - \frac{\Delta k}{k}.$$

PROOF. A lengthy computation shows that b satisfies the equation

$$(bg_{ij} - b_{ij})\partial_i\partial_j b = \Delta k + |\partial_i b_{jk}|^2 - |\partial_i b|^2 + k(b^2 - 4k).$$

At the point where b assumes its maximum we have $\partial_i b = 0$ and $\partial_i \partial_j b$ is a negative matrix. If M is convex then b_{ij} is a positive matrix, and so is $bg_{ij} - b_{ij}$. Also $|\partial_i b_{jk}|^2 \geqslant 0$. Hence at this point

$$\Delta k + k(b^2 - 4k) \leqslant 0, \qquad b^2 \leqslant 4k - \Delta k/k$$

and the theorem follows.

This theorem shows that the mean curvatures b are uniformly bounded. Since

$$|b_{ij}|^2 = b^2 - 2k$$

the second fundamental forms b_{ij} are also uniformly bounded. Then since

$$\partial_i \partial_j f^\alpha = -b_{ij} n^\alpha$$

and the n^α are unit vectors, the second derivatives $\partial_i \partial_j f^\alpha$ are also uniformly bounded. One is now in a position to use standard results about elliptic equations to estimate the higher derivatives as well. This completes the proof.

2.2. *Shallow water equations.*

2.2.1. EXAMPLE. The shallow water equations describe the motion of an incompressible inviscous fluid where the wave-length is long compared to the depth, so that the vertical component and variation of the velocity of the fluid may be neglected. This approximation is very good for tidal waves. To simplify the discussion we assume that the earth is a perfect sphere covered with water and neglect the rotation of the earth. We choose our units of length, time and mass so that the depth of the water at rest is 1, the acceleration due to gravity is 1, and the density of the water is 1. We let g_{ij} be the Riemannian metric on the earth sphere S and ∂_i the induced covariant derivative. The motion of the water is described by a tangent vector field $v = v^i \partial/\partial x^i$ on S giving the horizontal velocity (independent of the height) and a function h on S giving the height of the water. The shallow water equations for v and h are

$$(*) \qquad \begin{cases} \partial v^i/\partial t + v^j \partial_j v^i + g^{ij} \partial_j h = 0, \\ \partial h/\partial t + v^j \partial_j h + h \partial_j v^j = 0. \end{cases}$$

The first equation is Newton's law that force equals mass times acceleration. Since the density is one the force equals the acceleration. The term

$$\partial v^i/\partial t + v^j \partial_j v^i$$

is the acceleration of a particle moving in the water, which is the sum of the acceleration due to the change in the velocity at the point with time and acceleration due to the change in velocity when the particle moves to a region of different velocity. The force on the particle is

$$g^{ij} \partial_j h$$

which is the gradient of the function h. If the particle is on the surface it must gain potential energy to climb to a greater height; while if the particle is on the bottom it gains potential energy from an increase in the pressure, and the pressure equals the height. The force is the gradient of the potential energy.

The second equation is conservation of mass. The mass per unit area is the height, so the rate of increase of mass per unit area with time is $\partial h/\partial t$. On the other hand, the transport of mass is given by hv^j, so the rate at which mass flows out per unit area is given by its divergence

$$\partial_j(hv^j) = v^j\partial_j h + h\partial_j v^j$$

and the equation follows. Note the equations make sense on any Riemannian manifold. We shall prove the following results.

2.1.2. THEOREM. *The shallow water equations* (*) *admit a solution on any Riemannian manifold* (A) *with any initial conditions* v *and* $h > 0$ *at time* $t = 0$ *for a short time* $0 \leqslant t \leqslant \varepsilon$ (*where* ε *depends on the initial conditions*), *and* (B) *for a long time* $0 \leqslant t \leqslant T$ *with initial conditions* v *near* 0 *and* h *near* 1 (*how close depending on* T).

PROOF. First we modify the equations by replacing h by $k = 2\sqrt{h}$. This is reasonable since we want $h > 0$ on physical grounds anyway. (When $h = 0$ the bare earth is exposed, giving rise to an interesting free boundary problem which we shall not treat.) We also multiply the first equation by g_{ip} to lower the index. This casts our equations in the form

$$(**) \qquad \begin{cases} g_{ip}\left(\partial v^i/\partial t + v^j\partial_j v^i\right) + \frac{1}{2}k\partial_p k = 0, \\ \frac{1}{2}k\partial_j v^j + \left(\partial k/\partial t + v^j\partial_j k\right) = 0. \end{cases}$$

If our Riemannian manifold is S we let $M = S \times I$ where I is the time interval $0 \leqslant t \leqslant T$. Note that the tangent bundle TS over S pulls back to a bundle which we also call TS over M, while R can be regarded as a trivial bundle over S or M. We let $B = TS \times R$ be the product bundle. A section of B consists of a vector field $v^i \in TS$ and a function $k \in R$. The dual bundle $B^* = T^*S \times R$, and a section of B^* consists of a 1-form $w_p \in T^*S$ and a function $l \in R$.

Define an operator

$$P\colon \mathcal{C}^\infty(M, B) \to \mathcal{C}^\infty(M, B^*) \times \mathcal{C}^\infty(S, B)$$

by letting

$$P(v^i, k) = \left(w_p, l, v_0^i, k_0\right)$$

where $v_0^i = v^i \mid \{t = 0\}$ and $k_0 = k \mid \{t = 0\}$ and

$$g_{ip}\left(\partial v^i/\partial t + v^j\partial_j v^i\right) + \tfrac{1}{2}k\partial_p k = w_p,$$
$$\tfrac{1}{2}k\partial_j v^j + \left(\partial k/\partial t + v^j\partial_j k\right) = l.$$

We can compute the derivative of the operator P

$$DP(v^i, k)(\tilde{v}, \tilde{k}) = \left(\tilde{w}_p, \tilde{l}, \tilde{v}_0^i, \tilde{k}_0\right);$$

it is given by the equations

$$(***)\qquad \begin{cases} g_{ip}\left(\dfrac{\partial \tilde{v}^i}{\partial t} + v^j\partial_j\tilde{v}^i + \tilde{v}^j\partial_j v^i\right) + \dfrac{1}{2}k\partial_p\tilde{k} + \dfrac{1}{2}\tilde{k}\partial_p k = \tilde{w}_p, \\ \dfrac{1}{2}k\partial_j\tilde{v}^j + \dfrac{1}{2}\tilde{k}\partial_j v^j + \left(\dfrac{\partial\tilde{k}}{\partial t} + v^j\partial_j\tilde{k} + \tilde{v}^j\partial_j k\right) = \tilde{l} \end{cases}$$

with $\tilde{v}^i \mid \{t = 0\} = \tilde{v}_0^i$ and $\tilde{k} \mid \{t = 0\} = \tilde{k}_0$. Now we claim that $DP(v^i, k)$ is a linear symmetric system in $(\tilde{v}^i, \tilde{k})$. To check this we evaluate the symbol of DP in the direction of a cotangent vector ξ by replacing ∂_j by ξ and $\partial/\partial t$ by τ, and dropping the zero-order terms. This gives

$$\sigma DP(v, k)(\tau, \xi)(\tilde{v}, \tilde{k}) = (\tilde{w}, \tilde{l}),$$
$$\left(\tau + v^j\xi_j\right)g_{ip}\tilde{v}^i + \tfrac{1}{2}k\xi_p\tilde{k} = \tilde{w}_p,$$
$$\tfrac{1}{2}k\xi_j\tilde{v}^j + \left(\tau + v^j\xi_j\right)\tilde{k} = \tilde{l}$$

which shows the matrix of the symbol is

$$DP(v, k)(\tau, \xi) = \begin{pmatrix} [\tau + \langle v, \xi\rangle]I & \tfrac{1}{2}k\xi \\ \tfrac{1}{2}k\xi & \tau + \langle v, \xi\rangle \end{pmatrix}$$

which is symmetric. The function t will be our weight function. The covector dt has coordinates $\tau = 1$, $\xi = 0$ for which the symbol $\sigma DP(v, k)(1, 0) = I$ is the identity, so t is a positive weight function. Moreover dt is a normal covector pointing inward on $S \times 0$ and outward on $S \times 1$. Thus the boundary is nowhere characteristic and $\partial^+ M = S \times 0$ and $\partial^- M = S \times 1$. The theory of symmetric systems shows that $DP(v, k)$ is always invertible and has a smooth tame inverse. Then the Nash-Moser inverse function theorem implies that P is locally invertible in a neighborhood of any v and k.

To prove part (B) of the theorem, we use the inverse of P in a neighborhood of $v = 0$ and $k = 2$. When v_0 is close to 0 and k_0 is close to 2, and $w = 0$ and $l = 0$, we get a unique solution v and k of

$$P(v, k) = (0, 0, v_0, k_0)$$

which gives a solution of the shallow water equations for time $0 \leqslant t \leqslant T$. Note that v_0 and k_0 must get closer to 0 and 2 as T gets large.

To prove part (A) we use a clever device. The shallow water equations are formally solvable at $t = 0$. This means that for given initial data $v = v_0$ and $k = k_0$ at $t = 0$, we can solve for all the derivatives $\partial^n v/\partial t^n$ and $\partial^n k/\partial t^n$ at $t = 0$ by differentiating through the equation. We can then find smooth functions $\bar{v}$ and $\bar{k}$ on M which have the right time derivatives at $t = 0$. Let us define $\bar{w}$ and $\bar{l}$ by letting

$$P(\bar{v}, \bar{k}) = (\bar{w}, \bar{l}, v_0, k_0).$$

By our choice of $\bar{v}$ and $\bar{k}$, the functions $\bar{w}$ and $\bar{l}$ and all their derivatives vanish at $t = 0$. Extend $\bar{w}$ and $\bar{l}$ to vanish for $t \leqslant 0$. Then perform a small shift, so that

$$w(t) = \bar{w}(t - \varepsilon) \quad \text{and} \quad l(t) = \bar{l}(t - \varepsilon).$$

Then w and l vanish for $0 \leqslant t \leqslant \varepsilon$. Since P is invertible in a neighborhood of $(\bar{v}, \bar{k})$, we can solve the equation

$$P(v, k) = (w, l, v_0, k_0)$$

when $\varepsilon > 0$ is sufficiently small. Then v and k will solve the shallow water equations for a short time $0 \leqslant t \leqslant \varepsilon$. This completes the proof.

2.3. *Submanifolds of fixed volume.*

2.3.1. THEOREM. *Let F be a tame space and P: $U \subseteq F \to V$ a smooth map of an open set in F to a finite dimensional vector space. Suppose that for some $f_0 \in U$ the derivative $DP(f_0)$ is surjective. Then the level set $N = \{f \in U: P(f) = P(f_0)\}$ is a smooth tame submanifold in a neighborhood of f_0 with tangent space*

$$T_f N = \text{Null } DP(f).$$

PROOF. We may assume that $f_0 = 0$, that $F = G \times V$ and that the derivative of P at 0 is the projection on V. Define a map Q on a neighborhood of zero in $G \times V$ to itself by

$$Q(g, u) = (g, P(g, u)).$$

Then Q is a smooth tame map since the first factor is the identity and the second is finite dimensional. Moreover

$$DQ(g, u)(h, v) = (h, D_u P(g, u)v) = (h, w)$$

has solution

$$VQ(g, u)(h, w) = (h, v)$$

where

$$h = D_u P(g, u)^{-1} w.$$

Now $D_u P(0, 0)$ is the identity, and $D_u P(g, u)$ is a matrix in $L(V, V)$ whose entries depend smoothly on g and u. Hence the same is true of the inverse matrix. Thus h is a smooth tame function of g, u and w (since any map to a finite dimensional space is tame). By the Nash-Moser theorem, Q has a local smooth tame inverse. Since $Q^{-1}(G \times 0) = N$ we are done.

Let X be a finite dimensional Riemannian manifold. If S is a submanifold of X, we can define on S the mean curvature vector

$$k(S) \in \mathcal{C}^\infty(S, TX/TS)$$

with values in the normal bundle, the measure $d\mu(S)$, and the total measure

$$\mu(S) = \int_S d\mu(S).$$

Let $\mathcal{S}(X)$ be the smooth tame Fréchet manifold of all compact submanifolds S, and let $\mathcal{S}_r(X)$ be the closed subset of those with total measure r.

2.3.2. THEOREM. *If $S \in \mathcal{S}_r(X)$ and $k(S)$ is not identically zero, then $\mathcal{S}_r(X)$ is a smooth tame submanifold of codimension* 1 *in a neighborhood of S, with tangent space*

$$T_S \mathcal{S}_r(X) = \left\{ h \in \mathcal{C}^\infty(S, TX/TS): \int_S \langle k(S), h \rangle \, d\mu(S) = 0 \right\}.$$

REMARK. If S is a geodesic curve in X of length r then $\mathcal{S}_r(X)$ may not be a submanifold of codimension 1 near S. For a minimal isolated geodesic is an isolated point in $\mathcal{S}_r(X)$.

PROOF. If $S \in \mathcal{S}(X)$ and

$$h \in T_S\mathcal{S}(X) = \mathcal{C}^\infty(S, TX/TS)$$

then by differential geometry

$$D\mu(S)h = \int_S \langle k(S), h \rangle \, d\mu(S).$$

Pick a reference S_0 in $\mathcal{S}_r(X)$ and choose a diffeomorphism of a tubular neighborhood of the zero section in the normal bundle of S_0 into X. If f is a section of the normal bundle, let $S(f)$ be the corresponding submanifold. This gives a local chart on $\mathcal{S}(X)$ near S_0. Let $k(f)$ be the appropriate pull-back of the mean curvature $k(S(f))$ to S_0 under f. Then in our chart if $\mu(f) = \mu(S(f))$

$$D\mu(f)h = \int_{S_0} \langle k(f), h \rangle \, d\mu(S_0).$$

The mean curvature $k(f)$ is given by a nonlinear differential operator in f of degree 2, with $k(0) = k(S_0)$ not identically zero. Let F be the subspace of all sections f of the normal bundle to S_0 which are orthogonal to $k(0)$. Define a map P of a neighborhood of zero in $F \times R$ into $F \times R$ by letting

$$P(f, s) = (f, \mu(f + sk(0))).$$

Then P is a smooth tame map with derivative

$$DP(f, s)(h, t) = (h, D\mu(f + sk(0))(h + tk(0)))$$

and we have

$$D\mu(f + sk(0))(h + tk(0)) = \int_{S_0} \langle k(f + sk(0)), h + tk(0) \rangle \, d\mu(S_0)$$

which we can solve for t, provided

$$\int_S \langle k(f + sk(0)), k(0) \rangle \, d\mu(S_0) \neq 0$$

which happens for all f in a C^2 neighborhood of zero. Therefore P gives a smooth change of coordinate near S_0 making μ into one of the coordinate functions. This shows $\mathcal{S}_r(X) = \mu^{-1}(r)$ is a smooth submanifold with tangent space Null $D\mu$.

2.4. *Symplectic and contact structures.* The following result is useful in showing that a group acts transitively on a manifold. Let $\mathcal{G}$ be a tame Fréchet Lie group acting tamely on a tame Fréchet manifold $\mathcal{M}$ with action

$$A: \mathcal{G} \times \mathcal{M} \to \mathcal{M}.$$

For any $m \in \mathcal{M}$ there is a linear map

$$A'_m: G \to T_m\mathcal{M}$$

of the Lie algebra $G = T_1\mathcal{G}$ into the tangent space of $\mathcal{M}$ at m given by $A'_m = D_{\mathcal{G}}A(1, m)$, the partial derivative with respect to $\mathcal{G}$ of the action at the identity.

2.4.1. THEOREM. *Suppose $\mathcal{M}$ is connected, and for each $m \in \mathcal{M}$ the map A'_m: $G \to T_m\mathcal{M}$ is surjective with a tame linear right inverse. Then $\mathcal{G}$ acts transitively on $\mathcal{M}$.*

PROOF. Consider the map

$$P: \mathcal{G} \to \mathcal{M}, \qquad P(g) = A(g, m).$$

If L_g denotes left multiplication by g on $\mathcal{G}$ and A_g denotes multiplication by g on $\mathcal{M}$, there is a commutative diagram

$$\begin{array}{ccc} \mathcal{G} & \overset{L_g}{\to} & \mathcal{G} \\ P\downarrow & & P\downarrow \\ \mathcal{M} & \overset{A_g}{\to} & \mathcal{M} \end{array}$$

which shows that

$$DA_g \cdot DP(1) = DP(g) \cdot DL_g.$$

We can define a smooth tame vector bundle map VP: $T\mathcal{M} \to T\mathcal{G}$ which is a right inverse for DP by letting

$$VP(g) = DL_g \circ VP(1) \circ DA_{g^{-1}}$$

where $VP(1)$ is a tame linear right inverse for $DP(1) = D_{\mathcal{G}}A(1, m)$, which exists by hypothesis. Then it follows that P is locally surjective; hence $\mathcal{G}$ locally is transitive. If $\mathcal{M}$ is connected it follows that $\mathcal{G}$ is transitive on all of $\mathcal{M}$.

2.4.2. EXAMPLE. Let X be a compact manifold of even dimension $2n$. A symplectic form on X is a closed differential 2-form $\omega \in Z^2(X) \subseteq \wedge^2(X)$ such that

$$\omega^n = \underbrace{\omega \wedge \omega \wedge \cdots \wedge \omega}_{n \text{ times}} \neq 0$$

or equivalently $\omega = \omega_{ij}\, dx^i \wedge dx^j$ in local coordinate where ω_{ij} is an invertible matrix. If such a form exists then $H^2(X) \neq 0$. All Kähler manifolds admit symplectic forms. A diffeomorphism of X takes a symplectic form into another symplectic form in the same cohomology class. Locally this is the only restriction, as the following result shows.

2.4.3. THEOREM. *Let ω be a symplectic form on a compact manifold. Then any closed form near ω in the same cohomology class is also symplectic and is conjugate to ω by a small diffeomorphism.*

PROOF. A closed form close to a symplectic form is still symplectic. Let h be the cohomology class of ω in $H^2(X)$. Then h determines an affine subspace $B^2(X) + h$ in the linear subspace $Z^2(X)$ of closed forms of all those differing from ω by an exact form in $B^2(X)$. The exterior derivative d: $\wedge^1(X) \to B^2(X)$

is surjective and has a linear right inverse given by a Green's operator G: $B^2(X) \to \wedge^1(X)$ from Hodge theory. Since this is derived from the Green's operator of the elliptic system $\Delta = dd^* + d^*d$ it is clear that G is a tame linear map. We let our group be the diffeomorphism group $\mathcal{D}(X)$ of the manifold X, and we let it act on the affine space $B^2(X) + h$ of closed 2-forms in the cohomology class h. This is obtained by restricting the action on all of $\wedge^2(X)$. That action

$$A: \mathcal{D}(X) \times \wedge^2(X) \to \wedge^2(X)$$

has partial derivative with respect to the first factor at the identity $1 \in \mathcal{D}(X)$ in the direction of a tangent vector $v \in \mathcal{C}^\infty(TX) = T_1\mathcal{D}(X)$ given by

$$A'_\omega v = D_1 A(1, \omega)v = L_v\omega$$

where $L_v\omega$ is the Lie derivative of ω in the direction v given in local coordinates by

$$L_v\omega_{ij} = \frac{\partial \omega_{ij}}{\partial x^k} v^k + \omega_{kj}\frac{\partial v^k}{\partial x^i} + \omega_{ik}\frac{\partial v^k}{\partial x^j}.$$

When ω is closed so that $d\omega = 0$ or

$$\partial\omega_{ij}/\partial x^k + \partial\omega_{jk}/\partial x^i + \partial\omega_{hi}/\partial x^j = 0$$

then we see that

$$L_v\omega = d(v \perp \omega)$$

where $v \perp \omega$ is the contraction

$$(v \perp \omega)_j = v^i\omega_{ij}.$$

Note that the tangent space to the affine subspace $B^2(X) + h$ is just the linear subspace $B^2(X)$. We must show that the partial derivative of the action with respect to the first variable at the identity

$$A'_\omega: \mathcal{C}^\infty(TX) \to B^2(X)$$

given by $v \to d(v \perp \omega)$ is surjective with a tame right inverse. When ω is symplectic the equation $v \perp \omega = \alpha$ for any 1-form α has a unique solution $v = \alpha \perp \omega^{-1}$ where

$$(\alpha \perp \omega^{-1})^j = \alpha_i\omega^{-1ij}$$

and ω^{-1ij} is the matrix inverse to ω_{ij}. Then to solve $d(v \perp \omega) = \psi$ we take $v = G\psi \perp \omega^{-1}$ where G is the Green's operator described before. It is easy to see that this is a tame linear map. It now follows from the previous theorem that the identity component of the diffeomorphism group acts transitively on the component of the symplectic forms in the cohomology class h containing ω.

2.4.4. EXAMPLE. Let X be a compact manifold of dimension $n + 1$ and let B be a subbundle of the tangent bundle of codimension 1. There is a natural antisymmetric bilinear map

$$\lambda_B: B \times B \to TX/B$$

defined by letting

$$\lambda_B(u, v) = [u, v]/B$$

where $u, v \in B \subseteq TX$ are vector fields and $[u, v]$ is their Lie bracket.

2.4.5. LEMMA. *The map λ_B is a bilinear vector bundle map, i.e. $\lambda_B(u, v)$ does not involve derivatives of u and v.*

PROOF. We compute in local coordinates $x^1, \ldots, x^n$, y. Suppose that B is the subbundle of TX where

$$dy = b_i(x, y)\, dx^i$$

for some functions $b_i(x, y)$ defined in a local coordinate chart. If u and v lie in B then

$$u = u^i \partial/\partial x^i + u^i b_i \partial/\partial y, \qquad v = v^i \partial/\partial x^i + v^i b_i \partial/\partial y.$$

We now evaluate the commutator at a point where $b_i = 0$ and reduce mod B by ignoring the $\partial/\partial x^i$ components; this gives

$$[u, v] \equiv \left(\frac{\partial b_j}{\partial x^i} - \frac{\partial b_i}{\partial x^j} \right) u^i v^j \frac{\partial}{\partial y} \mod B$$

so that $\lambda(u, v) \equiv \lambda_{ij} u^i v^j \partial/\partial y \mod B$ where

$$\lambda_{ij} = \partial b_j/\partial x^i - \partial b_i/\partial x^j.$$

A contact structure is a subbundle B of TX of codimension 1 for which the map λ_B is nondegenerate, so that λ_{ij} is invertible. This can only happen if the dimension of X is odd, so that n is even, since λ_B is antisymmetric. The group of diffeomorphisms $\mathcal{D}(X)$ acts on the Fréchet manifold $\mathcal{B}$ of all contact structures B on X.

2.4.6. THEOREM. *If B is a contact structure on a compact manifold X, then any contact structure near B is conjugate to B by a diffeomorphism near the identity. The identity component of $\mathcal{D}(X)$ acts transitively on each component of $\mathcal{B}$.*

PROOF. The manifold $\mathcal{B}$ of all contact structures is an open subset of the Grassmannian manifold of all subbundles of TX of codimension 1. We can identify its tangent space as

$$T_B\mathcal{B} = \mathcal{C}^\infty(X, L(B, TX/B)),$$

the space of sections of the bundle of linear maps of B into TX/B. The action

$$A: \mathcal{D}(X) \times \mathcal{B} \to \mathcal{B}$$

has derivative $A'_B = D_1 A(1, B)$ which is a linear map

$$A'_B = \mathcal{C}^\infty(X, TX) \to \mathcal{C}^\infty(X, L(B, TX/B)).$$

In general A'_B will be a differential operator of degree 1 from TX to $L(B, TX/B)$, but on the subbundle $B \subseteq TX$ it has degree 0 and is given by λ_B, as the following lemma shows.

2.4.7. LEMMA. *If $v \in B$ and $w \in B$ then*

$$[A'_B(v)](w) = \lambda_B(w, v).$$

PROOF. If $v \in B$ then locally

$$v = v^i \partial/\partial x^i + b_i v^i \partial/\partial y.$$

We regard v as infinitesimally small and consider the diffeomorphism

$$\bar{x}^i = x^i + v^i, \qquad \bar{y} = y + b_i v^i.$$

Let $\bar{B}$ be the new bundle induced from B by this diffeomorphism. Then B is given by $dy = b_i\, dx^i$ and $\bar{B}$ is given by $d\bar{y} = \bar{b}_i\, dx^i$ where $\bar{b}_i = b_i + a_i$ for some infinitesimal change a_i in b_i. We proceed to expand the equation

$$d\bar{y} = \bar{b}_i(\bar{x}, \bar{y})\, dx^i$$

by replacing $\bar{x}^i$ by $x^i + v^i$, $\bar{y}$ by $y + b_i v^i$, and $\bar{b}_i$ by $b_i + a_i$, and throwing away quadratic terms in the infinitesimal quantities v^i and a_i. This gives

$$a_j = \frac{\partial b_i}{\partial x^j} v^i + \frac{\partial b_i}{\partial y} v^i b_j - \frac{\partial b_j}{\partial x^i} v^i - \frac{\partial b_j}{\partial y} b_i v^i.$$

Evaluating at a point where $b_i = 0$ and using our previous formula for λ_{ij} gives $a_i = \lambda_{ij} v^j$. Since $A'_B(v) = a$ we have

$$A'_B(v)(w) = \lambda_{ij} w^i v^j = \lambda_B(w, v)$$

which proves the lemma.

Now it is easy to see that if B is a contact structure then A'_B is surjective, and in fact the restriction of A'_B to sections of B is an isomorphism. For on $\mathcal{C}^\infty(B)$ the map A' is just induced by the linear vector bundle map $B \to L(B, TX/B)$ given by λ_B, and when λ_B is nondegenerate this is a vector bundle isomorphism. This proves the theorem.

2.5. *Volume preserving diffeomorphisms.* We can use the inverse function theorem to give a characterization of principal bundles. Let $\mathcal{G}$ be a tame Fréchet Lie group acting tamely on a tame Fréchet manifold $\mathfrak{M}$, and let P: $\mathfrak{M} \to \mathfrak{B}$ be a smooth tame map onto another tame Fréchet manifold $\mathfrak{B}$. We assume that $\mathcal{G}$ takes each fibre $P^{-1}(b)$ for $b \in \mathfrak{B}$ into itself and acts transitively and fixed-point free on the fibres. Let G be the Lie algebra of $\mathcal{G}$. Then for each $m \in \mathfrak{M}$ the action A of $\mathcal{G}$ on $\mathfrak{M}$ defines a map

$$A: \mathcal{G} \times \mathfrak{M} \to \mathfrak{M}$$

whose partial derivative defines a map

$$D_{\mathcal{G}} A(1, m): G = T_1\mathcal{G} \to T_m\mathfrak{M}.$$

If $G \times \mathfrak{M}$ is the trivial vector bundle over $\mathfrak{M}$ with fibre $\mathcal{G}$, then $D_{\mathcal{G}} A$ defines a linear vector bundle map of $\mathcal{G} \times \mathfrak{M}$ into $T\mathfrak{M}$. Likewise the derivative of P defines a map

$$DP(m): T_m\mathfrak{M} \to T_{P(m)}\mathfrak{B}$$

which induces a linear vector bundle map of $T\mathfrak{M}$ into the pull-back bundle $P^*T\mathfrak{B}$. This gives a sequence of tame linear vector bundle maps over $\mathfrak{M}$

$$0 \to G \times \mathfrak{M} \overset{D_{\mathcal{G}}A}{\to} T\mathfrak{M} \overset{DP}{\to} P^*T\mathfrak{B} \to 0.$$

2.5.1. THEOREM. *Suppose the above sequence is exact, and admits a smooth tame linear splitting over* $\mathfrak{M}$

$$0 \to P^*T\mathfrak{B} \overset{VP}{\to} T\mathfrak{M} \overset{V_{\mathcal{G}}A}{\to} G \times \mathfrak{M} \to 0$$

such that

$$V_{\mathcal{G}}A \cdot D_{\mathcal{G}}A = I, \qquad DP \cdot VP = I,$$
$$D_{\mathcal{G}}A \cdot V_{\mathcal{G}}A + VP \cdot DP = I.$$

Then $\mathfrak{M}$ *is a principal* $\mathcal{G}$*-bundle over* $\mathfrak{B}$ *with projection* P *and action* A.

PROOF. Fix $\bar{b} \in \mathfrak{B}$. In a neighborhood U of $\bar{b}$ we can define a smooth section

$$S\colon U \subseteq \mathfrak{B} \to \mathfrak{M}$$

with $PS = I$, since DP is surjective with a tame right inverse VP. We define a map

$$Q\colon \mathcal{G} \times (U \times \mathfrak{B}) \to \mathfrak{M}, \qquad Q(g, b) = A(g, S(b)).$$

We claim (by making U smaller if necessary) that Q is a diffeomorphism. Since Q gives a chart in which $\mathcal{G}$ acts by multiplication on the first factor and P acts by projection on the second, this will complete the proof. Since $\mathcal{G}$ takes fibres into themselves and acts transitively and fixed-point free on the fibres, we see that Q is an invertible map of $\mathcal{G} \times (U \subseteq \mathfrak{B})$ onto $P^{-1}(U) \subseteq \mathfrak{M}$. It remains to show that Q is a diffeomorphism. Since $\mathcal{G}$ acts transitively on the fibres, it is enough to prove this in a neighborhood of $(1, \bar{b})$.

To do this we first compute the derivative of Q. Letting $\tilde{g}$ be a tangent vector to $\mathcal{G}$ at $\mathfrak{g}$ and $\tilde{b}$ a tangent vector to $\mathfrak{B}$ at b we have

$$DQ(g, b)(\tilde{g}, \tilde{b}) = D_{\mathcal{G}}A(g, S(b))\tilde{g} + D_{\mathfrak{M}}A(g, S(b))DS(b)\tilde{b}.$$

Let $\tilde{m}$ be a tangent vector to $\mathfrak{M}$ at $m = Q(g, b)$, and let us solve the equation

$$DQ(g, b)(\tilde{g}, \tilde{b}) = \tilde{m}.$$

Since $PA(g, m) = P(m)$ we have

$$DP \cdot D_{\mathcal{G}}A = 0, \qquad DP \cdot D_{\mathfrak{M}}A = DP.$$

Therefore applying DP to the equation and remembering that $P \circ S = I$ so $DP \circ DS = I$ we have

$$\tilde{b} = DP(m)\tilde{m}$$

which solves for $\tilde{b}$. Then for $\tilde{g}$ we have the equation

$$D_{\mathcal{G}}A(g, S(b))\tilde{g} = \tilde{m} - D_{\mathfrak{M}}A(g, S(b))DS(b)\tilde{b}.$$

Since the term on the right lies in the null space of DP by our choice of $\tilde{b}$, and the sequence is exact, there exists a unique solution which is given tamely in terms of $\tilde{m}$, g and b by taking

$$\tilde{g} = V_{\mathcal{G}}A(m)\big[\tilde{m} - D_{\mathcal{M}}A(g, S(b))DS(b)\tilde{b}\big].$$

Since DQ is invertible with a tame inverse, so is Q in a neighborhood of $(1, \bar{b})$. This proves the theorem.

2.5.2. COROLLARY. *Let $\mathcal{G}$ be a tame Fréchet Lie group and $\mathcal{M}$ a tame Fréchet manifold on which $\mathcal{G}$ acts tamely and transitively. Suppose for some $m \in \mathcal{M}$ the subgroup $\mathcal{H}$ of elements of $\mathcal{G}$ which fix m is a closed tame submanifold, and that there is a short exact sequence of Fréchet spaces*

$$0 \to T_1\mathcal{H} \to T_1\mathcal{G} \to T_m\mathcal{M} \to 0$$

which admits a tame linear splitting. Then $\mathcal{G}$ is a principal $\mathcal{H}$-bundle over $\mathcal{M}$ with projection $P(g) = gm$.

PROOF. Since $\mathcal{G}$ acts transitively on $\mathcal{M}$, the fibres of P are all cosets of $\mathcal{H}$ in $\mathcal{G}$, and $\mathcal{H}$ acts transitively and fixed-point free on itself. If the sequence above splits at the identity $1 \in \mathcal{G}$, then it splits at every point in $\mathcal{G}$ using the group action. Then the result follows from the previous theorem.

We consider an example of a subgroup of the diffeomorphism group. Let X be a compact manifold. We say that a smooth positive measure μ on X is one which in local coordinates looks like $\mu = m(x)\,dx$ where $m(x)$ is a smooth positive function and dx is Lebesgue measure. Its total measure is

$$\mu(X) = \int_X d\mu.$$

We let $\mathcal{M}(X)$ be the Fréchet manifold of smooth positive measures on X of total measure 1. The smooth measures are sections of a vector bundle M over X, those of total measure 1 are an affine subspace, and the positive ones are an open subset. If μ and ν are smooth positive measures in $\mathcal{M}(X)$ then ν/μ is a smooth positive function of average value 1 with respect to μ. This provides a useful (global) coordinate chart.

2.5.3. THEOREM. *The diffeomorphism group $\mathcal{D}(X)$ acts transitively on the space $\mathcal{M}(X)$ of smooth positive measures of total measure one. For each $\mu \in \mathcal{M}(X)$ the subgroup $\mathcal{D}_\mu(X)$ of diffeomorphisms which preserve the measure μ is a closed tame Lie subgroup, and $\mathcal{D}(X)$ is a tame principal bundle over $\mathcal{M}(X)$ with fibre $\mathcal{D}_\mu(X)$ under the projection $P\phi = \phi\mu$. Hence $\mathcal{D}(X)/\mathcal{D}_\mu(X) = \mathcal{M}(X)$.*

PROOF. First note that $\mathcal{M}(X)$ is connected. For if μ and ν are smooth positive measures of total measure one, so is $(1 - t)\mu + t\nu$ for $0 \leqslant t \leqslant 1$. Next recall how to construct a coordinate chart near the identity 1 in $\mathcal{D}(X)$. We choose a tubular neighborhood U of the zero section in the tangent bundle TX and a diffeomorphism Φ of U to a neighborhood V of the diagonal in $X \times X$, which takes the zero vector at $x \in X$ into the point (x, x) on the diagonal, and which has for its derivative along the zero section in the vertical direction the

inclusion of TX into the second factor of $TX \times TX$. In a local coordinate chart

$$\Phi(x, v) = (x, \phi(x, v))$$

where $\phi(x, 0) = x$ and $D_v\phi(x, 0) = I$. We could construct Φ globally by patching such maps together with a partition of unity. Then Φ defines a diffeomorphism of a neighborhood $\tilde{U}$ of zero in $\mathcal{C}^\infty(X, TX)$ into a neighborhood $\tilde{V}$ of the identity in $\mathcal{D}(X)$ by letting $\Phi(v)$ be the diffeomorphism whose graph is the composition of Φ and v. In local coordinates

$$\Phi(v)\colon x \to \phi(x, v(x)).$$

Given a measure $\mu \in \mathcal{M}(X)$ and a diffeomorphism $\phi \in \mathcal{D}(X)$ we let $\phi\mu$ be the pull-back measure. This defines a smooth tame action of $\mathcal{D}(X)$ on $\mathcal{M}(X)$, as is clear from our estimates on compositions. We define a map

$$P\colon \mathcal{C}^\infty(X, TX) \to \mathcal{C}^\infty(X)$$

by letting

$$P(v) = \Phi(v)\mu/\mu.$$

2.5.4. LEMMA. *The map P is a nonlinear partial differential operator of degree 1. Its derivative $DP(v)w$ is a family of linear partial differential operators of degree 1 in w with coefficients which are nonlinear differential operators of degree 1 in v. At $v = 0$*

$$DP(0)w = \nabla_\mu w$$

is the divergence of the vector field w with respect to the measure μ.

PROOF. We write out $P(v)$ in local coordinates. The diffeomorphism

$$y = \phi(x, v(x))$$

has for its derivative the matrix

$$M(v) = \frac{\partial y}{\partial x} = \frac{\partial \phi}{\partial x}(x, v(x)) + \frac{\partial \phi}{\partial v}(x, v(x))\frac{\partial v}{\partial x}(x)$$

which is a nonlinear differential operator of degree 1 in v. Then so is the Jacobian $J(v) = \det M(v)$. Since $\mu = m(x)dx$ in local coordinates, we have

$$P(v) = J(v) \cdot m(\phi(x, v(x)))/m(x)$$

which is also a nonlinear differential operator of degree 1 in v. Hence P is a smooth tame map of the type considered before. Its derivative may be computed as follows. If $M(v)$ is a matrix depending smoothly on v and $J(v)$ is its determinant then

$$DJ(v)w = J(v) \cdot \operatorname{trace}\{M(v)^{-1} \cdot DM(v)w\}$$

as can easily be seen by diagonalizing the matrix (or at least upper-triangularizing). Then we get the formula

$$DP(v)w = P(v)\left[\operatorname{trace}\{M(v)^{-1} \cdot DM(v)w\} + Q(v)w\right]$$

where

$$DM(v)w = \frac{\partial^2\phi}{\partial x \partial v}(x, v(x))w(x) + \frac{\partial^2\phi}{\partial v^2}(x, v(x))\frac{\partial v}{\partial x}(x)w(x) + \frac{\partial\phi}{\partial v}(x, v(x))\frac{\partial w}{\partial x}(x)$$

is a differential operator of degree 1 linear in w with coefficients depending nonlinearly on v, and where

$$Q(v)w = \frac{\partial \log m}{\partial x}(\phi(x, v(x)))\frac{\partial\phi}{\partial v}(x, v(x))w(x)$$

is an operator of degree zero linear in w and nonlinear in v. At $v = 0$ we have $\partial\phi(x,0)/\partial v = I$ so $\partial^2\phi(x,0)/\partial x \partial w = 0$. Then the formula simplifies to

$$DP(0)w = \operatorname{trace}\frac{\partial w}{\partial x} + \frac{\partial \log m}{\partial x}, \qquad w = \nabla_\mu w,$$

which is the formula for the divergence of w with respect to the measure μ. It is the rate at which the measure expands or contracts under a flow in the direction w. This proves the lemma.

Next we choose a Riemannian metric $g = \{g_{ij}\}$ on X with measure μ, so that $m = \sqrt{\det g_{ij}}$. We can then take the gradient $\nabla_g f$ of a function f on X with respect to the measure g, defined by

$$\nabla_g f^i = g^{ij}\partial f/\partial x^j.$$

Note that $\nabla_\mu \nabla_g f = \Delta_g f$ is the Laplacian of f with respect to the metric g. Let us let $\mathcal{C}_0^\infty(X)$ denote the closed subspace of $\mathcal{C}^\infty(X)$ of functions of average value zero with respect to the measure μ. Also let V denote the closed subspace of $\mathcal{C}^\infty(X, TX)$ of vector fields v on X which are divergence-free with respect to μ, so that $\nabla_\mu v = 0$. We define a map on some neighborhood U of zero in the product

$$Q\colon U \subseteq V \times \mathcal{C}_0^\infty(X) \to V \times \mathcal{C}_0^\infty(X)$$

by letting

$$Q(v, f) = \big(v, P(v + \nabla_g f) - 1\big).$$

We claim Q is invertible in a neighborhood of zero by the Nash-Moser inverse function theorem. To see this we can compute its derivative

$$DQ(v, f)(w, h) = \big(w, DP(v + \nabla_g f)(w + \nabla_g h)\big).$$

To solve $DQ(v, f)(w, h) = (w, k)$ we need to find h with

$$DP(v + \nabla_g f)\nabla_g h = k - DP(v + \nabla_g f)w.$$

Now the operator $DP(v + \nabla_g f)\nabla_g h$ is a linear differential operator of degree 2 in h with coefficients which are smooth nonlinear differential operators of degree 1 in v and 2 in f. When $v = 0$ and $f = 0$ the operator is just $\nabla_\mu \nabla_g h = \Delta_g h$. Since the Laplacian is elliptic, the operator remains elliptic for small v and f. Moreover the operator always has the constant functions for its

null space, and takes its range in the functions of average value zero with respect to the measure μ. Therefore the operator defines an invertible linear map of $\mathcal{C}_0^\infty(X)$ to itself. Moreover the family of inverses forms a smooth tame family of linear maps by our argument for the Green's operator of an elliptic family. Hence DQ is always an invertible family with a smooth tame family of inverses. Then we can apply the inverse function theorem.

Since Q is invertible, we obtain a coordinate chart on a neighborhood of the identity in $\mathcal{G}$ by composing Q^{-1} with the chart Φ introduced before. In this chart with domain in $V \times \mathcal{C}_0^\infty(X)$ the image of $V \times 0$ corresponds to the subgroup $\mathcal{D}_\mu(X)$. This shows that $\mathcal{D}_\mu(X)$ is a closed tame submanifold near the identity, and hence everywhere using the group operation. Moreover in our chart the second factor describes the change in the measure μ under the diffeomorphism. Hence the action of $\mathcal{D}(X)$ on $\mathcal{M}(X)$ is locally transitive, and then globally so since $\mathcal{M}(X)$ is connected. The rest of the theorem follows from Corollary 2.5.2.

2.5.5. *Problem.* If X is a compact manifold with a symplectic form ω, is the subgroup of the diffeomorphism group $\mathcal{D}_\omega(X)$ which preserves ω a smooth tame submanifold?

2.5.6. *Problem.* If X is a compact manifold with a contact structure B, is the subgroup of the diffeomorphism group $\mathcal{D}_B(X)$ which preserves the contact structure B a smooth tame submanifold?

III.3. Generalizations.

3.1. *The Nash-Moser theorem for exact sequences.* Our first generalization is useful in problems involving deformation of structures. Let F, G and H be tame Fréchet spaces and let P and Q be two smooth tame maps between open subsets U, V and W,

$$U \subseteq F \xrightarrow{P} V \subseteq G \xrightarrow{Q} W \subseteq H$$

such that the composition $QP = 0$.

3.1.1. THEOREM. *Suppose that for each $f \in F$ the image of $DP(f)$ is the entire null space of $DQ(P(f))$. Suppose moreover that we can find two smooth tame maps*

$$VP\colon (U \subseteq F) \times G \to F, \qquad VQ\colon (U \subseteq F) \times H \to G$$

with $VP(f)g$ and $VQ(f)h$ linear separately in g and h such that

$$DP(f)VP(f)g + VQ(f)DQ(P(f))g = g$$

for all $f \in U \subseteq F$ and all $g \in G$. Then for any $f_0 \in U$ the image of a neighborhood of f_0 fills out a neighborhood of $g_0 = P(f_0)$ in the subset of $V \subseteq G$ where $Q(g) = 0$. Moreover we can find a smooth tame map $S\colon V' \subseteq G \to U' \subseteq F$ from a neighborhood V' of g_0 to a neighborhood U' of f_0 such that $P(S(g)) = g$ whenever $g \in V'$ and $Q(g) = 0$.

PROOF. We refer the reader to our paper [**4**, §2.6].

We now give a version of the above theorem which is invariantly defined for manifolds, where the condition $Q(g) = 0$ does not make much sense.

3.1.2. THEOREM. *Let $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ be smooth tame Fréchet manifolds and let P, Q and R be smooth tame maps*

$$\mathcal{F} \xrightarrow{P} \mathcal{G} \underset{R}{\overset{Q}{\rightrightarrows}} \mathcal{H}$$

such that $Q(P(f)) = R(P(f))$ for all $f \in \mathcal{F}$. Let $T\mathcal{F}$ be the tangent bundle to $\mathcal{F}$, $T\mathcal{G}$ the pull-back to $\mathcal{F}$ of the tangent bundle over $\mathcal{G}$ by the map P, and $T\mathcal{H}$ the pull-back to $\mathcal{F}$ of the tangent bundle to $\mathcal{H}$ by the map $Q \circ P = R \circ P$. Then there is a sequence of linear maps of vector bundles over $\mathcal{F}$

$$T\mathcal{F} \xrightarrow{DP} T\mathcal{G} \xrightarrow{DQ-DR} T\mathcal{H}$$

with $(DQ - DR) \circ DP = 0$. Suppose that this sequence is exact, and admits a splitting by smooth tame linear maps V and W of the vector bundles

$$T\mathcal{F} \underset{V}{\overset{DP}{\rightleftarrows}} T\mathcal{G} \underset{W}{\overset{DQ-DR}{\rightleftarrows}} T\mathcal{H}$$

such that $DP \circ V + W \circ (DQ - DR) = I$ (the identity). Then for every f_0 in $\mathcal{F}$ the image of a neighborhood of f_0 fills up a neighborhood of $g_0 = P(f_0)$ in the subset of $\mathcal{G}$ where $Q(g) = R(g)$. Moreover we can find a smooth tame map S of a neighborhood of g_0 in g to a neighborhood of f_0 in $\mathcal{F}$ such that $P(S(g)) = g$ whenever $Q(g) = R(g)$.

PROOF. Choose local coordinate charts and apply the previous result to P and $Q - R$.

3.2. *Embedding surfaces of negative curvature.*

3.2.1. EXAMPLE. Consider a compact surface M with smooth boundary embedded in R^3 with strictly negative curvature. In this case the second fundamental form b_{ij} has two characteristic directions in which it is zero. The characteristic curves of M are the curves in the characteristic directions of b_{ij}.

3.2.2. DEFINITION. We say M is in proper position if ∂M has two components and each characteristic curve starts in one boundary component and ends in the other without being tangent to the boundary. Note that if M is in proper position so is any nearby surface $\tilde{M}$. If $y = f(x)$ is a positive convex function on $a \leqslant x \leqslant b$ and we revolve its graph around the x-axis the resulting surface of revolution is in proper position. We will prove the following result.

3.2.3. THEOREM. *Let M be a compact surface with boundary of strictly negative curvature in proper position. Then* (A) *any small isometric motion of one boundary curve extends uniquely to a small isometric motion of the whole surface, and* (B) *any metric close to the given one also admits an isometric embedding close to the given one, and* (C) *the embedding of the new metric is uniquely determined by the embedding of one boundary component, which may be any isometric embedding close to the given one.*

PROOF. First observe that we can orient each characteristic curve so it goes from one boundary component $\partial^+ M$ to the other one $\partial^- M$. At each point there is the curve going to the left and the one going to the right. By continuity each curve is a left or right going curve at all its points. The left going curves

give a diffeomorphism between M and $\partial^+ M \times I$ where I is an interval, and so do the right going curves. Thus M is an annulus.

We can choose left and right unit vector fields y^i and z^i on M tangent to the left and right going characteristic curves. Then we have

$$g_{ij}y^iy^j = 1, \quad g_{ij}z^iz^j = 1,$$
$$b_{ij}y^iy^j = 0, \quad b_{ij}z^iz^j = 0,$$

and if μ_{ij} is the measure tensor then

$$\mu_{ij}y^iz^j > 0.$$

3.2.4. LEMMA. *We can find a weight function p on M such that $p = 0$ on $\partial^+ M$ and $p = 1$ on $\partial^- M$, and we have*

$$y^i\partial_i p > 0 \quad \textit{and} \quad z^i\partial_i p > 0$$

everywhere on M, so that p is strictly increasing along the characteristic curves.

PROOF. Given any point in M, follow forward along the left and right going characteristic curves to $\partial^- M$, and let s^- be the distance separating the endpoints along $\partial^- M$. (To make s^- continuous it may be necessary to count the distance as though it were on the universal cover of M, in case we go around $\partial^- M$ several times. Can this ever really happen?) Likewise follow back along the characteristic curves and let s^+ be the separation on $\partial^+ M$. Then put

$$p = \frac{2}{\pi} \arctan \frac{s^+}{s^-}.$$

If we move in the forward direction along either a left or right going characteristic curve, it is clear that s^+ will increase and s^- will decrease. Hence p has the required properties. We do not actually need the part that p is constant on the boundaries, but we include it anyway. Note that the converse holds; if p is constant on the boundary components and strictly increasing along the characteristic curves then M is in proper position. Note also that for a surface of revolution around the x-axis we could use $p = x$.

We wish to transform our variational problem into a symmetric system. To accomplish this we introduce the tensor c^{ijk} which is completely symmetric in all three indices given by

$$c^{ijk} = y^iy^jy^k + z^iz^jz^k.$$

We can then define a linear transformation L of the bundle $L^2_S(TM)$ of symmetric bilinear forms into the bundle TM of tangent vectors by letting $L(h_{jk}) = v^i$ where $v^i = c^{ijk}h_{jk}$.

3.2.5. LEMMA. *The linear transformation L is surjective with kernel given by the one dimensional subbundle B spanned by the second fundamental form b_{ij}. Hence L gives an isomorphism of the quotient bundle $Q = L^2_S(TM)/B$ onto TM.*

PROOF. Since $b_{jk}y^jy^k = 0$ and $b_{jk}z^jz^k = 0$ we have $c^{ijk}b_{jk} = 0$. Hence B is in the kernel. Now a basis for $L^2_S(TM)$ is given by the bilinear forms

$$y_jy_k, \qquad z_jz_k, \qquad y_jz_k + y_kz_j$$

where $y_j = g_{jl}y^l$ and $z_j = g_{jl}z^l$ are the dual basis, and b_{jk} is a multiple of $y_jz_k + y_kz_j$. We shall show that L is injective on the subspace spanned by y_jy_k and z_jz_k. Indeed

$$c^{ijk}y_jy_k = y^i \quad \text{and} \quad c^{ijk}z_jz_k = z^i.$$

Since y^i and z^i are linearly independent, the lemma follows.

Let $\mathcal{F}(M)$ be the Fréchet manifold of embeddings of M into R^3, $G(M)$ the Fréchet space of symmetric bilinear forms on M and $\mathcal{G}(M)$ the open subset of $G(M)$ of Riemannian metrics. Similar definitions apply to $\mathcal{F}(\partial^+ M)$, $G(\partial^+ M)$ and $\mathcal{G}(\partial^+ M)$. We define a complex of smooth tame maps P and Q

$$\mathcal{F}(M) \xrightarrow{P} \mathcal{G}(M) \times \mathcal{F}(\partial^+ M) \xrightarrow{Q} G(\partial^+ M)$$

by letting $P(f^\alpha) = (g_{ij}, h^\alpha)$ and $Q(h^\alpha) = k_{ij}$ where f^α is an embedding of M into R^3,

$$g_{ij} = g_{\alpha\beta}\partial_i f^\alpha \partial_j f^\beta$$

is the metric induced on M by the Euclidean metric $g_{\alpha\beta}$ on R^3 via the embedding f^α, $h^\alpha = f^\alpha \mid \partial^+ M$ is the restriction of the embedding to $\partial^+ M$, and

$$k_{ij} = g_{\alpha\beta}\partial_i h^\alpha \partial_j h^\beta - g_{ij} \mid \partial^+ M$$

is the difference between the metric induced on $\partial^+ M$ by the embedding h^α of $\partial^+ M$ into R^3 and the metric obtained by restricting the metric g_{ij} on M to the boundary component $\partial^+ M$. Note that $QP = 0$. We want to show that P is surjective and the image of P fills out the null space of Q in a neighborhood of a proper embedding. This says that for any metric g_{ij} on M near the given one and for any embedding h^α of $\partial^+ M$ near the given one which induces the same metric on $\partial^+ M$ as g_{ij} there is a unique embedding f^α near the given one which induces the metric g_{ij} on M and agrees with h^α on $\partial^+ M$. Both assertions follow from the following lemma by the Nash-Moser theorems for injective maps and for exact sequences.

3.2.6. LEMMA. *The linearized sequence DP and DQ is a short exact sequence which admits a smooth tame splitting by maps VP and VQ so that*

$$VP \circ DP = I, \qquad DP \circ VP + VQ \circ DQ = I, \qquad DQ \circ VQ = I$$

in the diagram

$$0 \to T\mathcal{F}(M) \underset{VP}{\overset{DP}{\rightleftarrows}} T\mathcal{G}(M) \times T\mathcal{F}(\partial^+ M) \underset{VQ}{\overset{DQ}{\rightleftarrows}} TG(\partial^+ M) \to 0.$$

PROOF. We let $\tilde{f}^\alpha$, $\tilde{g}_{ij}$, $\tilde{h}^\alpha$, and $\tilde{k}_{ij}$ denote tangent vectors at f^α, g_{ij}, h_α and k_{ij} respectively. Near $\partial^+ M$ we choose coordinates x^1 and x^2 so that M is given by $x^1 \geqslant 0$. Then x^1 is the coordinate function on $\partial^+ M$ and k_{ij} has only the component k_{11}. Since the boundary is not characteristic, we know $b_{11} \neq 0$ in such coordinates. The formulas for DP are

$$\tilde{g}_{ij} = g_{\alpha\beta}\partial_i \tilde{f}^\alpha \partial_j f^\beta + g_{\alpha\beta}\partial_i f^\alpha \partial_j \tilde{f}^\beta, \qquad \tilde{h}^\alpha = \tilde{f}^\alpha \mid \partial^+ M$$

and the formula for DQ is

$$\tilde{k}_{11} = 2g_{\alpha\beta}\partial_1 h^\alpha \partial_1 \tilde{h}^\beta - \tilde{g}_{11} \,|\, \partial^+ M.$$

We shall show that for all $\tilde{g}_{ij}$ and $\tilde{h}^\alpha$ with $\tilde{k}_{11} = 0$ there is a unique $\tilde{f}^\alpha$ which solves the equations.

As before we decompose $\tilde{f}^\alpha$ into its tangential and normal components. Instead of the tangential vector components $\tilde{v}^i$ we use the dual vectors $\tilde{v}_i = g_{ij}\tilde{v}^j$ lowering indices. Thus we let

$$\tilde{f}^\alpha = g^{ij}\tilde{v}_i \partial_j f^\alpha + \tilde{w} n^\alpha$$

where $\tilde{v}_i$ are the tangential part and $\tilde{w}$ the normal part of $\tilde{f}^\alpha$. Likewise we let

$$\tilde{h}^\alpha = g^{ij}\tilde{r}_i \partial_j f^\alpha + \tilde{s} n^\alpha$$

where $\tilde{r}_i$ is the tangential part and $\tilde{s}$ the normal part of $\tilde{h}^\alpha$. Then as before the equations for DP become

$$\partial_i \tilde{v}_j + \partial_j \tilde{v}_i + 2b_{ij}\tilde{w} = \tilde{g}_{ij} \quad \text{on } M,$$
$$\tilde{v}_i \,|\, \partial^+ M = \tilde{r}_i \quad \text{and} \quad \tilde{w} \,|\, \partial^+ M = \tilde{s}$$

while the equations for DQ become

$$2\partial_1 \tilde{r}_1 + 2b_{11}\tilde{s} - \tilde{g}_{11} \,|\, \partial^+ M = \tilde{k}_{11}.$$

Given $\tilde{g}_{ij}$, $\tilde{r}_i$ and $\tilde{s}$ solving the equations for DQ, we must show that there is a unique solution for $\tilde{v}_i$ and $\tilde{w}$ of the equations for DP. Our procedure is the following. Given $\tilde{g}_{ij}$ and $\tilde{r}_i$ we let $\tilde{v}_i$ be the solution of the equations

$$(*) \qquad \begin{cases} \partial_i \tilde{v}_j + \partial_j \tilde{v}_i \equiv \tilde{g}_{ij} \mod b_{ij}, \\ \tilde{v}_i \,|\, \partial^+ M = \tilde{r}_i. \end{cases}$$

We will show this solution exists and is unique. We can then solve for $\tilde{w}$ algebraically with

$$\partial_i \tilde{v}_j + \partial_j \tilde{v}_i + 2b_{ij}\tilde{w} = \tilde{g}_{ij}.$$

We take the $\tilde{f}^\alpha$ determined by this $\tilde{v}_i$ and $\tilde{w}$ to be $\tilde{f}^\alpha = VP(\tilde{g}_{ij}, \tilde{r}_i, \tilde{s})$. Note it is independent of $\tilde{s}$. If $k_{11} = 0$ then we will automatically have $\tilde{w} \,|\, \partial^+ M = \tilde{s}$. More generally we can define the map

$$(\tilde{g}_{ij}, \tilde{r}_i, \tilde{s}) = VQ(\tilde{k}_{11})$$

by letting

$$\tilde{s} = \tilde{k}_{11}/2b_{11}$$

and putting $\tilde{g}_{ij} = 0$ and $\tilde{r}_i = 0$. It is easy to check that this gives a splitting.

We can transform the equations $(*)$ into a symmetric system using the transformation L defined before. Put

$$\tilde{l}^k = c^{ijk}\tilde{g}_{ij}.$$

Then $L(\tilde{g}_{ij}) = \tilde{l}^k$ gives an isomorphism of the quotient bundle $Q = L^2_S(TM)/B$ onto the tangent bundle TM. The equations (*) are equivalent to the equations

$$(**) \qquad \begin{cases} 2c^{ijk}\partial_i \tilde{v}_j = \tilde{l}^k & \text{on } M, \\ \tilde{v}_j \mid \partial^+ M = \tilde{r}_j. \end{cases}$$

The equations define a first order linear operator from $\tilde{v}_j \in T^*M$ into the dual bundle $\tilde{l}^k \in TM$. The matrix of the symbol at a cotangent vector ζ_i is just $2c^{ijk}\zeta_i$, as we see by replacing ∂_i by ζ_i, and this is clearly symmetric in j and k since c^{ijk} is fully symmetric. Therefore we have a symmetric system as defined before.

3.2.7. LEMMA. *If p is a function with $y^i\partial_i p > 0$ and $z^i\partial_i p > 0$ then p is a positive weight function for the symmetric system defined by*

$$c^{ijk} = y^i y^j y^k + z^i z^j z^k$$

provided y^i and z^i span the tangent space at each point.

PROOF. We must check that $c^{ijk}\partial_i p = m^{jk}$ is a positive-definite matrix. Let x_j be any nonzero covector. Then

$$c^{ijk}\partial_i p\, x_j x_k = y^i\partial_i p\, y^j x_j y^k x_k + z^i\partial_i p\, z^j x_j z^k x_k.$$

Now $y^i\partial_i p$ and $z^i\partial_i p$ are positive, and if y^j and z^j span the tangent space then either $y^j x_j$ or $z^j x_j$ is not zero. It follows that $c^{ijk}\partial_i p\, x_j x_k > 0$.

Finally we must check that if ν_i is an inward pointing normal covector at the boundary then $c^{ijk}\nu_i$ is positive-definite on $\partial^+ M$ and negative-definite on $\partial^- M$. On $\partial^+ M$ we know that $y_i\nu_i > 0$ and $z^i\nu_i > 0$, so as before if $x_i \neq 0$ then

$$c^{ijk}\nu_i x_j x_k = y^i\nu_i y^j x_j y^k x_k + z^i\nu_i z^j x_j z^k x_k > 0$$

since y^j and z^j span the tangent space. The inequality reverses on $\partial^- M$ since $y^i\nu_i < 0$ and $z^i\nu_i < 0$.

It follows now by the theory of symmetric systems in §II.3.4 that the equations (**) have a unique solution $\tilde{v}_j$ which is a smooth tame function of $\tilde{l}^k$ and the coefficients c^{ijk}. Then the splitting maps VP and VQ are smooth tame functions of the embedding f^α and the data $\tilde{g}_{ij}$, $\tilde{h}^\alpha$ for VP and $\tilde{k}_{ij}$ for VQ. This proves the theorem.

3.3. *The implicit function theorem with quadratic error.* Our other generalization of the Nash-Moser theorem is an implicit function theorem with quadratic remainder term of the sort introduced by Zehnder [**24**].

3.3.1. THEOREM. *Let F, G and H be tame Fréchet spaces and let A be a smooth tame map defined on an open set U in $F \times G$ to H,*

$$A\colon U \subseteq F \times G \to H.$$

Suppose that whenever $A(f, g) = 0$ the partial derivative $D_f A(f, g)$ is surjective, and there is a smooth tame map $V(f, g)h$ linear in h,

$$V\colon (U \subseteq F \times G) \times H \to F,$$

and a smooth tame map $Q(f, g)\{h, k\}$ *bilinear in* h *and* k,

$$Q\colon (U \subseteq F \times G) \times H \times H \to F,$$

such that for all (f, g) *in* U *and all* $h \in H$ *we have*

$$D_f A(f, g)V(f, g)h = h + Q(f, g)\{A(f, g), h\}$$

so that V *is an approximate right inverse for* $D_f A$ *with quadratic error* Q. *Then if* $A(f_0, g_0) = 0$ *for some* $(f_0, g_0) \in U$, *we can find neighborhoods of* f_0 *and* g_0 *such that for all* g *in the neighborhood of* g_0 *there exists an* f *in the neighborhood of* f_0 *with* $A(f, g) = 0$. *Moreover the solution* $f = B(g)$ *is defined by a smooth tame map* B.

PROOF. We could prove this by a slight generalization of our argument for the inverse function theorem, but to spare the reader the tedious details we merely reduce it to an application of the theorem on near-projections in [**4**, §2.1]. We may assume $f_0 = 0$ and $g_0 = 0$.

3.3.2. LEMMA. *The map*

$$P(f, g, h) = (f - V(f, g)A(f, g), g, h - A(f, g))$$

is a near-projection in a neighborhood of $(0, 0, 0)$.

PROOF. We let

$$\Delta f = V(f, g)A(f, g), \quad \Delta g = 0, \quad \Delta h = A(f, g).$$

Using the formula for the quadratic error it is easy to compute

$$DP(f, g, h)(\Delta f, \Delta g, \Delta h) = (k, 0, l)$$

where

$$k = V(f, g)Q(f, g)\{\Delta h, \Delta h\} - D_f V(f, g)\{\Delta f, \Delta h\},$$
$$l = Q(f, g)\{\Delta h, \Delta h\}$$

so that k and l are quadratic in Δf, and Δh. This proves P is a near-projection.

It follows that the algorithm

$$f_{n+1} = f_n + S_n V(f_n, g)A(f_n, g),$$
$$g_{n+1} = g, \qquad h_{n+1} = h_n + S_n A(f_n, g)$$

induced by the near-projection with initial data $f_0 = 0$, $g_0 = g$, and $h_0 = 0$ converges as $n \to \infty$. The f_n converge to a solution f of $A(f, g) = 0$, and the solution f is a smooth tame function of g.

If instead $V(f, g)$ is an approximate left inverse for $D_f A(f, g)$ with a quadratic error term, then the solution of $A(f, g) = 0$ for a given g may not exist, but if it does it is unique. In particular we have the following result.

3.3.3. THEOREM. *Let* F, G *and* H *be tame Fréchet spaces and*

$$A\colon U \subseteq F \times G \to H$$

a smooth tame map with $A(f_0, g_0) = 0$. *Let*

$$V\colon (U \subseteq F \times G) \times H \to F \quad \textit{and} \quad Q\colon (U \subseteq F \times G) \times H \times H \to F$$

be smooth tame maps with $V(f, g)h$ linear in h and $Q(f, g)\{h, k\}$ bilinear in h and k, such that for all $(f, g) \in U$ and all h

$$D_f A(f, g)V(f, g)h = h + Q(f, g)\{A(f, g), h\}.$$

Then we can find neighborhoods of f_0 and g_0 such that if f_1 and f_2 are in the neighborhood of f_0 and g is in the neighborhood of g_0 with $A(f_1, g) = 0$ and $A(f_2, g) = 0$ then $f_1 = f_2$.

PROOF. By Taylor's theorem with integral remainder we can find a smooth tame map

$$B(f_1, f_2, g)\{h, k\} = \int_0^1 (1 - t)D_f^2 A((1 - t)f_1 + tf_2, g)\{h, k\}\, dt$$

such that

$$A(f_2, g) = A(f_1, g) + D_f A(f_1, g)(f_2 - f_1) + E$$

where

$$E = B(f_1, f_2, g)\{(f_2 - f_1), (f_2 - f_1)\}.$$

We can write

$$f_2 - f_1 = V(f_1, g)D_f A(f_1, g)(f_2 - f_1) + Q(f_1, g)\{A(f_1, g), f_2 - f_1\}.$$

Now if $A(f_1, g) = 0$ and $A(f_2, g) = 0$ then

$$f_2 - f_1 = -V(f_1, g)E.$$

Then taking a suitable degree r and base b for our tame estimates we have

$$\| f_2 - f_1 \|_b \leq C \| f_2 - f_1 \|_{b+r}^2$$

when f_1 and f_2 lie in a neighborhood of f_0. By interpolation

$$\| f_2 - f_1 \|_{b+r}^2 \leq C \| f_2 - f_1 \|_{b+2r} \| f_2 - f_1 \|_b.$$

For any $\varepsilon > 0$ we have $\| f_2 - f_1 \|_{b+2r} < \varepsilon$ when f_1 and f_2 lie in a small enough neighborhood of f_0. This makes

$$\| f_2 - f_1 \|_b \leq C\varepsilon \| f_2 - f_1 \|_b$$

with the constant C independent of $\varepsilon > 0$. Taking $\varepsilon < 1/C$ we get $\| f_2 - f_1 \|_b = 0$ so $f_1 = f_2$.

There is an interesting global version of this theorem for manifolds and vector bundles, in which the quadratic error comes from the choice of a connection on the vector bundle.

. 3.3.4. THEOREM. *Let $\mathcal{F}$ and $\mathcal{G}$ be smooth tame Fréchet manifolds and let $\mathcal{V}$ be a smooth tame vector bundle over $\mathcal{F}$ with a smooth tame connection Γ. Let $P: \mathcal{F} \to \mathcal{G}$ be a smooth tame map of manifolds and let $Q: \mathcal{F} \to \mathcal{V}$ be a smooth tame section of the vector bundle. Let $T\mathcal{F}$ be the tangent bundle to $\mathcal{F}$ and let $P^*T\mathcal{G}$ be the pull-back of the tangent bundle to $\mathcal{G}$ by the map P. Then we have smooth tame linear vector bundle maps of bundles over $\mathcal{F}$*

$$DP: T\mathcal{F} \to P^*T\mathcal{G}, \qquad D_\Gamma Q: T\mathcal{F} \to \mathcal{V}$$

where DP is the derivative of P and $D_\Gamma Q$ is the covariant derivative of the section Q with respect to the connection Γ. Suppose that

$$DP \times D_\Gamma Q: T\mathcal{F} \to P^*T\mathcal{G} \times \mathcal{V}$$

is a smooth tame linear vector bundle isomorphism of bundles over $\mathcal{F}$ with a smooth tame inverse V. Suppose moreover that for some $f_0 \in \mathcal{F}$ we have $P(f_0) = g_0$ and $Q(f_0) = 0$. Then we can find neighborhoods of f_0 and g_0 such that for all g in the neighborhood of g_0 there exists a unique f in the neighborhood of f_0 with $P(f) = g$ and $Q(f) = 0$. Moreover the solution $f = S(g)$ is defined by a smooth tame map from a neighborhood of g_0 to a neighborhood of f_0 with $S(g_0) = f_0$.

If $DP \times D_\Gamma Q$ is merely surjective with a smooth tame right inverse, then the solution exists in the neighborhoods and is given by a smooth tame map S but it may not be unique. On the other hand if $DP \times D_\Gamma Q$ is merely injective with a smooth tame left inverse, then the solution may not exist, but if it does it is unique in the neighborhoods.

PROOF. We compute the local representatives. Locally $\mathcal{F}$ is an open set U in a Fréchet space F and G is an open set $\tilde{U}$ in a Fréchet space G, while $\mathcal{V}$ is locally a product $U \times H$ for some Fréchet space H. Locally P is given by a map

$$P: U \subseteq F \to \tilde{U} \subseteq G$$

and Q is given by a map

$$Q: U \subseteq F \to H.$$

The connection Γ is given locally by a map

$$\Gamma: (U \subseteq F) \times F \times H \to H$$

where $\Gamma(f)\{(f', h)\}$ is bilinear in f' and h. Locally $DP(f)f'$ is just the derivative of P, while

$$D_\Gamma Q(f)f' = DQ(f)f' + \Gamma(f)\{f', Q(f)\}$$

where $D_\Gamma Q$ is the covariant derivative of Q with respect to the connection Γ and DQ is the ordinary derivative of Q in our coordinate chart on $\mathcal{V}$. We define a map

$$A: U \times \tilde{U} \subseteq F \times G \to H \times H$$

by letting

$$A(f, g) = (P(f) - g, Q(f)).$$

Then

$$D_f A(f, g)f' = (DP(f)f', DQ(f)f')$$

and hence

$$D_f A(f, g)f' = DP \times D_\Gamma Q(f)f' - (0, \Gamma(f)\{f', Q(f)\}).$$

If $DP \times D_\Gamma Q$ is invertible then its inverse provides an approximate inverse for $D_f A$ with a quadratic error term arising from the connection term $\Gamma(f)\{f', Q(f)\}$, since $Q(f)$ is just the second term in $A(f, g)$. If we have only

a left or right inverse for $DP \times D_\Gamma Q$ then we get only a left or right approximate inverse for $D_f A$ with a quadratic error. We can now apply our previous results.

3.4. *A free-boundary problem.*

3.4.1. EXAMPLE. We consider a free boundary problem. (See Schaeffer [**19**] and Acker [**1**], who proves this result by other techniques.) Let B be a closed curve in the plane. We seek another closed curve C outside of B and a smooth function f on the closed annular region A between B and C with

$$(*) \qquad \begin{cases} \Delta f = 0 & \text{on } A, \\ f = 1 & \text{on } B, \\ f = 0 & \text{on } C, \\ \partial f/\partial n + 1 = 0 & \text{on } C \end{cases}$$

where $\partial f/\partial n$ is the outward normal derivative. Note that since the outer boundary C is free or undetermined we can impose an extra boundary condition on C. We prove the following result.

3.4.2. THEOREM. *For every smooth convex curve B in the plane there exists a unique solution of the free boundary problem* (*). *The curve C is also smooth and convex and the function f is smooth.*

Note that if B is not convex the solution may fail to exist, as when B is a horseshoe. We conjecture that the same result holds in R^n. The inverse function theorem part of the proof goes through there, but our a priori estimates come from conformal mapping.

Our boundary value problem has the following interesting physical interpretation. Imagine a perfect fluid flowing around an obstacle in the plane with convex boundary B. We look for a flow which is at rest outside of an annulus A with a jump discontinuity in the velocity on the outer boundary curve C which is to be determined. Suppose the fluid has unit density. We can write the velocity v in terms of a harmonic stream potential function f, so that $v = \text{curl } f$ and $\Delta f = 0$ in A. Since v is tangent to the boundary curves B and C we must have f constant on B and C. If the flow is stagnant outside C the pressure is constant there. Then particles on the boundary curve C experience only a normal force from the pressure, so the speed $|v|$ is constant on C. Since f is constant on C, $|v| = |\partial f/\partial n|$ where $\partial f/\partial n$ is the normal derivative. Thus $\partial f/\partial n$ is constant on C also. The difference between f on B and C is the mass flow rate across any curve cutting A from B to C, which we call the circulation. By choosing our units of length and time we can make the velocity on C and the circulation both equal to 1. This gives our equations (*) for the free boundary problem. Thus there exists a unique flow around any convex obstacle in the plane which is stagnant outside a compact set and has an arbitrarily given outer velocity and circulation. Physically it is hard to realize this solution for long due to viscosity.

PROOF. We start by using the inverse function theorem to show that the set of curves B for which we can solve the free boundary problem (*) is open. Let $\mathcal{C}$ be the Fréchet manifold of all annular domains in the plane with smooth

boundary, and likewise let $\mathcal{B}$ and $\mathcal{C}$ be the Fréchet manifolds of all boundary curves B and C. Then $\mathcal{A}$ is the open subset of $\mathcal{B} \times \mathcal{C}$ consisting of all pairs (B, C) with B inside C. This gives a smooth tame projection map $P: \mathcal{A} \to \mathcal{B}$. Let $\mathcal{C}^\infty(\mathcal{A})$ denote the Fréchet vector bundle over $\mathcal{A}$ whose fibre over a region $A \in \mathcal{A}$ is the Fréchet space $\mathcal{C}^\infty(A)$ of smooth functions on A. Likewise $\mathcal{C}^\infty(B)$ and $\mathcal{C}^\infty(\mathcal{C})$ are Fréchet vector bundles over $\mathcal{B}$ and $\mathcal{C}$, or over $\mathcal{A}$ by pull-back. There is a linear vector bundle map

$$L: \mathcal{C}^\infty(\mathcal{A}) \to \mathcal{C}^\infty(\mathcal{A}) \times \mathcal{C}^\infty(\mathcal{B}) \times \mathcal{C}^\infty(\mathcal{C})$$

of vector bundles over $\mathcal{A}$ defined by

$$Lf = (\Delta f, f|B, f|C).$$

Since the Dirichlet problem is always elliptic and invertible, we see that L is a smooth tame isomorphism. Taking $f = L^{-1}(0, 0, 1)$ defines a smooth tame section S of the vector bundle $\mathcal{C}^\infty(\mathcal{A})$ over $\mathcal{A}$, where for each $A \in \mathcal{A}$ the function $f = S(A)$ is the solution of $\Delta f = 0$ on A, $f = 0$ on B and $f = 1$ on C.

Using S we can define a smooth tame section Q of the bundle $\mathcal{C}^\infty(\mathcal{C})$ over $\mathcal{A}$, by letting $Q(A) = k$ where

$$k = \partial f/\partial n + 1 \quad \text{on } C$$

and $f = S(A)$. Note that the free boundary problem $(*)$ has a solution on A exactly when $Q(A) = 0$.

3.4.3. LEMMA. *If $\mathcal{B}$ is convex and the free boundary problem $(*)$ has a solution for B, then it also has a solution for all $\tilde{B}$ near B. If A is an annulus with inner boundary B and $Q(A) = 0$, then for every curve $\tilde{B}$ near B there exists a unique annulus $\tilde{A}$ near A with $\tilde{B}$ the inner boundary of $\tilde{A}$ and $Q(\tilde{A}) = 0$.*

PROOF. We use the Nash-Moser implicit function theorem. The bundles $\mathcal{C}^\infty(\mathcal{A})$ and $\mathcal{C}^\infty(\mathcal{B})$ and $\mathcal{C}^\infty(\mathcal{C})$ all acquire connections as explained previously. Computing the covariant derivative of Q using the connection on $\mathcal{C}^\infty(\mathcal{C})$, we must show that

$$DP \times DQ: T\mathcal{A} \to T\mathcal{B} \times \mathcal{C}^\infty(\mathcal{C})$$

is a smooth tame linear isomorphism of vector bundles over $\mathcal{A}$. Now $T\mathcal{A} = T\mathcal{B} \times T\mathcal{C}$ since $\mathcal{A}$ is an open subset of $\mathcal{B} \times \mathcal{C}$, while $T\mathcal{B} = \mathcal{C}^\infty(\mathcal{B})$ and $T\mathcal{C} = \mathcal{C}^\infty(\mathcal{C})$ using the unit normal vector fields to trivialize the normal bundle. Since $P: \mathcal{A} \to \mathcal{B}$ is projection on the first factor, so is $DP: T\mathcal{A} \to T\mathcal{B}$. Hence it suffices to invert the restriction of DQ to $T\mathcal{C} = \mathcal{C}^\infty(\mathcal{C})$.

We shall now compute

$$DQ: \mathcal{C}^\infty(\mathcal{C}) \to \mathcal{C}^\infty(\mathcal{C}).$$

Let $\tilde{h} \in \mathcal{C}^\infty(\mathcal{C})$ denote an infinitesimal variation of the outer boundary C in the direction of the outward normal and let $\tilde{f} = DS(h, 0) \in \mathcal{C}^\infty(\mathcal{A})$ be the induced variation on the function f, keeping the inner boundary B fixed. The variation of f on C is the sum of two terms, one due to the variation of f and one due to the variation of C, and in order to keep $f = 1$ on C we must have

$$\tilde{f} + (\partial f/\partial n)\tilde{h} = 0.$$

Since $\partial f/\partial n + 1 = 0$ on C we get $\tilde{f} = \tilde{h}$. Thus $\tilde{f}$ is the solution of the Dirichlet problem

$$(1) \qquad \begin{cases} \Delta \tilde{f} = 0 & \text{on } A, \\ \tilde{f} = 0 & \text{on } B, \\ \tilde{f} = \tilde{h} & \text{on } C. \end{cases}$$

Since $k = \partial f/\partial n + 1$ on C, the variation $\tilde{k}$ in k will be the variation in $\partial f/\partial n$. This has three parts, one due to varying f, one due to varying the boundary, and one due to varying the normal direction. We get

$$\tilde{k} = \frac{\partial \tilde{f}}{\partial n} + \frac{\partial^2 f}{\partial n^2}\tilde{h} - \nabla f \cdot \nabla \tilde{h}$$

since ∇f is perpendicular to the boundary and $\nabla \tilde{h}$ is taken along the boundary we can drop the third term. For the second term we observe that when $f = 0$ and $\partial f/\partial n + 1 = 0$ on C and $\Delta f = 0$ in A then $\partial^2 f/\partial n^2 = m$ where m is the curvature of C. (In higher dimensions m is the near curvature, the trace of the second fundamental form.) Then

$$(2) \qquad \tilde{k} = \partial \tilde{f}/\partial n + m\tilde{f}$$

using $\tilde{f} = \tilde{h}$ on C to eliminate $\tilde{h}$. The operator $DQ(\tilde{h}) = \tilde{k}$ is obtained by solving (1) for $\tilde{f}$ and finding $\tilde{k}$ by (2). To invert DQ, we must solve the elliptic boundary value problem

$$(3) \qquad \begin{cases} \Delta \tilde{f} = 0 & \text{on } A, \\ \tilde{f} = 0 & \text{on } B, \\ \partial \tilde{f}/\partial n + m\tilde{f} = \tilde{k} & \text{on } C \end{cases}$$

and take $\tilde{h} = \tilde{f} \mid C$. The above problem differs by the lower order term $m\tilde{f}$ from Laplace's equation with Dirichlet conditions on B and Neumann conditions on C. Hence it has index zero, and it is invertible if and only if its null space is zero.

3.4.4. LEMMA. *If C is convex then the elliptic system* (3) *is invertible.*

PROOF. This follows from the following energy estimate when $\tilde{k} = 0$, which is obtained by integrating by parts in $\tilde{f}\Delta\tilde{f}$.

$$\iint_A |\nabla \tilde{f}|^2 + \int_C m\tilde{f}^2 = 0.$$

If C is convex then $m \geqslant 0$ and $\nabla \tilde{f} = 0$, so $\tilde{f} = 0$.

The proof of Lemma 3.4.3 will be complete as soon as we make the following observation.

3.4.5. LEMMA. *If the free boundary problem* (*) *admits a solution with a convex inner boundary B then the outer boundary C is also convex.*

PROOF. By the theory of conformal mapping the region A can be mapped smoothly and conformally onto a standard annulus. Passing to the universal cover we get a conformal map of the strip $0 \leqslant v \leqslant 1$ in the uv-plane onto the

annulus A in the xy-plane with $v = 0$ mapping onto C and $v = 1$ mapping onto B and such that the map is periodic with some period p in u, such that $u \to u + p$ corresponds to going once around the annulus. In two dimensions Laplace's equation $\Delta f = 0$ is invariant under conformal transformations so we see that $v = f(x, y)$ is our solution function and u is the conjugate function. The flat metric in the xy-plane pulls back to a periodic conformally flat metric in the uv-plane

$$ds^2 = dx^2 + dy^2 = e^{-2g}(du^2 + dv^2)$$

where g is a smooth function of u and v periodic in u with period p. Since $|\nabla f|^2 = 1$ on C we must have $g = 0$ along $v = 0$. A basis of unit vectors in the g-metric is given by

$$T = e^g \partial/\partial u, \qquad N = e^g \partial/\partial v$$

where T is tangent and N normal to the level curves $f = v =$ constant. Since N is an inward pointing normal we have $T \cdot N = -mT$ where m is the curvature of the level curve. It is easy to compute

$$m = e^g \partial g/\partial v.$$

If B is convex then $m \geqslant 0$ along $v = 1$ and hence $\partial g/\partial v > 0$ along $v = 1$. The metric is conformally flat if and only if $\Delta g = 0$. It now follows from the maximum principle that $g \geqslant 0$ and $\partial g/\partial v \geqslant 0$ on the whole strip. (*Hint.* The function g has its maximum on the boundary, and it isn't along $v = 1$ by the inequality. Also $\Delta(\partial g/\partial v) = 0$ and $(\partial g/\partial v)(\partial g/\partial v) = -\partial^2 g/\partial u^2 = 0$ along $v = 0$ so we can apply the same argument to $\partial g/\partial v$.) Therefore $m \geqslant 0$ in the whole strip, so all the level curves of f are convex, including the outer boundary C where $f = 0$. This completes the proof of Lemma 3.4.5 and our application of the inverse function theorem.

To finish the proof we need a priori estimates on the solution f and its derivatives, which we shall briefly sketch. Let $M = \max_B m$ be the maximum curvature of the given inner boundary B. We make estimates in terms of constants $C(M)$ depending only on M.

3.4.6. LEMMA. *We have estimates*

$$\max g \leqslant C(M) \quad \textit{and} \quad \max m \leqslant C(M).$$

PROOF. We know $\Delta g = 0$ on $0 \leqslant v \leqslant 1$ and $g = 0$ on $v = 0$ and $e^g \partial g/\partial v = m \leqslant M$ on $v = 1$. Since $g \geqslant 0$ we have $\partial g/\partial v \leqslant M$ on $v = 1$ also. As before $\partial^2 g/\partial v^2 = 0$ on $v = 0$ and $\Delta(\partial g/\partial v) = 0$ on $0 \leqslant v \leqslant 1$. Then by the maximum principle $\partial g/\partial v \leqslant M$ on all of $0 \leqslant v \leqslant 1$. It is then easy to bound g and $m = e^g \partial g/\partial v$ by constants $C(M)$ depending on M. Note that the estimate $0 \leqslant g \leqslant C(M)$ gives estimates on the maximum and minimum width of the annulus A in terms of the maximum curvature M of the inner boundary B, while the estimate $0 \leqslant m \leqslant C(M)$ shows the outer boundary C is convex and estimates its maximum curvature in terms of M.

We can also estimate the period p. Note that p is just the perimeter of the outer boundary C, so p is at least the perimeter of B, which in turn is at least $2\pi/M$. On the other hand

$$\text{area } A = \int_{u=0}^{p}\int_{v=0}^{1} e^{2g}\,du\,dv \geqslant p$$

and we can estimate the area of A in terms of the diameter of B and the maximum width of A. Hence p can be estimated above and below by the maximum curvature M of B and the diameter of B.

Now we can estimate higher derivatives. Let $\|f\|_n$ denote the supremum norm of f and its derivatives of order at most n. Likewise we define $\|B\|_n$ using a coordinate chart on B in a neighborhood of some $\bar{B}$.

3.4.7. LEMMA. *We can make an estimate for* $n \geqslant 2$,

$$\|f\|_n \leqslant C(\|B\|_n)$$

where $C(\|B\|_n)$ *is a constant depending only on* $\|B\|_n$ *and* n, *and the estimate holds for all* B *in a neighborhood of* $\bar{B}$.

PROOF. Since $n \geqslant 2$, $\|B\|_n$ estimates the maximum curvature M of B, and its diameter, so our previous estimates apply. If D is any partial derivative then $\Delta Df = 0$ in A so Df attains its maximum on B or C. First we estimate on C in our new coordinate system uv where $f = v$. Then estimating $\|f\|_n$ reduces to estimating $\|g\|_{n-1}$. Since $\Delta g = 0$ on $0 \leqslant v \leqslant 1$, $g = 0$ on $v = 0$, and $g \leqslant C(M)$ on $0 \leqslant v \leqslant 1$, while the period p is bounded above and below, it is easy to estimate $\|g\|_n$ on $v = 0$ in the usual way.

The same estimate applies on $0 \leqslant v \leqslant 1 - \varepsilon$ for any $\varepsilon > 0$. We can choose $\delta > 0$ smaller than our estimate for the minimum width of A. Then the curve B_δ at distance δ from B lies in A, and in $0 \leqslant f \leqslant 1 - \varepsilon$ for $\varepsilon > 0$ small enough. We know $\Delta f = 0$ between B and B_δ, $f = 1$ on B and all the derivatives of f are estimated on B_δ. Then it is easy to derive estimates for the derivatives of f on B. One way to do this is to use a construction like our coordinate chart on $\mathcal{Q}$ to pull back each annulus A_δ, which will be a smooth tame function of B. The pull-back of f to $\bar{A}_\delta$ satisfies the Laplace equation in the variable metric and we can estimate the Dirichlet boundary data, so the tame estimates on the higher derivatives of f we proved earlier will hold. To estimate the C^n norm of f we use the $C^{n+\alpha}$ norm for $0 < \alpha < 1$. This requires the $C^{n-2+\alpha}$ norm of the induced metric, which can be estimated by its C^{n-1} norm. This in turn requires the C^n norm of the boundary curve B.

Next we derive estimates for the boundary curve C. Pick a point 0 inside $\bar{B}$ and take it as the origin for polar coordinates r and θ. For all B in a neighborhood of $\bar{B}$ the origin 0 lies inside B also. Write B as the graph of $r = b(\theta)$ and C as the graph of $r = c(\theta)$. Since B and C are convex and enclose the origin 0, the functions $b(\theta)$ and $c(\theta)$ are single-valued and smooth. The norm $\|B\|_n$ of the last lemma is equivalent to $\|b\|_n$.

3.4.8. LEMMA. *We can estimate for all* $n \geqslant 2$

$$\|c\|_n \leqslant C(\|b\|_n)$$

where $C(\|b\|_n)$ is a constant depending only on $\|b\|_n$ and the estimate holds for all B in a neighborhood of $\bar{B}$.

PROOF. We know that $|\nabla f| = e^{-g}$ and $g \leqslant C(M) \leqslant C(\|b\|_2)$, so $|\nabla f| \geqslant \varepsilon(\|b\|_2) > 0$. Also the maximum and minimum values of $c(\theta)$ are bounded in terms of $\|b\|_2$ since the width of A is. It is now an easy consequence of the implicit function theorem in finite dimensions to write an nth derivative of $c(\theta)$ in terms of nth derivatives of f along C. Combining with the previous lemma we get our estimate.

3.4.9. LEMMA. *Let B_t be a smooth path of convex curves in B parametrized by t for $0 \leqslant t \leqslant 1$. Suppose for $t = 0$ the free boundary problem has a solution f_0 which is smooth on an annulus $A_0 \in \mathcal{A}$. Then we can find a unique smooth path A_t in $\mathcal{A}$ and a path $f_t \in \mathcal{C}^\infty(\mathcal{A}_t)$ of sections over A_t which solve the free boundary problem and such that A_t has inner boundary B_t.*

PROOF. The inverse function theorem supplies us with the smooth paths A_t and f_t for $0 \leqslant t \leqslant \varepsilon$. If we can't continue it to $t = 1$, there will be a maximal ω such that the smooth solution exists on $0 \leqslant t \leqslant \omega$. We take $\bar{B} = B_\omega$ in the previous lemmas. Then for all t in $\omega - \delta < t < \omega$ for some $\delta > 0$ we have uniform estimates on $\|f_t\|_n$ and $\|c_t\|_n$, where B_t and C_t are the graphs of $r = b_t(\theta)$ and $r = c_t(\theta)$ in polar coordinates about some point 0 inside B_ω. Then for some sequence $t_j \to \omega$ we will have $C_{t_j} \to C_\omega$ and $f_{t_j} \to f_\omega$ on the annulus A_ω between B_ω and C_ω, by Ascoli's theorem. (After the C_{t_j} converge to C_ω, we may use our coordinate chart on $\mathcal{A}$ around A_ω to pull the functions f_{t_j} on A_{t_j} back to A_ω, and the pull-backs $f_{t_j}^*$ still have uniform bounds for $\|f_{t_j}^*\|_n$ on A_ω.) It is immediate that f_ω solves the free boundary problem on A_ω with inner boundary B_ω. By the implicit function theorem we get a smooth path $\tilde{f}_t$ of solutions on a smooth path $\tilde{A}_t$ of domains which agrees with f_ω and A_ω at $t = \omega$, and is the unique solution in a neighborhood of f_ω in $\mathcal{C}^\infty(\mathcal{A})$. Since the f_{t_j} converge to f_ω in $\mathcal{C}^\infty(\mathcal{A})$, the f_{t_j} and $\tilde{f}_{t_j}$ must agree when j is large enough. Hence $f_t = \tilde{f}_t$ for all t near ω, and the solution extends past ω. This gives a contradiction if ω is maximal.

3.4.10. LEMMA. *When B is a circle there exists a unique smooth solution to the free boundary problem.*

PROOF. If B is the circle $r = b$ in polar coordinates then C is the circle $r = c$ where c is the unique solution of the equation

$$ce^{-1/c} = b$$

and f is the function

$$f = \log(c/r)/\log(c/b).$$

(Note the function $y = xe^{-1/x}$ has $y' = (1 + 1/x)e^{-1/x} > 0$ for $0 < x < \infty$ and $y \to 0$ when $x \to 0$ while $y \to \infty$ as $x \to \infty$.)

To see that the solution is unique, we look at the associated metric function g introduced in Lemma 3.4.5. We know that g is periodic in the strip $0 \leqslant v \leqslant 1$

and satisfies $\Delta g = 0$ on the strip, while $g = 0$ along the boundary $v = 0$. On $v = 1$ we have

$$e^g \partial g/\partial v = 1/b$$

since the curvature of B is $1/b$ when B is a circle of radius b. [Note that in general we do not know the curvature as a function of u on $v = 1$ because the diffeomorphism attaching the line $v = 1$ to the inner boundary B is unknown, but when B is a circle the curvature is constant and this doesn't matter.] Then we can differentiate with respect to u to get

$$\begin{cases} \Delta\left(\dfrac{\partial g}{\partial u}\right) = 0 & \text{on } 0 \leqslant v \leqslant 1, \\ \dfrac{\partial g}{\partial u} = 0 & \text{on } v = 0, \\ e^g \dfrac{\partial}{\partial v}\left(\dfrac{\partial g}{\partial u}\right) + \dfrac{1}{b}\dfrac{\partial g}{\partial u} = 0 & \text{on } v = 1. \end{cases}$$

It now follows from the maximum principle that $\partial g/\partial u = 0$. [Note that $\partial g/\partial u$ has its maximum on the boundary, and it isn't on $v = 1$ since there e^g and $(\partial/\partial v)(\partial g/\partial u)$ and $\partial g/\partial u$ would all be positive. One needs to throw in an ε.] Then g is a function only of v, and we have the ordinary differential equation

$$\begin{cases} d^2g/dv^2 = 0 & \text{on } 0 \leqslant v \leqslant 1, \\ g = 0 & \text{on } v = 0, \\ e^g dg/dv = 1/b & \text{on } v = 1 \end{cases}$$

whose only solutions are of the form $g = av$. Since a $e^a = b$ we see that a is uniquely determined. The strip with metric $ds^2 = e^{-2g}(du^2 + dv^2)$ must map isometrically onto the annulus A, and such a map is unique up to a rotation which leaves $f = v$ invariant. Hence the solution is unique when B is a circle.

It is now easy to complete the proof that the solution exists and is unique for any convex B, using Lemmas 3.4.9 and 3.4.10. We can join any convex curve to a circle through a path of convex curves. By Lemma 3.4.7 the solution for the circle continues to a solution for B. If there were two distinct solutions for B, we could continue them back along the path by Lemma 3.4.9 to two distinct solutions for the circle, contradicting Lemma 3.4.8. This proves the theorem.

BIBLIOGRAPHY

1. A. Acker, *A free boundary optimization problem*. I, Siam J. Math. Anal. **9** (1978), 1179–1191; II, **11** (1980), 201–209.

2. J. T. Beale, *The existence of solitary water waves*, Comm. Pure Appl. Math. **30** (1977), 373–389.

3. N. Desolneux-Moulis, *Theorie du degre dans certains espaces de Fréchet, d'apres R. S. Hamilton*, Bull. Soc. Math. France, Mem. 46, Journees sur la Geometrie de la Dimension Infinie, 1976, p. 191.

4. R. S. Hamilton, *Deformation of complex structures on manifolds with boundary*.

I. *The stable case*, J. Differential Geometry **12** (1977), 1–45.

II. *Families of non-coercive boundary value problems*, J. Differential Geometry **14** (1979), 409–473.

III. *Extension of complex structures*, preprint, Cornell Univ.

5. _____, *Deformation theory of foliations*, preprint, Cornell Univ.

6. Lars Hörmander, *The boundary problems of physical geodesy*, Arch. Rational Mech. Anal. **62** (1976), 1–52.

7. _____, *Implicit function theorems*, lecture notes, Stanford Univ., 1977.

8. H. Jacobowitz, *Implicit function theorems and isometric embeddings*, Ann. of Math. (2) **95** (1972), 191–225.

9. M. Kuranishi, *Deformations of isolated singularities and* $\bar{\partial}_b$, J. Differential Geometry.

10. O. A. Ladyzenskaja and N. N. Ural'ceva, *Equations aux derivees partielles de type elliptique*, Dunod, Paris, 1968, p. 450.

11. Martin Lo, *Isometric embeddings and deformations of surfaces with boundary in* R^3, thesis, Cornell Univ., 1981.

12. S. Lojasiewics, Jr. and E. Zehnder, *An inverse function theorem in Fréchet spaces*, preprint.

13. J. Moser, *A rapidly convergent iteration method and non-linear differential equations*, Ann. Scuola Norm. Sup. Pisa (3) **20** (1966), 499–535.

14. J. Nash, *The embedding problem for Riemannian manifolds*, Ann. of Math. (2) **63** (1956), 20–63.

15. L. Nirenberg, *The Weyl and Minkowski problems in differential geometry in the large*, Comm. Pure Appl. Math. **6** (1953), 337–394.

16. _____, *An abstract form of the non-linear Cauchy-Kowalewski Theorem*, J. Differential Geometry **6** (1972), 561–576.

17. P. H. Rabinowitz, *Periodic solutions of non-linear hyperbolic partial differential equations*. I, Comm. Pure Appl. Math. **20** (1967), 145–205; II, **22** (1968), 15–39.

18. David G. Schaeffer, *The capacitor problem*, Indiana Univ. Math. J. **24** (1974/75), 1143–1167.

19. _____, *A stability theorem for the obstacle problem*, Adv. in Math. **17** (1975), 34–47.

20. F. Sergeraert, *Une generalization du théorème des fonctions implicites de Nash*, C. R. Acad. Sci. Paris Ser. A **270** (1970), 861–863.

21. _____, *Un théorème de fonctions implicites sur certains espaces de Fréchet et quelques applications*, Ann. Sci. École Norm. Sup., 4^e serie, **5** (1972), 599–660.

22. _____, *Une extension d'un théorème de fonctions implicites de Hamilton*, Bull. Soc. Math. France, Mem. 46, Journees sur la Geometrie de la Dimension Infinite, 1976, p. 191.

23. M. Spivak, *A comprehensive introduction to differential geometry*, vol. 5, Publish or Perish, Berkeley, Calif., 1979.

24. E. Zehnder, *Generalized implicit function theorems with applications to some small divisor problems*. I, Comm. Pure Appl. Math. **28** (1975), 91–141; II, **29** (1966), 49–113.

25. _____, *Moser's implicit function theorem in the framework of analytic smoothing*, preprint.

DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY, ITHACA, NEW YORK 14853