NOTES ON THE ISOMETRIC EMBEDDING PROBLEM AND THE NASH-MOSER IMPLICIT FUNCTION THEOREM

BEN ANDREWS

Contents

1. Introduction and summary

1.1. **General remarks.** The isometric embedding problem originates from the historical development of differential geometry: Early works considered the relatively concrete situation of curves and surfaces in space, and submanifolds of Euclidean spaces of higher dimension. The more abstract notion of a Riemannian manifold arose later, following Gauss's Theorema Egregium stating that the Gauss curvature of a surface depends only on the induced metric, and Riemann's work extending this to higher dimensions and developing the intrinsic geometry associated with a metric tensor prescribed in local coordinates.

Schläfli [44] discussed in 1873 the (local) question of whether a metric given in local coordinates always comes from an embedding into some Euclidean space, and conjectured that it should be possible to do so into $\mathbb{R}^{\frac{n(n+1)}{2}}$. This dimension seems plausible, since the number of independent components of the metric tensor at each point is equal to $\frac{n(n+1)}{2}$.

The local question (at least for real-analytic metrics) was solved for the 2-dimensional case in 1926 by Janet [25] , and Cartan [4] extended this to all n in 1927 as an application of his work on exterior differential systems. I should point out that the corresponding problem for C^k metrics is quite different — there is a counterexample due to Pogorelov [43], giving a $C^{1,1}$ metric on the plane which cannot be locally isometrically embedded into $\mathbb{R}^3$. If the metric has positive or negative curvature at a point, then local isometric embedding is possible, and C. S. Lin has proved that it is still possible if the curvature near the point is non-negative [29] or if the curvature is zero but the derivative is non-zero [30]. A very recent preprint of Nadirashvilli gives an example of a C^∞ metric which cannot be locally isometrically embedded into $\mathbb{R}^3$, which seems to close the question.

The question of globally embedding a manifold is a natural extension of the local question, but could not have been formulated precisely until after Weyl's precise definition of differentiable manifolds [48] in 1912, brought into common use after the work of Whitney in the 1930s. Whitney ([50]–[52]) proved that any compact manifold of dimension n can be embedded (without requiring isometry) into $\mathbb{R}^{2n}$, and immersed into $\mathbb{R}^{2n-1}$.

The general result was finally proved by John Nash [37] in 1954 using methods that seem to be entirely without precedent. He showed that any compact manifold with a metric of class C^k, $k \geq 3$, can be isometrically embedded in $\mathbb{R}^N$ where $N = \frac{n(3n+11)}{2}$. The dimension requirement has been gradually reduced over the years, particularly through work of Gromov [9], who proved one can take $N = n^2+10n+3$ for $k > 2$, or $N = (n+2)(n+3)/2$ if $k \geq 4$.

The hard analytic part of Nash's proof was taken up by others and fashioned into a more general theorem (or method) now called the Nash-Moser implicit function theorem. This was done by several authors including J. Schwartz ([45],[46]), J. Moser ([33]–[34]), L. Nirenberg [40], L. Hörmander [21]–[23], H. Jacobowitz [24], E. Zehnder [54]–[55], and R. Hamilton [16]. The results apply to a range of problems, including the solution of a wide variety of nonlinear elliptic and parabolic equations, and most famously to the proof of the KAM theorem on existence of invariant tori in Hamiltonian systems obtained by perturbations of integrable systems.

There is an interesting postscript to this story: Matthias Günther [12]–[14] discovered in around 1987 that one can circumvent the difficulties which Nash encountered, so that the remarkable Nash-Moser iteration method is not required. Using this observation he achieves isometric embeddings into Euclidean space of dimension $N = \max\{\frac{n(n+3)}{2} + 5, \frac{n(n+5)}{2}\}$.

Note that in the case $n = 2$, Nash gives an isometric embedding of a compact surface into $\mathbb{R}^{17}$, Gromov (and Günther) into $\mathbb{R}^{10}$. Gromov used different methods particular to the two-dimensional case to show that every compact surface isometrically embeds in $\mathbb{R}^5$. This cannot be improved in general, since the standard metric on the real projective plane cannot be embedded isometrically in $\mathbb{R}^4$. Conceivably it might be possible to reduce this to $\mathbb{R}^4$ for oriented surfaces, but again no better since compact surfaces with non-positive curvature cannot be embedded in $\mathbb{R}^3$. Spheres with non-negative curvature can always be embedded in $\mathbb{R}^3$ as ovaloids (boundaries of convex bodies), thanks to the Weyl embedding problem proved by Weyl [49], Lewy [28], Aleksandrov [1], Pogorelov [42], Nirenberg[39], E. Heinz [17]–[18], P.-F. Guan and Y.-Y. Li [15].

The plan is to start by outlining Nash's proof of the isometric embedding problem, which includes a number of good ideas beyond the perturbation result: Important aspects of this are setting up a framework so that the local perturbation result can be applied, and proving existence of approximate isometric embeddings. Then we will return to the proof of the perturbation result, by two different methods: First using a method of Hörmander which essentially provides a model for the Nash-Moser method, and a second, more special to the isometric embedding problem, due to Günther.

1.2. **The isometric embedding problem.** Let (M, g) be a (compact) Riemannian manifold of dimension n. Given any map $F = (F^1, \ldots, F^N) : M \to \mathbb{R}^N$, there is an induced metric tensor on M given in any local coordinates by

$$(g_F)_{ij} = \frac{\partial F}{\partial x^i} \cdot \frac{\partial F}{\partial x^j} = \sum_{r=1}^{N} \frac{\partial F^r}{\partial x^i} \frac{\partial F^r}{\partial x^j}.$$

This is a Riemannian metric provided F is an immersion. The isometric embedding problem, simply stated, is to find a one-to-one function F such that $g_F = g$.

1.3. **Perturbation of embeddings.** Nash's strategy, which is also the strategy of later authors, is to consider the problem of perturbing a given isometric immersion (or embedding) to achieve some desired (suitably small) change in the metric.

Suppose h is the desired change in the metric (i.e. a symmetric tensor on M). Then we can try to choose a map $V : M \to \mathbb{R}^N$ such that $g_{F+V} = g_F + h$, which means

$$\sum_{r=1}^{N} \frac{\partial F^r}{\partial x^i} \frac{\partial V^r}{\partial x^j} + \sum_{r=1}^{N} \frac{\partial V^r}{\partial x^i} \frac{\partial F^r}{\partial x^j} + \sum_{r=1}^{N} \frac{\partial V^r}{\partial x^i} \frac{\partial V^r}{\partial x^j} = h_{ij}.$$

This gives a first order system of partial differential equations which must be satisfied by the variation V. Nash simplified this to some extent by considering variations that are normal to the embedding F, which means imposing the extra equations

$$\sum_{r=1}^{N} V^r \frac{\partial F^r}{\partial x^j} = 0$$

for $j = 1, \dots, n$. The simplification results from differentiating this equation with respect to x^i, to give

$$\sum_{r=1}^{N} \frac{\partial V^r}{\partial x^i} \frac{\partial F^r}{\partial x^j} = -\sum_{r=1}^{N} V^r \frac{\partial^2 F^r}{\partial x^i \partial x^j}.$$

Substituting this into the perturbation equation gives the rather simpler result

$$-2\sum_{r=1}^{N} V^r \frac{\partial^2 F^r}{\partial x^i \partial x^j} + \sum_{r=1}^{N} \frac{\partial V^r}{\partial x^i} \frac{\partial V^r}{\partial x^j} = h_{ij}.$$

Note that the last term on the left is quadratic in V, so if V is small then this should be insignificant. The key observation is that the remaining terms form an (algebraic, not differential) linear system of equations for the components of V. In particular, if we write down the corresponding infinitesimal problem coming from considering the above perturbation problem for th_{ij} as $t \to 0$, then the infinitesimal variation $W = \frac{\partial V}{\partial t}|_{t=0}$ satisfies the equations

$$\sum_{r=1}^{N} W^r \frac{\partial^2 F^r}{\partial x^i \partial x^j} = -\frac{1}{2} h_{ij}$$

and

$$\sum_{r=1}^{N} W^r \frac{\partial F^r}{\partial x^j} = 0.$$

We therefore have at each point of M a system of $n(n+3)/2$ linear equations in the N unknowns W^r (n of these come from the normal variation condition, and the remaining $n(n+1)/2$ from the equation for each component of the symmetric tensor h_{ij}). Clearly the system cannot be solved in general if $N < \frac{n(n+3)}{2}$, while if $N > \frac{n(n+3)}{2}$ then any solution will be non-unique. If $N \geq \frac{n(n+3)}{2}$ then a solution exists provided the $n(n+3)/2$ vectors

$$\frac{\partial F}{\partial x^i}, \qquad i = 1, \dots, n$$

and

$$\frac{\partial^2 F}{\partial x^i \partial x^j}, \qquad 1 \leq i \leq j \leq n$$

are linearly independent.

1.4. **Freeness of embeddings and immersions.** A map satisfying this condition everywhere is called a *free* immersion. Recall the definition of the second fundamental form,

$$\frac{\partial^2 F}{\partial x^i \partial x^j} = -I\!I^\alpha_{ij}\nu_\alpha + \Gamma_{ij}{}^k \frac{\partial F}{\partial x^j},$$

where ν_α, $\alpha = 1, \dots N - n$ are a basis for the normal space of the map F, and $\Gamma_{ij}{}^k$ are the connection coefficients. From this we see that the condition that F is free is equivalent to saying that the second fundamental form is an injective map from the bundle of symmetric 2-tensors on M to the normal bundle of M at each point. In particular, freeness is independent of the choice of local coordinates.

If F is free, then there exists a solution W of the infinitesimal perturbation problem. If $N > \frac{n(n+3)}{2}$ then this solution is not unique, but we can pick out a preferred solution by asking that W have the shortest length possible at each point. The solution W at each point is unique up to an element of the orthogonal complement of $TM \oplus \mathrm{span} I\!I$, so the shortest length is achieved precisely when W is in $\mathrm{span} I\!I$. Then we can write

$$W = \frac{1}{2}\left(G^{-1}\right)^{ij,kl} h_{ij} I\!I_{kl}{}^\alpha \nu_\alpha.$$

Here

$$G_{ij,kl} = I\!I_{ij}{}^\alpha I\!I_{kl}{}^\beta \nu_\alpha \cdot \nu_\beta,$$

and by assumption $G_{ij,kl}$ is a positive definite bilinear form on the space of symmetric $(0,2)$-tensors, and so has an inverse G^{-1} which is a positive definite bilinear form on the space of symmetric $(2,0)$-tensors. Note that G^{-1} is a rational function of the coefficients of G, which is itself quadratic in the components of the second fundamental form. It follows that W is as regular as h and $I\!I$ are.

1.5. **Nash's perturbation result.** I will now state Nash's perturbation result, but defer the proof until later (this is the part of the proof which contains the hard analysis).

Theorem 1.1. *Let M be a compact manifold with a free real-analytic embedding F into $\mathbb{R}^N$. If h is a C^k symmetric $(2,0)$-tensor field on M with $k \geq 3$, which is sufficiently small in C^3, then there exists a C^k map $V : M \to \mathbb{R}^N$ such that $g_{F+V} = g_F + h$.*

Thus we can perturb about real-analytic free embeddings. In fact real-analyticity is not at all necessary, we can take F to be C^∞, or less regular if the metric we are trying to attain is less regular (see the precise statements in Lectures 5–6). Also, the closeness condition can be weakened to $C^{2,\alpha}$ instead of C^3.

The regularity of V given in the Theorem seems at first sight to be worse than could be expected — roughly speaking, the metric g is constructed from first derivatives of the embedding F, so if the metric is C^k then we might expect the embedding to be C^{k+1}. However this is not true in general, and the regularity cannot be improved without further assumptions.

To see this, consider the expressions for the intrinsic curvature tensor of the induced metric. This can be computed from the metric tensor itself, as follows: By definition, the covariant derivatives of the coordinate vector field $\partial_i = \frac{\partial}{\partial x^i}$ are given by

$$\nabla_{\partial_i}\partial_k = \Gamma_{ik}{}^p\partial_p,$$

where the Christoffel symbol $\Gamma_{ik}{}^p$ is given by

$$\Gamma_{ik}{}^p = \frac{1}{2}g^{pq}\left(\frac{\partial}{\partial x^i}g_{kq} + \frac{\partial}{\partial x^k}g_{iq} - \frac{\partial}{\partial x^q}g_{ik}\right).$$

The curvature tensor is then given by

$$\begin{aligned} R_{ijkl} &= g\left(\nabla_{\partial_j}\nabla_{\partial_i}\partial_k - \nabla_{\partial_i}\nabla_{\partial_j}\partial_k, \partial_l\right) \\ &= \left(\frac{\partial}{\partial x^j}\Gamma_{ik}{}^q - \frac{\partial}{\partial x^i}\Gamma_{jk}{}^q + \Gamma_{ik}{}^p\Gamma_{jp}{}^q - \Gamma_{jk}{}^p\Gamma_{ip}{}^q\right) g_{ql} \end{aligned}$$

This involves second derivatives of the metric tensor, so can be expected to be C^{k-2} if the metric is C^k.

Alternatively, we can compute the intrinsic curvature from the extrinsic curvature via the Gauss equation:

$$R_{ijkl} = \left(I\!I^{\alpha}_{ik}I\!I^{\beta}_{jl} - I\!I^{\alpha}_{jk}I\!I^{\beta}_{il}\right)\nu_\alpha \cdot \nu_\beta.$$

The second fundamental form is defined in terms of second derivatives of the embedding, and so is no worse than C^{k-1} if the embedding is C^{k+1}. Therefore to show the embedding cannot be C^{k+1}, we simply need to find a metric for which the intrinsic curvature is indeed no more regular than C^{k-2} (it is possible that one could get miraculous cancellations in the first expression so that the result was in fact C^{k-1} for a C^k metric). In two dimensions, take the metric $g = e^{2f(x)}\left(dx^2 + dy^2\right)$, where f is C^k. Then we find the scalar curvature is given by

$$R = -e^{-2f}f''$$

which is no better than C^{k-2}. In higher dimensions take the product of this with a flat metric.

1.6. **Loss of differentiability.** I will try to indicate where the difficulties lie in proving the perturbation result. We have set up the equations for the perturbation problem, and showed (at least if the immersion we are perturbing about is free) that there exists a solution of the infinitesimal problem. Usually in such circumstances we would hope to apply an implicit function theorem to show that there is in fact a solution.

Let us formalise things a little more: We have a fixed starting embedding F which is free and can be assumed to be quite regular (even real analytic). Consider the map which takes a C^k section V of the normal bundle of $F(M)$ to the C^{k-1} symmetric tensor $h = g_{(F+V)} - g_F$. This is a smooth map from the Banach space $C^k(NM)$ to the Banach space $C^{k-1}(S_2M)$, where S_2M is the bundle of symmetric 2-tensors on M. We have computed the derivative of this map about the zero section. It looks like we have just shown that the derivative is surjective, but look more closely: What we have shown is that any infinitesimal perturbation h of the metric can be obtained by some infinitesimal variation W in the normal bundle, but our expression above shows that if h is C^{k-1} then the variation W is also C^{k-1}, not C^k. So in fact the derivative of the above map between Banach spaces is not surjective, and we cannot apply the implicit function theorem for Banach spaces to find a local inverse for the map. Instead we have shown that the derivative maps onto the smaller subspace of C^k variations of the metric, but that is no good because the map does not give us a C^k variation of the metric in general.

This is the phenomenon of loss of differentiability which is the key analytic difficulty which Nash managed to overcome, and which is addressed in the Nash-Moser implicit function theorem, or 'hard implicit function theorem' as it is also known.

2. Setting up the isometric embedding

In this section I will show how the local perturbation result can be used to prove the global isometric embedding theorem. The results here are largely geometric, and involve a number of nice tricks.

2.1. **Difficulties in applying the perturbation result.** In order to apply the local perturbation result to obtain an isometric embedding of a given metric, the obvious thing to try to do is find embeddings for which the induced metric is close to the given one. But this is a tall order: We would need the embedding to be real analytic, and the induced metric would have to be sufficiently close in C^3 to the desired one so that we could apply the perturbation result. Unfortunately, 'sufficiently close' is not spelled out. If one inspects the proof of the perturbation result, it becomes clear that 'sufficiently close' depends on some estimate for the freeness of the initial embedding. But the approximate isometric embeddings that we will construct (in the next lecture) have very poor control on their freeness: These are obtained by 'twisting' a collection of maps around very tight circles, with better approximations being produced by tighter and tighter circles. So to get a good approximation to the metric, the embedding will typically have very large second fundamental form.

2.2. **Nash's y and z embeddings.** Nash uses the following trick to get around the problem, at the expense of increasing the dimension of the Euclidean space we embed into: Suppose we have two embeddings, F_y and F_z, into Euclidean spaces $\mathbb{R}^N$ and $\mathbb{R}^{N'}$. Then consider the map $(F_y, F_z) : M \to \mathbb{R}^{N+N'}$. The induced metric of this is equal to $g_{F_y} + g_{F_z}$.

The idea is this: First choose an embedding F_z, which Nash calls the 'z-embedding', which is real analytic and free, and so can be perturbed locally to get any nearby metric which is sufficiently close in C^3 (here 'sufficiently close' means within some fixed distance which will not change from now on, since we will always be perturbing about this fixed embedding). By scaling, ensure that the induced metric g_{F_z} is strictly less than the desired metric g (Gromov and Rokhlin call this a 'strictly short' embedding).

Then we try to choose an embedding F_y (the 'y-embedding') for which the induced metric is close to $\tilde{g} = g - g_{F_z}$ — in fact, what we need is that it is 'sufficiently close' in C^3 to $\tilde{g}$, in precisely the sense of the previous paragraph.

It is clear that this would suffice to prove the existence of an isometric embedding. This leaves us with two problems to tackle: Approximate isometric embeddings, and existence of free embeddings.

2.3. **Existence of free embeddings.** We will prove the following:

Theorem 2.1. *A compact manifold M of dimension n has a C^∞ free embedding into $\mathbb{R}^N$, where $N = \frac{n(n+5)}{2}$.*

This was proved by Nash [37], but we will use different methods following Gromov-Rokhlin [11], with an argument essentially following the proof of Whitney's 'easy' embedding theorem [50]. Let us recall this first:

Theorem 2.2. *A compact manifold M of dimension n has a C^∞ embedding into $\mathbb{R}^{2n+1}$ and an immersion into $\mathbb{R}^{2n}$.*

Whitney's proof follows the following steps: First, show that M can be embedded into some Euclidean space, without controlling the dimension. Then show that the projection of the resulting submanifold onto some space of one dimension lower is an embedding (or immersion) if the dimension is not too small.

The first step is easy: Take a finite cover of M by charts $x_i = (x_i^1, \ldots, x_i^n) : U_i \to B_1(0) \subset \mathbb{R}^n$, $i = 1, \ldots, r$, such that the smaller sets $W_i = x_i^{-1}(B_{1/3}(0))$ also cover M. Let f be a smooth function on $\mathbb{R}^n$ which is identically 1 on $B_{1/3}(0)$, identically zero outside $B_{2/3}(0)$, and strictly between 0 and 1 on $B_{2/3}(0)\backslash\overline{B_{1/3}(0)}$. Then for each i, $f \circ x_i$ extends by zero to a smooth function on M. Define $F : M \to \mathbb{R}^{r(n+1)}$ by

$$F = (f \circ x_1, \ldots, f \circ x_r, x_1 f \circ x_1, x_2 f \circ x_1, \ldots, x_r f \circ x_r).$$

F is one-to-one, since if $F(x) = F(y)$ then there is some i such that $x \in W_i$, but then $f \circ x_i(x) = f \circ x_i(y) = 1$, and therefore by definition of f, $y \in W_i$. But also $x_i(x) = x_i(y)$, so $x = y$ since x_i is one-to-one on W_i. F is an immersion, since in W_i F has as some of its components $x_i f \circ x_i = x_i$, which has derivative in the chart x_i equal to the identity. Therefore F is an embedding.

It remains to prove that if $N > 2n + 1$ and M^n is a compact submanifold of $\mathbb{R}^N$, then there is some $v \in S^{N-1}$ such that the orthogonal projection π_v onto the $(N-1)$-dimensional subspace orthogonal to v, given by

$$\pi_v(x) = x - (x \cdot v)v$$

is an embedding on M. Similarly in $N > 2n$ then there is some v such that π_v is an immersion on M.

To see this, consider the map h from $M \times M\backslash\Delta$, where Δ is the diagonal, given by

$$h(x, y) \mapsto \frac{x - y}{|x - y|}.$$

π_v is one-to-one provided h never takes the value v. Also consider the map k from the unit sphere bundle $SM = \{(p, w) : \ p \in M,\ w \in T_pM,\ |w| = 1\}$ to S^{N-1} given by

$$k(p, w) = w.$$

Then π_v is an immersion provided k never takes the value v.

Note that $M \times M\backslash\Delta$ is a manifold of dimension $2n$, while SM is a manifold of dimension $2n - 1$. We use the fact that a smooth map from a manifold M to a manifold N has image of measure zero if N has larger dimension than M. It follows that there exists a point $v \in S^{N-1}$ which is not in the image of h or k provided $N - 1 > \max\{2n, 2n - 1\}$ (i.e. $N > 2n + 1$) and there is $v \in S^{N-1}$ not in the image of k provided $N > 2n$.

This completes the proof of Theorem 2.2. Note that this gives a C^r embedding for $r = 1, \ldots, \infty$, but does not extend to the real-analytic case. Whitney [53] proved that any C^r manifold carries a compatible real-analytic structure and has a real-analytic embedding for that structure. The question of whether a manifold with a given real-analytic structure has a real-analytic embedding was not resolved until later, by Morrey in 1958 [32], after Nash's proof of the isometric embedding theorem. However, the case of interest to the isometric embedding problem, that of embedding a real-analytic manifold which carries a real-analytic Riemannian metric, was proved by Bochner [2] in 1937 (it is a consequence of Morrey's theorem that this is the general case!).

To prove Theorem 2.1 we use the same steps: First find a free embedding into some Euclidean space of large dimension, then show that the dimension can be reduced if it is too large.

To accomplish the first step, first note that we can find a free embedding of an open set in $\mathbb{R}^n$ into $\mathbb{R}^{\frac{n(n+3)}{2}}$ as follows: Take an orthonormal basis $\{e_i\}_{1\leq i\leq n} \cup \{e_{ij}\}_{1\leq i\leq j\leq n}$ for $\mathbb{R}^{\frac{n(n+3)}{2}}$, and define

$$F(x^1, \ldots, x^n) = \sum_{i=1}^{n} x^i e_i + \sum_{1\leq i\leq j\leq n} x^i x^j e_{ij}.$$

This is clearly an embedding since the first n components are. To see that it is free, note that

$$e_i = \frac{\partial F}{\partial x^i} - \sum_{j=1}^{n} x^j \frac{\partial^2 F}{\partial x^i \partial x^j}; \quad e_{ij} = \frac{\partial^2 F}{\partial x^i \partial x^j}, \quad (i \neq j); \quad e_{ii} = \frac{1}{2}\frac{\partial^2 F}{(\partial x^i)^2}.$$

Therefore the first and second derivatives of F span the whole space, and hence give an isomorphism at each point.

The first step is easy, given Theorem 2.2, since we can embed M as a submanifold of some Euclidean space, then take a free embedding of that Euclidean space in a higher-dimensional Euclidean space. The restriction of a free map to a submanifold is clearly free.

It remains to prove that some projection π_v is a free embedding if the dimension is large. The embeddedness condition holds provided v is not in the image of the maps h and k defined above.

Define the 2-jet space $J^2_p M^*$ at $p \in M$ to be the set of equivalence classes of germs of smooth functions about p, where two function germs are equivalent if their first and second derivatives agree at p in any chart. This is a vector space of dimension $\frac{n(n+3)}{2}$, and has a natural basis in any chart $(x^1, \ldots, x^n)$ for M about p, given by the equivalence classes of the functions x^i for $i = 1, \ldots n$ and $x^i x^j$ for $1 \leq i \leq j \leq n$.

There is a corresponding dual basis for the dual space J^2_pM, which we denote e_i, $1 \le i \le n$ and e_{ij}, $1 \le i \le j \le n$.

Let SJ^2M be the unit sphere bundle of J^2M. This is a manifold of dimension $n + \frac{n(n+3)}{2} - 1$. Consider the map $j : SJ^2M \to S^{N-1}$ given by restricting the map above. Then it is clear that π_v is a free map provided v is not in the image of j, so π_v is a free embedding provided v is not in the image of h, k or j. This can be guaranteed for some v as long as the target has higher dimension than the source, which means $N - 1 > \max\{2n, 2n-1, n + \frac{n(n+3)}{2} - 1\}$, which means $N \ge n + \frac{n(n+3)}{2} = \frac{n(n+5)}{2}$.

This proves Theorem 2.1. The proof works without change to give C^r free embeddings of C^r manifolds, $r \ge 1$, and also for real-analytic manifolds as long as we assume Morrey's embedding result.

Other methods (more closely related to those of Whitney's first proof) give a somewhat more powerful result: First consider the simpler embedding and immersion problems: The condition that a map F be an immersion can be expressed in terms of its 1-jet J^1F, which is a section of the 1-jet bundle $j^1(M, \mathbb{R}^N) \simeq \bigoplus^N T^*M$. The requirement is that the one-jet avoid the submanifolds A_k consisting of 1-jets which are rank k, for each $k = 0, \dots, n-1$. The largest of these is A_{n-1}, which has dimension $nN + n - 1 - N$ at each point, so dimension $nN + 2n - 1 - N$ within the 1-jet bundle. If the dimensions of the section j^1F and A_{n-1} sum to less than the total dimension of the 1-jet space, then transversality implies that they are disjoint. This is true provided $n + nN + 2n - 1 - N < n + nN$, which means $N > 2n - 1$, or $N \ge 2n$. The transversality theorem (see [20]) then implies that the set of C^k immersions is *residual* in $C^k(M, \mathbb{R}^N)$, which means it is an intersection of open dense sets, which by the Baire category theorem is dense. Thus every C^k map into $\mathbb{R}^N$ can be approximated in C^k by immersions if $N \ge 2n$.

Similarly, the one-to-one condition amounts to the map from $M \times M \backslash \Delta$ to $\mathbb{R}^N$ given by $(x, y) \mapsto F(x) - F(y)$ avoiding zero, which is generically true if its dimension is less than N, so $N > 2n$ or $N \ge 2n + 1$. Thus any C^k map from M into $\mathbb{R}^N$ with $N \ge 2n + 1$ can be approximated in C^k by embeddings.

Now turn to the case of free maps: In this case we require that the 2-jet of F avoid certain submanifolds in the 2-jet bundle, so this holds generically provided the submanifold has codimension greater than n. This submanifold consist of 2-jets which have rank k, for $k = 0, \dots, \frac{n(n+3)}{2} - 1$. The largest of these has dimension $\frac{n(n+3)}{2} - 1 + N\frac{n(n+3)}{2} - N$ at each point, so our requirement becomes $2n + \frac{n(n+3)}{2} -$

$1+N\frac{n(n+3)}{2}-N<n+N\frac{n(n+3)}{2}$, or $N\geq\frac{n(n+5)}{2}$. So for N in this range, any C^k map from M to $\mathbb{R}^N$ (with $k>2$) can be approximated in C^k by free embeddings.

It is interesting to note that Whitney improved the result of Theorem 2.2 much later, in 1944, to give [51] an embedding into $\mathbb{R}^{2n}$ and [52] an immersion into $\mathbb{R}^{2n-1}$ (for $n>1$). These results are much more difficult than the earlier ones (they are known as the 'hard' Whitney embedding theorems). It seems plausible that methods similar to this later work of Whitney (particularly that on immersions) might give a free embedding into a lower dimension than the proof above produced. However, in general no such improvement is possible: Eliashberg [5] showed that if $n=2^{k+1}$ with $k\geq 1$, then $\mathbb{R}P^{2^k}\times\mathbb{R}P^{2^k}$ cannot be freely mapped into $\mathbb{R}^{\frac{n(n+5)}{2}-1}$.

Some improvement in embedding dimension for particular manifold dimensions may be possible by the following approach: There are topological characterisations of when a manifold M^n can be immersed in $\mathbb{R}^{n+k}$ for $k<n$, due to Hirsch [19], and sometimes called the Smale-Hirsch Theorem. Consider $GL(n)$ acting on the space $V_{n,n+k}$ of n-frames in $\mathbb{R}^{n+k}$ in the obvious way. Associated to this action there is a bundle B with fibre given by $V_{n,n+k}$, defined by $B=(F(M)\times V_{n,n+k})/GL(n)$, where $F(M)$ is the frame bundle of M and $GL(n)$ acts separately on each factor. The theorem states that M^n can be immersed into $\mathbb{R}^{n+k}$ (with $k\geq 1$) if and only if B has a non-vanishing section. Note that this condition is equivalent to the existence of some k-dimensional vector bundle B' over M such that $TM\oplus B'$ is trivial. Hirsch shows in particular that every compact 3-manifold can be immersed in $\mathbb{R}^4$ (since 3-manifolds are parallelizable), and that every compact 5-manifold can be immersed in $\mathbb{R}^8$. Eliashberg and Gromov [6] prove that a manifold M^n can be freely mapped into $\mathbb{R}^{\frac{n(n+3)}{2}+k}$ (with $k\geq 1$) if and only if there is a bundle P over M of dimension k such that $TM\oplus S_2M\oplus P$ is trivial, where S_2M is the bundle of symmetric 2-tensors on M.

3. Approximate isometric embeddings

In this section we continue the process of setting up the isometric embedding problem by constructing embeddings which are approximately isometric. The argument we give yields an isometric embedding into a high-dimensional Euclidean space, modulo the local perturbation result. At this stage the dimension required depends on the metric g and not only on the dimension n of the manifold, but this will be corrected in the next section.

3.1. **The Nash Twist.** Here is another one of Nash's good ideas, which makes the construction of approximate embeddings fairly easy. The idea is the following: Suppose we can express the desired C^r metric g in the form

$$g_{ij} = \sum_{k=1}^{m} (a^k)^2 \frac{\partial f_k}{\partial x^i} \cdot \frac{\partial f_k}{\partial x^j} \tag{3.1}$$

where $a^k \in C^r(M)$ is positive and f_k is C^∞ (or analytic) for $k = 1, \ldots, m$. Then define a map $y_\lambda : M \to \mathbb{R}^{2m}$ as follows:

$$\begin{aligned} y_\lambda^k &= \frac{a^k}{\lambda} \sin(\lambda f_k), \quad k = 1, \ldots, m; \\ y_\lambda^{m+k} &= \frac{a^k}{\lambda} \cos(\lambda f_k), \quad k = 1, \ldots, m. \end{aligned}$$

Roughly speaking, the map y_λ takes each component of the map $f = (f_1, \ldots, f_m)$ and winds it around a circle with radius λ^{-1}, then scales the result by the weight a_k. If λ is large, then a^k is close to constant on each traverse of the circle, so the speed of motion is approximately a^k times the rate of change of f_k along any curve in M. Computing more precisely, the induced metric $g_\lambda = g_{y_\lambda}$ given by

$$
\begin{aligned}
(g_\lambda)_{ij} &= \sum_{k=1}^m \left(\frac{\partial y_\lambda^k}{\partial x^i} \frac{\partial y_\lambda^k}{\partial x^j} + \frac{\partial y_\lambda^{m+k}}{\partial x^i} \frac{\partial y_\lambda^{m+k}}{\partial x^i} \right) \\
&= \sum_{k=1}^m \Bigg((a^k)^2 \cos^2(\lambda f_k) \frac{\partial f_k}{\partial x^i} \frac{\partial f_k}{\partial x^j} \\
&\qquad + \frac{a^k \sin(\lambda f_k)\cos(\lambda f_k)}{\lambda} \left(\frac{\partial a^k}{\partial x^i} \frac{\partial f_k}{\partial x^j} + \frac{\partial a^k}{\partial x^j} \frac{\partial f_k}{\partial x^i} \right) \\
&\qquad + \frac{\sin^2(\lambda f_k)}{\lambda^2} \frac{\partial f^k}{\partial x^j} \frac{\partial f_k}{\partial x^i} \\
&\qquad + (a^k)^2 \sin^2(\lambda f_k) \frac{\partial f_k}{\partial x^i} \frac{\partial f_k}{\partial x^j} \\
&\qquad - \frac{a^k \sin(\lambda f_k)\cos(\lambda f_k)}{\lambda} \left(\frac{\partial a^k}{\partial x^i} \frac{\partial f_k}{\partial x^j} + \frac{\partial a^k}{\partial x^j} \frac{\partial f_k}{\partial x^i} \right) \\
&\qquad + \frac{\cos^2(\lambda f_k)}{\lambda^2} \frac{\partial a^k}{\partial x^j} \frac{\partial a^k}{\partial x^i} \Bigg) \\
&= \sum_{k=1}^m \left((a^k)^2 \frac{\partial f_k}{\partial x^i} \frac{\partial f_k}{\partial x^j} + \frac{1}{\lambda^2} \frac{\partial a^k}{\partial x^j} \frac{\partial a^k}{\partial x^i} \right) \\
&= g_{ij} + \frac{1}{\lambda^2} \sum_{k=1}^m \frac{\partial a^k}{\partial x^j} \frac{\partial a^k}{\partial x^i}.
\end{aligned}
$$

If we take λ large, this is a good approximation for g in C^{r-1}.

3.2. **Applying the Nash Twist.** To make use of this observation, we need to express g in the given form. Nash's approach is the following: Construct a collection of functions f_k, $k = 1, \ldots, m$ such that the symmetric bilinear forms

$$\frac{\partial f_k}{\partial x^i} \cdot \frac{\partial f_k}{\partial x^j}$$

for $k = 1, \ldots, m$ span the space of symmetric bilinear forms at each point. Nash showed that this can be done with $m = \frac{n(n+3)}{2}$. Then any metric can be expressed as a linear combination of these (with C^r coefficients if we insist that the sum of the squared norm of the coefficients is as small as possible), and any metric which is sufficiently close (in C^0) to the metric

$$\gamma_{ij} = \sum_{i=1}^m \frac{\partial f_k}{\partial x^i} \cdot \frac{\partial f_k}{\partial x^j}$$

has coefficients which are positive in this decomposition.

This shifts some of the problem back to the construction of the free z-embedding, which must be chosen in such a way that $g - g_z$ is close to γ in C^0 to allow it to be approximated by the metric of the y-embedding.

3.3. **Existence of Full maps.** Let us now construct a collection of functions $f = (f_1, \dots, f_m)$ satisfying the requirements of the previous section, so that the metric elements $df_j \cdot df_j$, $j = 1, \dots, m$ span the space of symmetric 2-tensors at each point of M (let us agree that such a map be called *full*). This is easy if we don't care about the dimension: Let F be an immersion of M into $\mathbb{R}^N$ (we can take $N = 2n$ by the easy Whitney theorem) and take the collection of functions $f_{ij} = F_i + F_j$, $1 \leq i \leq j \leq N$. At any point of M, some n of the functions F_i (say $i = 1, \dots, n$) are suitable as local coordinates for M, and then the collection of $\frac{n(n+1)}{2}$ functions f_{ij} for $1 \leq i \leq j \leq n$ have metric elements

$$(g_{ij})_{kl} = \frac{\partial f_{ij}}{\partial x^k}\frac{\partial f_{ij}}{\partial x^l} = (\delta_{ik} + \delta_{jk})(\delta_{il} + \delta_{jl}).$$

These span the space of symmetric bilinear forms, since

$$\frac{1}{4}(g_{ii})_{kl} = \delta_{ik}\delta_{il}$$

and

$$(g_{ij})_{kl} - \frac{1}{4}(g_{ii})_{kl} - \frac{1}{4}(g_{jj})_{kl} = \delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il}.$$

Therefore a general symmetric bilinear form with coefficients a_{kl} at a point of W_α can be expressed as

$$\frac{1}{4}\sum_{i=1}^{n} a_{ii}g_{ii} + \sum_{1\leq i<j\leq n} a_{ij}\left(g_{ij} - \frac{1}{4}g_{ii} - \frac{1}{4}g_{jj}\right).$$

This gives a full map into $\mathbb{R}^{n(2n+1)}$.

A better result can be obtained using a transversality argument: The condition of fullness of a map F into $\mathbb{R}^N$ says that at each point the 1-jet of F (i.e. the derivative, locally an $n \times N$ matrix at each point of M) avoids a certain union of submanifolds in the space of 1-jets, namely the submanifolds for which the span of the metric elements have rank k, for each $k < \frac{n(n+1)}{2}$, in the space of symmetric bilinear forms. The largest of these (with $k = \frac{n(n+1)}{2} - 1$) is defined by N equations in $Nn + \frac{n(n+1)}{2} - 1$ variables, and has dimension $N(n-1) + \frac{n(n+1)}{2} - 1$. The 1-jet of F is a section of the 1-jet bundle, hence of dimension n, and we want this to avoid the submanifold of dimension $N(n-1) + \frac{n(n+1)}{2} - 1 + n$, so we ask that the sum of these dimensions be less than the dimension of the 1-jet space, which is $n + Nn$, so that transversality implies disjointness. This gives the requirement

$$N(n-1) + \frac{n(n+1)}{2} + 2n - 1 < n + Nn,$$

which means $N \geq \frac{n(n+3)}{2}$. The transversality theorem therefore implies that any C^k map from M to $\mathbb{R}^{\frac{n(n+3)}{2}}$ can be approximated in C^k by full maps (provided $k > 1$).

3.4. **Isometric embedding in high dimensions.** I will avoid using Nash's approach for now, and instead take a different approach which requires a larger dimension.

Lemma 3.1. *Let M^n be a compact C^∞ manifold, and g a C^k metric on M, $k \geq 1$. Let $F : M \to \mathbb{R}^N$ be a C^∞ immersion. Then there exists a finite collection of unit vectors $e_1, \ldots, e_r$ in $\mathbb{R}^N$ and C^k non-negative functions $a_1, \ldots, a_r$ on M such that*

$$g_{kl} = \sum_{i=1}^{r} a_i^2 \frac{\partial}{\partial x^k}(F \cdot e_i) \frac{\partial}{\partial x^l}(F \cdot e_i).$$

Proof. For each $z \in M$, g is a positive definite symmetric bilinear form, so (since all such are similar) we can choose vectors $e_1(z), \ldots, e_n(z) \in D_zF(T_zM)$ such that

$$g_{kl}(z) = \sum_{1 \leq i \leq j \leq n} \frac{\partial}{\partial x^k}(F \cdot (e_i(z) + e_j(z))) \frac{\partial}{\partial x^l}(F \cdot (e_i(z) + e_j(z)))$$

Since the bilinear forms $\frac{\partial}{\partial x^k}(F \cdot (e_i(z) + e_j(z))) \frac{\partial}{\partial x^l}(F \cdot (e_i(z) + e_j(z)))$ are a basis for the space of bilinear forms, and g is continuous, it remains true for y in a neighbourhood U_z of z that

$$g_{kl}(y) = \sum_{1 \leq i \leq j \leq n} \beta_{ij}^2(z, y) \frac{\partial}{\partial x^k}(F \cdot (e_i(z) + e_j(z))) \frac{\partial}{\partial x^l}(F \cdot (e_i(z) + e_j(z)))$$

where $\beta_{ij}(z, y)$ is positive for each $1 \leq i \leq j \leq n$ and each $y \in U_z$. Cover M by a finite number of such regions (given by some choice of $z_1, \ldots, z_m$) and choose a collection of smooth non-negative functions f_α, $\alpha = 1, \ldots, m$ with $\mathrm{supp} f_\alpha \subset U_{z_\alpha}$ and $\sum_\alpha f_\alpha^2 = 1$ everywhere. Then

$$g_{kl} = \sum_{\alpha; 1 \leq i \leq j \leq n} f_\alpha^2 \beta_{ij}^2(z_\alpha, .) \frac{\partial}{\partial x^k}(F \cdot (e_i(z_\alpha) + e_j(z_\alpha))) \frac{\partial}{\partial x^l}(F \cdot (e_i(z_\alpha) + e_j(z_\alpha)))$$

∎

We can now apply the Nash twist to get an embedding with metric approximating the desired metric g in C^{k-1}. This completes the proof of the isometric embedding theorem, at least if we don't care what dimension the embedding space should be, and modulo the proof of the local perturbation result.

3.5. **Nash's argument.** A few words on Nash's argument, which will explain where his embedding dimension comes from: First, construct a full map into $\mathbb{R}^{\frac{n(n+3)}{2}}$, and scale to make it short for g (that is, so that the induced metric γ is strictly smaller than g in every direction). Nash then wants to construct the z-embedding, which should be a free (real-analytic) embedding with metric close in C^0 to $g - \gamma$. This is done as follows: Start with any embedding (say, into $\mathbb{R}^{2n}$ as given by Whitney's theorem) which is short for $g-\gamma$. Nash proves (in an earlier paper [36]) that this can be perturbed an arbitrarily small amount in C^0 to give a C^1 isometric embedding of the metric $g - \gamma$. Now, at the expense of moving to the higher-dimensional space $\mathbb{R}^{\frac{n(n+5)}{2}}$ we can approximate the resulting embedding in C^1 by analytic free embeddings (first approximate in C^1 by a C^k map, $k > 2$, then approximate that by a C^k free embedding using the genericity result, then approximate the result in C^k by a real-analytic map — since the freeness, immersion and one-to-one conditions are open in C^2, the resulting map will be a free embedding if the last approximation is close enough). Sufficiently close C^1 approximation ensures that the metric g_z is close in C^0 to $g - \gamma$, so that $g - g_z$ is close to γ and the coefficients of $g - g_z$ are positive with respect to the full map we started with. Then the Nash twist can be used to construct the y-embedding into $\mathbb{R}^{n(n+3)}$ with arbitrarily close C^k approximation to $g - g_z$. If this approximation is good enough in C^3, then we can perturb the z embedding to give $g_z = g - g_y$, which completes the proof. The resulting embedding, given by combining the y and z embeddings, is into a Euclidean space of dimension $n(n+3) + \frac{n(n+5)}{2} = \frac{n(3n+11)}{2}$.

All we are missing to carry out this approach is the C^1-isometric embedding result. Since the idea of this is closely related to the Nash twist we have just seen, I will make a few remarks on this result and its proof.

3.6. C^1 **isometric embeddings.** The main result of the paper [36] is as follows:

Theorem 3.2. *Let (M^n, g) be a complete Riemannian manifold (g continuous), and $F : M \to \mathbb{R}^{n+k}$, $k \geq 2$ a strictly short immersion* *(embedding). Then for any* $\varepsilon > 0$ *there exists a* C^1 *immersion (embedding)* F' *with* $|F - F'|_{C^0} < \varepsilon$ *and* $g_{F'} = g$.

Kuiper [26]–[27] later improved this to allow $k \geq 1$ — thus any compact Riemannian 2-manifold can be C^1-isometrically immersed in $\mathbb{R}^3$. C^1 isometric embeddings are therefore very different animals to smoother ones — the main point being that curvature does not make sense for such immersions, so all of the usual obstructions to isometric immersion are gone.

The method of proof is as follows: We carry out a sequence of ‘stages’ in each of which we improve the aproximation to isometry, roughly decreasing the error in the metric by half while keeping the immersion strictly short.

In each stage we begin by writing the difference $g - g_F$ in the form (3.1), where each of the coefficents a_k is compactly supported in some coordinate chart (this is provided by our construction above). Then for each term in the expansion we try to do some analogue of the Nash twist to remove most of the error from that term. Instead of ‘twisting’ in $2m$ dimensions as we did above, we twist in $n + 2$ dimensions by choosing (on the support of a_k) a pair of smooth orthonormal vectors normal to M, say e_1 and e_2, and taking

$$F_\lambda = F + \frac{a^k}{\sqrt{2}\lambda}\left(\sin(\lambda f_k)e_1 + \cos(\lambda f_k)e_2\right).$$

One can check that (if F is smooth), the induced metric of F_λ is a good approximation to $g_F + \frac{1}{2}a_k^2 df_k^2$ (in C^0). Now repeat this for each term, and we have completed our first stage. The factor $\frac{1}{2}$ keeps the map strictly short, but we can be sure of removing roughly half the error. Now repeat the process indefinitely — at each stage we are left with a smooth immersion (embedding), but the smoothness deteriorates as the stages progress. However, since the metric is converging, we have control on the map in C^1.

Kuiper’s modification works by constructing ‘corrugations’ or ‘ripples’ instead of twisting around in two dimensions, which is why he needs only $k \geq 1$.

4. Smoothing operators on manifolds

This section is required to prepare for the proof of the perturbation result. Roughly speaking, the idea of the proof is to adapt Newton’s method by introducing some smoothing at each iteration step. To do this we need to devise smoothing operators which give the best possible estimates.

4.1. **The required estimates.** In the following we will fix a real-analytic embedding $\bar{F}$ of M into some Euclidean space, and let $\bar{g}$ be the induced metric. This will be used to define all notions of smoothness, including norms on C^r spaces, and so on.

We need to construct smoothing operators T_N for some parameter N on our manifold, with sufficiently good properties. Here large N corresponds to less smoothing and better approximation, while small N means more smoothing and consequently a worse approximation. The two properties we will need are the following: First, a smoothed function $T_N u$ should have derivatives of all orders, with bounds depending on lower derivatives of u:

$$|T_N u|_{C^r} \leq C N^{r-s} |u|_{C^s}, \qquad r \geq s.$$

Secondly, the approximation of the smoothed function $T_N u$ to the original one should be good in C^k if u is more regular than C^k:

$$|T_N u - u|_{C^s} \leq C N^{s-r} |u|_{C^r}, \qquad r \geq s.$$

Constructing such an operator takes some care, as we will see.

4.2. **Mollifications.** The standard way of choosing a smoothing operator is to take mollifications in coordinate charts, patched together with a partition of unity. This gives good smoothing properties, but the approximation is not as good as we require. In fact mollification does not give the properties we need, even on $\mathbb{R}$: Consider the function $u(x) = \frac{x^2}{1+x^2}$, and compute its mollifications for N large:

$$T_N u(0) = \int_{B^1(0)} \rho(y) u(y/N) dy \simeq C N^{-2}$$

for N large. Therefore $|T_N u - u|_{C^0} \geq C N^{-2}$. However u is bounded in C^3, so we should expect $|T_N u - u|_{C^0} \leq C N^{-3}$ for N large. It can be seen fairly easily that smoothing by mollification gives the desired approximation estimate only for $r - s \leq 2$.

4.3. **Reduction to the Euclidean case.** First we reduce the problem to finding suitable operators on $\mathbb{R}^d$.

The embedding $\bar{F}$ has a tubular neighbourhood V_a on which there is a smooth nearest-point projection π onto M, with positive radius a (say half of the smallest radius of curvature of the embedding). Take a C^∞ function η which is non-negative, identically equal to 1 on $V_{a/2}$, and zero outside V_a.

Let P be the operator which extends a function f on M to a compactly supported function on $\mathbb{R}^d$ by taking $Pf(y) = \eta(y) f(\pi y)$ for each point y in V_a, and $Pf(y) = 0$ outside V_a This clearly satisfies the inequalities

$$|Pf|_{C^k(\mathbb{R}^d)} \leq C|f|_{C^k(M)}.$$

Also let ι be the operator which takes a function on $\mathbb{R}^d$ to a function on M by restricting to $\bar{F}(M)$. Then we again have

$$|\iota f|_{C^k(M)} \leq C|f|_{C^k(\mathbb{R}^d)}.$$

Now suppose we have smoothing operators $\tilde{T}_N$ on $\mathbb{R}^d$ which satisfy the desired inequalities. Then we have

$$|\iota\tilde{T}_N Pu|_{C^r(M)} \leq C|\tilde{T}_N Pu|_{C^r(\mathbb{R}^d)} \leq CN^{r-s}|Pu|_{C^s(\mathbb{R}^d)} \leq CN^{r-s}|u|_{C^s(M)},$$

so the smoothing estimates hold for $T_N = \iota\tilde{T}_N P$, and

$$\begin{aligned} |T_N u - u|_{C^s(M)} &= |\iota(\tilde{T}_N - I)Pu|_{C^s(M)} \\ &\leq C|(\tilde{T}_N - I)Pu|_{C^s(\mathbb{R}^d)} \\ &\leq CN^{s-r}|Pu|_{C^r(\mathbb{R}^d)} \\ &\leq CN^{s-r}|u|_{C^r(M)}. \end{aligned}$$

4.4. **Nash's smoothing operators.** Nash's idea is to use convolution, but not with a compactly supported bump function as is normally used in mollifications. Instead we define a radially symmetric function K by taking its Fourier transform $\hat{K}$ to be a compactly supported radially symmetric C^∞ bump function, equal to a positive constant in the ball of radius 1/2, and vanishing outside the ball of radius 1. This guarantees that K is smooth and decreases rapidly at infinity, since

$$\|D^\beta x^\alpha K\|_{L^2} = \|\xi^\beta D^\alpha \hat{K}\|_{L^2} < \infty$$

for any multiindices α and β. Note also that K is real since $\hat{K}$ is even. By scaling we can ensure that $\int_{\mathbb{R}^d} K(y)\mathrm{d}y^d = 1$. The crucial point about this choice is that the resulting function has no moments, i.e. for any multiindex α with $|\alpha| > 0$,

$$\int_{\mathbb{R}^d} K(y)y^\alpha \mathrm{d}y^d = 0.$$

Next we define

$$\tilde{T}_N u(x) = \int_{\mathbb{R}^d} K(y)u(x + y/N)\mathrm{d}y^d = \int_{\mathbb{R}^d} K_N(y - x)u(y)\mathrm{d}y^n$$

where $K_N(y) = N^d K(Ny)$. Note that the Fourier transform of K_N is given by $\hat{K}_N(\xi) = \hat{K}(\xi/N)$.

4.5. **Smoothing estimates.** It is easy to see that this gives the desired smoothing properties, since we can write

$$\begin{aligned} D^\alpha \tilde{T}_N u(x) &= \int_{\mathbb{R}^d} K(y) D^\alpha u(x+y/N) \mathrm{d}y^n \\ &= (-1)^{|\beta|} N^{|\beta|} \int_{\mathbb{R}^d} D^\beta K(y) D^\gamma u(x+y/N) \mathrm{d}y^n, \end{aligned}$$

whenever $\beta + \gamma = \alpha$, and hence

$$|\tilde{T}_N u|_{C^r} \leq C N^{r-s} |u|_{C^s}$$

for integers $r \geq s$. We also have

$$\begin{aligned} &D^\alpha \tilde{T}_N u(x_2) - D^\alpha \tilde{T}_N u(x_1) \\ &\qquad = \pm N^{|\beta|} \int_{\mathbb{R}^d} D^\beta K(y) \left(D^\gamma u(x_2 + y/N) - D^\gamma u(x_1 + y/N) \right) \mathrm{d}y^d \end{aligned}$$

so that for $\sigma \in [0,1]$

$$|\tilde{T}_N u|_{C^{r,\sigma}} \leq C N^{r-s} |u|_{C^{s,\sigma}}.$$

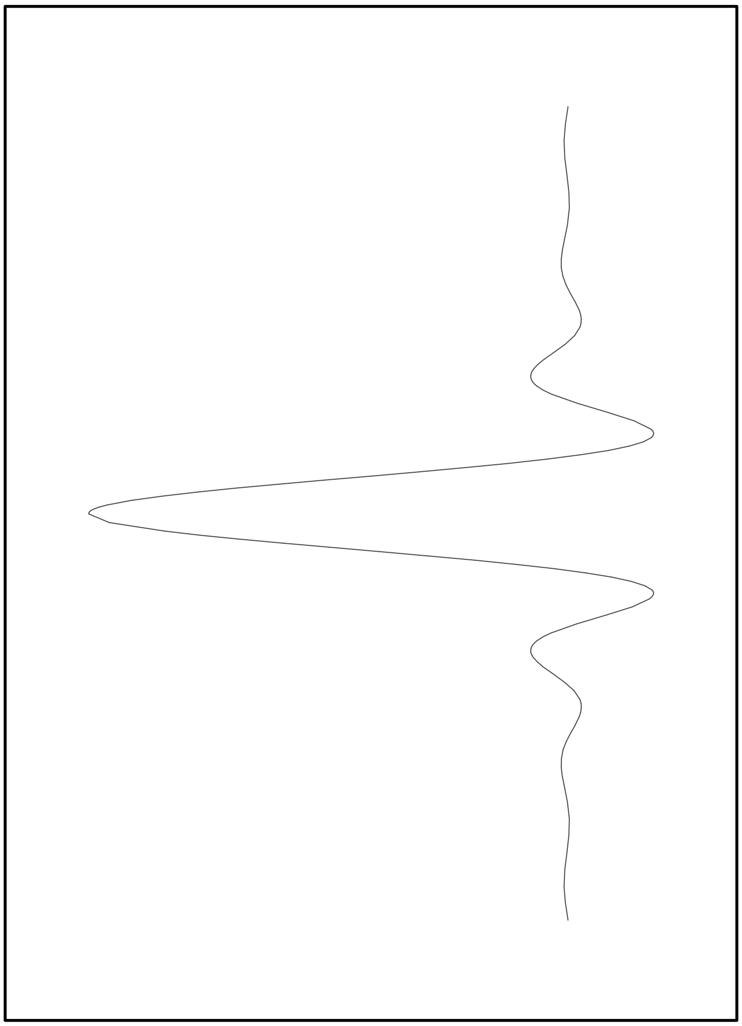
The most general estimate now follows by interpolation: We know that for any function $f \in C^{1,\sigma}$ and any $\mu \in (\sigma, 1+\sigma)$,

$$[f]_{C^\mu} \leq C [f]_{C^{0,\sigma}}^{1+\sigma-\mu} [f]_{C^{1,\sigma}}^{\mu-\sigma}$$

so we can estimate for $r + \mu \geq s + \sigma$

$$\begin{aligned} |\tilde{T}_N u|_{C^{r+\mu}} &\leq C |\tilde{T}_N u|_{C^{r,\sigma}}^{1+\sigma-\mu} |\tilde{T}_N u|_{C^{r+1,\sigma}}^{\mu-\sigma} \\ &\leq C (N^{r-s})^{1+\sigma-\mu} (N^{r+1-s})^{\mu-\sigma} |u|_{C^{s,\sigma}} \\ &\leq C N^{r+\mu-s-\sigma} |u|_{C^{s,\sigma}} \end{aligned}$$

This extends the regularity estimates to arbitrary real exponents.



A plot of K in the one-dimensional case.

4.6. **Approximation estimates.** The approximation property is more difficult to prove. First note that since K has integral equal to 1, the limit of $\tilde{T}_N u$ as $N \to \infty$ is u, so it is enough to control how $\tilde{T}_N u$ changes as N varies.

We can write

$$\frac{d}{dN}\tilde{T}_N u(x) = \int_{\mathbb{R}^d} \frac{d}{dN} K_N(y-x)\, u(y) \mathrm{d}y^n = \int_{\mathbb{R}^d} L_N(y-x) u(y)\, \mathrm{d}y^n,$$

where $L_N(x) = N^{d-1}L(Nx)$ and $L(x) = dK(x) + x^i D_i K(x)$. Note that the Fourier transform $\hat{L}$ of L is equal to $-\xi^i D_i \hat{K}$, which is radially symmetric, and is non-zero only on the annular region $A = B_1(0)\backslash B_{1/2}(0)$. The set A can be covered by the n open sets $A_j = \{\frac{1}{2}d^{-1/2} < |\xi_j| < 1\}$. Take a smooth partition of unity with respect to this cover, say $\{\rho_j\}_{j=1}^d$, with supp$\rho_j \subseteq A_j$ for each j, and each ρ_j even. Then we can write $L = \sum_j L_j$, where each of the functions L_j is a smooth, rapidly decreasing function with Fourier transform $\hat{L}_j = \rho_j \hat{L}$. The beauty of this construction is the following: For each j and each positive integer r define $H_{j,r}$ to be the real, smooth, rapidly decreasing function with Fourier transform $\hat{H}_{j,r} = (i\xi_j)^{-r}\,\hat{L}_j$. This works because the support of $\hat{L}_j$ is away from the ξ_j axis. But then applying the Fourier transform to this definition, we have

$$\frac{\partial^r H_{j,r}}{(\partial x^j)^r} = L_j,$$

and therefore for any positive integer r,

$$\begin{aligned}\frac{d}{dN}\tilde{T}_N u(x) &= N^{-1}\int_{\mathbb{R}^d} L(y)u(x+y/N)\mathrm{d}y^d\\ &= N^{-1}\sum_{j=1}^d \int_{\mathbb{R}^d} L_j(y)u(x+y/N)\mathrm{d}y^d\\ &= N^{-1}\sum_{j=1}^d \int_{\mathbb{R}^d} \frac{\partial^r H_{j,r}}{(\partial y^j)^r}u(x+y/N)\mathrm{d}y^d\\ &= (-1)^r N^{-r-1}\sum_{j=1}^d \int_{\mathbb{R}^d} H_{j,r}(y)D_j^r u(x+y/N)\mathrm{d}y^d.\end{aligned}$$

Taking the C^s norm for any $s \geq 0$, we obtain

$$\left|\frac{d}{dN}\tilde{T}_N u\right|_{C^s} \leq CN^{-r-1}\,|u|_{C^{r+s}}\,.$$

Integrating from N to ∞ gives

$$\left|\tilde{T}_N u - u\right|_{C^s} \leq C\int_N^\infty N^{-r-1}\,dN\,|u|_{C^{r+s}} \leq CN^{-r}\,|u|_{C^{r+s}}\,,$$

which is the required approximation estimate in the case of integer exponents. In the general case we proceed as before, first by noting that the argument above gives

$$\left|\frac{d}{dN}\tilde{T}_N u\right|_{C^{s,\mu}} \leq CN^{s-r-1}\,|u|_{C^{r,\mu}}$$

and then interpolating the $C^{s,\sigma}$ norm of $\frac{d}{dN}\tilde{T}_N u$ between the $C^{s,\mu}$ and $C^{s\pm 1,\mu}$ norms (depending whether σ is greater than or less than μ) to obtain

$$\left|\frac{d}{dN}\tilde{T}_N u\right|_{C^{s,\sigma}} \leq CN^{s+\sigma-r-\mu-1}\left|u\right|_{C^{r,\mu}}.$$

Integrating from N to ∞ then gives the desired approximation estimate

$$\left|\tilde{T}_N u - u\right|_{C^{s,\sigma}} \leq CN^{s+\sigma-r-\mu}\left|u\right|_{C^{r,\mu}},$$

provided $s+\sigma \leq r+\mu$.

4.7. **Approximating tensors.** So far everything has been done for approximating functions. We also need to be able to approximate metrics, and this is done as follows: Given our embedding $\bar{F}$, the tangent space for M at each point can be identified with a subspace of the embedding space $\mathbb{R}^d$. The metric can be extended to a bilinear form on $\mathbb{R}^d$ at each point by taking the action on any normal vector to be zero.

The metric is then represented by a $d \times d$ matrix at each point, and we think of this as a collection of $\frac{d(d+1)}{2}$ real functions on M. We approximate each of these as before, then restrict back to the tangent plane to obtain our approximations. This gives the same kinds of estimates as for the approximations of functions.

Similar remarks apply for approximating arbitrary tensors on a manifold.

5. Perturbation result after Hörmander

I will now turn to an argument based on one of Hörmander [22], which is somewhat easier to motivate and understand than Nash's original argument, and avoids some technicalities such as the short-time existence of Nash's continuous deformation process. This is simplest in the case where the desired metric change is in a non-integer Hölder space. The integer case is not treated by Hörmander, but I include a proof here using a slight extension of his argument. I also provide a proof that more regular metrics are in fact attained by more regular embeddings.

5.1. **Decomposition into frequency bands.** The key idea in Hörmander's proof is to break up the total desired metric change into pieces corresponding to various 'frequency bands', and then feed these pieces in one at a time with a level of smoothing suited to each piece.

In order to get a good decomposition into pieces, we define for each positive integer j an operator R_j as follows:

$$R_j f = T_{e^{j+1}} f - T_{e^j} f. \tag{5.1}$$

For $j = 0$ we take $R_0 f = T_1 f$.

The operators R_j have good estimates: First, for $j > 0$ we have

$$\begin{aligned} \|R_j f\|_{C^r} &\le \int_{e^j}^{e^{j+1}} \left\| \frac{dT}{dN} f \right\|_{C^s} dN \\ &\le C_{r,s} \int_{e^j}^{e^{j+1}} N^{r-s-1}\, dN \|f\|_{C^s} \\ &\le C_{r,s} \frac{N^{r-s}}{r-s}\bigg|_{e^j}^{e^{j+1}} \|f\|_{C^s} \\ &\le \frac{C_{r,s}(e^{r-s}-1)}{r-s} e^{j(r-s)} \|f\|_{C^s} \\ &= \tilde{C}_{r,s} e^{j(r-s)} \|f\|_{C^s}. \end{aligned} \tag{5.2}$$

This estimates holds (with constants depending on r and s) for any values of r and s.

We can write formally

$$f = \sum_{j=0}^{\infty} R_j f,$$

since the partial sum to k terms is just T_{e^k}. This converges to f as $k \to \infty$, at least in the C^β sense for $\beta < \alpha$ if f is C^α. In fact if f is C^k then the sum converges in C^k, but the same result is not true for $C^{k,\sigma}$, $\sigma > 0$ (Exercise: A function f on a compact manifold M is continuous if and only if $T_N f$ approaches f uniformly as $N \to \infty$).

Recalling our definition of the smoothing operators, we can give the operator R_j an interpretation: T_N truncates the Fourier transform of (the extension of) f to the ball of radius N (give or take a bit of smoothing), so R_j is more or less the operator which takes that part of f which has Fourier transform in the shell between radii e^j and e^{j+1}.

5.2. **A characterisation of $C^{k,\alpha}$ functions.** There is a kind of converse to the observation of the previous section, which will be very helpful in the proof we outline below: Suppose we have a sequence of functions u_j, $j = 0, 1, \ldots$ satisfying the following estimates for some constant M:

$$\|u_j\|_{C^r} \le M e^{j(r-s)} \tag{5.3}$$

for every r in some range $[r_1, r_2]$, where $r_1 < s < r_2$. Then the sum $\sum_{j=0}^{\infty} u_j$ converges in C^r for $r < s$ (the sum is absolutely convergent for such r). Let the limit be u. Below we assume that $s = k + \sigma$ where $0 < \sigma < 1$ — we will consider integer cases later.

Theorem 5.1. *Assume s is not an integer. If $u = \sum_j u_j$, and $\|u_j\|_r \leq Me^{j(r-s)}$ for all j and all $r \in [r_1, r_2]$, then u is in C^s, and $\|u\|_{C^s} \leq CM$ for some constant C depending on r_1 and r_2 and the smoothing constants.*

In fact we will get a little more: If we consider the infinum of M over all such decompositions of u into pieces satisfying the estimate (5.3), this is comparable to $\|u\|_{C^s}$.

Note first that the sum converges absolutely in C^k, so in particular the limit is C^k and the C^k norm can be estimated by

$$\|u\|_{C^k} \leq \sum_{j=0}^{\infty} Me^{-j\sigma} \leq \frac{M}{1 - e^{-\sigma}}.$$

We need to obtain a $C^{0,\sigma}$ estimate for any kth derivative of u. To get this, we write $S_j u = \sum_{i=0}^{j} u_i$, choose $\mu = \min\{r_2 - k, 1\} > \sigma$ and leave j to be chosen, and obtain

$$\begin{aligned}
|D^k u(x) - D^k u(y)| &\leq |D^k u(x) - D^k S_j u(x)| \\
&\quad + |D^k S_j u(x) - D^k S_j u(y)| + |D^k S_j u(y) - D^k u(y)| \\
&\leq 2\left\|\sum_{i=j}^{\infty} u_i\right\|_{C^k} + \left\|\sum_{i=0}^{j-1} u_i\right\|_{C^{k,\mu}} |y - x|^\mu \\
&\leq 2M \sum_{i=j}^{\infty} e^{-i\sigma} + M \sum_{i=0}^{j-1} e^{i(\mu-\sigma)} |y - x|^\mu \\
&\leq 2M \frac{e^{-j\sigma}}{1 - e^{-\sigma}} + M \frac{e^{j(\mu-\sigma)}}{e^{\mu-\sigma} - 1} |y - x|^\mu \\
&\leq M|y - x|^\sigma \left(\frac{2\left(e^j|y - x|\right)^{-\sigma}}{1 - e^{-\sigma}} + \frac{\left(e^j|y - x|\right)^{\mu-\sigma}}{e^{\mu-\sigma} - 1} \right).
\end{aligned}$$

Now we choose j to be the integer closest to $-\log|y - x|$, so that

$$\frac{1}{\sqrt{e}} \leq e^j|y - x| \leq \sqrt{e},$$

so that the bracket in the last line becomes a constant depending only on σ and $\mu - \sigma$. This proves the Theorem.

5.3. **The approximation process.** Hörmander's method is to decompose the desired metric change $h = g - g_{F_0}$ into frequency bands as above, then feed these in one at a time to Newton's method, each time smoothing at a length-scale suited to the frequency band. Precisely, we start at the embedding F_0, and take a sequence of adjustments $\dot{F}_j$, $j = 0, 1, \ldots,$. The embedding $F_0 + \sum_{j=0}^{k-1} \dot{F}_j$ is denoted by F_k, and for convenience we denote by u_k the total correction $F_k - F_0 = \sum_{j=0}^{k-1} \dot{F}_j$. Then the corrections are defined by

$$\dot{F}_k = L_{F_0+v_k} h_k, \tag{5.4}$$

where $h_k = R_k h$ is the kth frequency band of the desired metric change, and v_k is a smoothing of u_k:

$$v_k = T_{e^k} u_k. \tag{5.5}$$

Here also L is the operator we derived in section 4 of Lecture 2 as an inverse for the linearised problem:

$$L_F h = \frac{1}{2} \left(G_F^{-1}\right)^{ij,kl} h_{ij} I\!I^F_{kl}, \tag{5.6}$$

where $I\!I^F$ is the second fundamental form of the embedding F, and G_F is the metric induced on the bundle of symmetric 2-tensors by this, i.e.

$$(G_F)_{ij,kl} = I\!I^F_{ij} \cdot I\!I^F_{kl}.$$

After having completed this for all k, we will have achieved a new embedding which is much closer to having the desired metric change — it will turn out that if the desired metric change is C^s with sufficiently small norm δ, then the error after this sequence of corrections is bounded in C^s, with norm at most $C\delta^2$.

5.4. **Estimating compositions and products.** In analyzing the behaviour of the iteration we have just defined, it is useful to observe two facts. First, we have a result that simplifies the estimation of products of functions:

Lemma 5.2. *Suppose ϕ and ψ are C^r functions. Then*

$$\|\phi\psi\|_r \leq C \left(\|\phi\|_0 \|\psi\|_r + \|\phi\|_r \|\psi\|_0\right)$$

where C may depend on r.

The starting point in the proof is the following interpolation estimate (see [21])

$$\|u\|_s \leq C \|u\|_0^{1-s/r} \|u\|_r^{s/r} \tag{5.7}$$

for any $0 < s < r$.

If we compute a derivative of $\phi\psi$ in the direction of a multi-index α, we get terms like this:

$$D^\alpha(\phi\psi) = \sum_{\beta+\gamma=\alpha} D^\beta\phi D^\gamma\psi.$$

If $r = k + \sigma$, with k an integer and $\sigma \in [0, 1)$, then

$$\begin{aligned}
\|\phi\psi\|_r &= \sum_{|\alpha|\le k} \|D^\alpha(\phi\psi)\|_\sigma \\
&\le \sum_{p+q\le k} \left(\|\phi\|_{p+\sigma}\|\psi\|_q + \|\phi\|_p\|\psi\|_{q+\sigma}\right) \\
&\le C \sum_{p+q\le k} \|\phi\|_0^{\frac{q}{p+q+\sigma}} \|\phi\|_{p+q+\sigma}^{\frac{p+\sigma}{p+q+\sigma}} \|\psi\|_0^{\frac{p+\sigma}{p+q+\sigma}} \|\psi\|_{p+q+\sigma}^{\frac{q}{p+q+\sigma}} \\
&\quad + C \sum_{p+q\le k} \|\phi\|_0^{\frac{q+\sigma}{p+q+\sigma}} \|\phi\|_{p+q+\sigma}^{\frac{p}{p+q+\sigma}} \|\psi\|_0^{\frac{p}{p+q+\sigma}} \|\psi\|_{p+q+\sigma}^{\frac{q+\sigma}{p+q+\sigma}} \\
&\le C \sum_{j\le k} \left(\|\phi\|_0\|\psi\|_{j+\sigma} + \|\phi\|_{j+\sigma}\|\psi\|_0\right) \\
&\le C \left(\|\phi\|_0\|\psi\|_r + \|\phi\|_r\|\psi\|_0\right).
\end{aligned}$$

Here we used the interpolation estimate (5.7) to obtain the third line, then Young's inequality to get the second-last line.

The other fact we need is the following, which simplifies the estimation of compositions of the type appearing in our iteration:

Lemma 5.3. *Suppose ψ is a smooth map on an open bounded set U, and f maps into U. Then for any $r \ge 0$,*

$$\|\psi \circ f\|_r \le C\left(1 + \|f\|_r\right).$$

Here the constant C depends on bounds for derivatives of ψ up to order r, and on the bound for U.

This holds quite generally, but we will be applying it in estimating terms in the operator L, which we think of as a smooth function of $f = D^2F$ on a suitable region U where G_F is bounded from below. This gives the following:

Corollary 5.4. *Given a free embedding F_0, there exists $\delta > 0$ and $C < \infty$ such that for any $F \in C^{r+2}$ with $\|F - F_0\|_2 < \delta$,*

$$\|L_F h\|_r \le C\left(\|h\|_r + \|h\|_0\|F\|_{r+2}\right)$$

This follows from the form (5.6) of the operator L, and uses both Lemma 5.2 and 5.3. The crucial point in the above result is that the derivatives of F appear only in a linear way in the estimate, even though high derivatives of $L_F h$ will typically result in products of many terms involving derivatives of F.

Now I will prove Lemma 5.3. If we compute a kth derivative of a composition, we get something of the following form:

$$D^k(\psi \circ f) = \sum_{i=1}^{k} D^i\psi * \sum_{j_1+\cdots+j_i=k} D^{j_1} f * \cdots * D^{j_i} f,$$

where $A * B$ represents a linear combination of terms obtained by contracting tensor A with tensor B. The interpolation estimate can be applied to each term involving derivatives of f, yielding for $\sigma \in [0,1)$ (with $k+\sigma \le r$)

$$\begin{aligned}
&\|D^{j_1} f * \cdots * D^{j_i} f\|_\sigma \\
&\qquad \le C \sum_{l=1}^{i} D^{j_1} f\|_0 \ldots D^{j_{l-1}} f\|_0 \|D^{j_l} f\|_\sigma \|D^{j_{l+1}} f\|_0 \ldots \|D^{j_i} f\|_0 \\
&\qquad \le C \sum_{l=1}^{i} \|f\|_0^{1-j_l+\sigma/r} \|f\|_r^{\frac{j_l+\sigma}{r}} \prod_{m\neq l} \|f\|_0^{1-j_m/r} \|f\|_k^{j_m/r} \\
&\qquad = C\|f\|_0^{i-\frac{j_1+\cdots+j_i+\sigma}{r}} \|f\|_r^{\frac{j_1+\cdots+j_i+\sigma}{r}} \\
&\qquad \le C\|f\|_0^{i-1}\|f\|_r.
\end{aligned}$$

The Lemma follows (using Lemma 5.2), since $D^i\psi$ is bounded in C^σ and $\|f\|_0$ is bounded by assumption.

5.5. **Controlling the embeddings.** We will first show that the embeddings can be controlled quite strongly throughout the sequence of corrections (5.4)–(5.5), and converge to a C^s embedding. In fact the following estimates hold for each j:

$$\|\dot{F}_j\|_{C^r} \le C_1 e^{j(r-s)} \|h\|_{C^s} \tag{5.8}$$

for $r_1 \le r \le r_2$;

$$\|u_j\|_{C^s} \le C_2 \|h\|_{C^s}, \tag{5.9}$$

where C_2 can be assumed to be small enough that any map with $\|F - F_0\|_s \le C_2$ is a free embedding with G_F bounded away from zero; and furthermore

$$\|v_j\|_{C^r} \le C_3 e^{j(r-s)} \|h\|_{C^s} \tag{5.10}$$

for $s < r \le r_2 + 2$; finally

$$\|u_j - v_j\|_{C^r} \le C_4 e^{j(r-s)} \|h\|_{C^s} \tag{5.11}$$

for all $r \leq r_2$. Here $C_1, \ldots, C_4$ are constants independent of j and h. In proving these we will assume that $\|h\|_s$ is sufficiently small, say less than a constant $\delta < 1$.

For $j = 0$ we have the last three inequalities trivially, since $u_0 = v_0 = 0$. We will proceed by induction: Suppose we have the last three inequalities for $0 \leq j \leq k$ and the first one for $0 \leq j \leq k-1$. We will prove the first inequality for $j = k$ and the last three for $j = k+1$.

To prove the first we use Corollary 5.4)(writing $A = \|F_0\|_{r_2+2}$ and using Corollary 5.4)

$$\begin{aligned} \|\dot{F}_k\|_r &= \|L_{F_0+v_k} h_k\|_r \\ &\leq C\left(\|F_0 + v_k\|_{r+2}\|h_k\|_0 + \|F_0 + v\|_2\|h_k\|_r\right) \\ &\leq C\left((A + C_3 e^{k(r+2-s)_+}\|h\|_s)e^{-ks}\|h\|_s + e^{k(r-s)}\|h\|_s\right) \\ &\leq C e^{k(r-s)}\|h\|_s\left(1 + Ae^{-kr} + C_3\delta e^{k((r+2-s)_+-r)}\right) \end{aligned} \tag{5.12}$$

If $s \geq 2$, the exponentials in the bracket are all bounded by 1. Choose $C_1 > C(1+A)$, and choose δ sufficiently small to ensure $C_1 > C(1 + A + C_3\delta)$ — this can be done whatever the value of C_3 may be. This proves the first estimate for $j = k$.

To prove the second for $j = k+1$, we note that $u_{k+1} = \sum_{j=0}^{k} \dot{F}_j$. The estimate we have just proved shows that this sum satisfies the assumptions of Theorem 5.1, so u_{k+1} has C^s norm bounded by CC_1, so we must choose C_2 larger than this.

The third estimate follows from the estimates for the smoothing operator, giving

$$\|v_{k+1}\|_{C^r} \leq C_{r,s} e^{(k+1)(r-s)}\|u_{k+1}\|_{C^s}$$

for any $r \geq s$, and we can take the constant $C_{r,s}$ uniform on bounded intervals of r, in particular for $s \leq r \leq r_2 + 2$. This gives the third estimate provided $C_3 \geq CC_2$.

For $r = 0$ (or more generally any $r < s$) the fourth estimate also follows from the second by the approximation estimates for the smoothing operator:

$$\|u_{k+1} - v_{k+1}\|_0 \leq Ce^{-(k+1)s}\|u_{k+1}\|_s.$$

For $r = r_2$ we get a similar estimate by much cruder means:

$$\begin{aligned}\|u_{k+1} - v_{k+1}\|_{r_2} &\le \|u_{k+1}\|_{r_2} + \|v_{k+1}\|_{r_2}\\ &\le (1+C)\|u_{k+1}\|_{r_2}\\ &\le (1+C)\left\|\sum_{j=0}^{k} \dot F_j\right\|_{r_2}\\ &\le (1+C)C_1\sum_{j=0}^{k} e^{j(r_2-s)}\|h\|_s\\ &\le \frac{(1+C)C_1}{e^{r_2-s}-1}e^{(k+1)(r_2-s)}\|h\|_s.\end{aligned}$$

Interpolation gives the estimate for each $r \in [r_1, r_2]$, with a constant comparable to the larger of the above two. Thus we are in business provided $C_4 \ge C(C_1 + C_2)$. This completes the induction, and establishes the bounds for every j.

It follows (from Theorem 5.1) that as $k \to \infty$ the embeddings $F_k = F_0 + u_k$ converge (in C^r for $r < s$) to a limit F_∞ which is C^s.

5.6. **Controlling the errors.** Now we will turn to controlling the errors in the metric accumulated over the sequence of corrections. This is not too hard: Let us compute the change in the metric $\dot g_k = g_{F_{k+1}} - g_{F_k}$ in each step:

$$\dot g = h_k + D(\dot F_k)\otimes D(\dot F_k) + D(u_k - v_k)\otimes D(\dot F_k) = g_k + E_k + E_k'. \tag{5.13}$$

The second and third terms are the error terms that we need to control.

For the second term we have the estimate

$$\|E_k\|_{C^r} \le C\|\dot F_k\|_{C^1}\|\dot F_k\|_{C^{r+1}} \le CC_1^2 e^{k(r-(2s-2))}\|h\|_s^2.$$

for $r \in [r_1, r_2]$. For the third term we have

$$\begin{aligned}\|E_k'\|_{C^r} &\le C\left(\|u_k - v_k\|_1\|\dot F_k\|_{r+1} + \|u_k - v_k\|_{r+1}\|\dot F_k\|_1\right)\\ &\le CC_1C_4e^{k(r-(2s-2))}\|h\|_s^2\end{aligned}$$

provided $r_1 \le 1$ and $r_2 > s+1$.

Combining these, we have

$$\|\dot g_k - h_k\|_r \le C(C_1^2 + C_1C_4)e^{k(r-(2s-2))}\|h\|_s^2$$

for all k, from which we deduce (by Theorem 5.1) that

$$\|\sum_{k=0}^{\infty}(\dot g_k - h_k)\|_{2s-2} \le C\|h\|_s^2.$$

Thus the metric of the limit F_∞ is g_0+h+E, where $\|E\|_{C^{2s-2}} \le C\|h\|_s^2$.

5.7. **Continuity.** A slightly more detailed look at the above argument also gives us that the embedding F_∞ we end up with, and the metric $g_0 + h + E(h)$, depend continuously on h in C^s. If we take two C^s bilinear forms h and k (with norm less than δ), then the corresponding embeddings at each step, $F_{k,i} = F_0 + u_{k,i}$ and $F_{h,i} = F_0 + u_{h,i}$, satisfy the estimates

$$\|\dot{F}_{h,i} - \dot{F}_{k,i}\|_r \le Ce^{i(r-s)}\|h-k\|_s, \qquad r_1 \le r \le r_2; \tag{5.14}$$

$$\|u_{k,i} - u_{h,i}\|_r \le Ce^{i(r-s)}\|h-k\|_s, \qquad s \le r \le r_2; \tag{5.15}$$

$$\|v_{k,i} - v_{h,i}\|_r \le Ce^{i(r-s)}\|h-k\|_s, \qquad s \le r \le r_2+2; \tag{5.16}$$

(5.17)

$$\|u_{k,i} - v_{k,i} - u_{h,i} + v_{h,i}\|_r \le Ce^{i(r-s)}\|h-k\|_s, \qquad r \le r_2+1.$$

This follows by a straightforward induction argument using the estimates (5.8)–(5.11).

From this we find (by an argument very similar to that above) that the errors $E(h)$ and $E(k)$ in the metrics of the two limiting embeddings satisfy

$$\|E(h) - E(k)\|_{2s-2} \le C\|h-k\|_s\left(\|h\|_s + \|k\|_s\right). \tag{5.18}$$

In particular, E is a continuous map into C^{2s-2}. Similar arguments show that E is differentiable.

5.8. **Removing the errors.** The metric we end up with by feeding in a desired metric change h is given by $h+E(h)$, where $E(h)$ is bounded in C^{2s-2}, hence compact in C^s (since $s > 2$), with norm bounded by $\|h\|_s^2$. It follows from the Schauder fixed point theorem that this takes on all values in a neighbourhood of the origin in C^s: To find a solution of $h+E(h) = \varphi$, we solve the equation $-E(\varphi+v) = v$, so that $\varphi + v + E(\varphi+v) = \varphi$. The map $E(\varphi + .)$ is compact and continuous from $B_\delta \subset C^s$ into C^s, and maps the ball of radius δ' inside the ball of radius $C(\delta')^2$ in C^s for $\delta' < \delta$ if $\|\varphi\|_s < \delta'$. Choosing δ' sufficiently small so that $C\delta^2 < \delta$, we get a fixed point of $E(\varphi + .)$ in $B_{\delta'}$ (see Corollary 11.2 in Gilbarg-Trudinger [7]).

Remark 5.5. In this proof the original embedding F_0 must be bounded in C^{s_2+2}, which means we can perturb in C^s provided $s > 2$ and the initial embedding is $C^{s'}$ with $s' > s+3$.

5.9. **Remarks on integer cases.** Next let us consider what happens in cases where s is an integer. Here we have two different interpretations of the space C^s — either C^s or $C^{s-1,1}$. We will deal with both of these cases.

The main difficulty is that Theorem 5.1 does not apply in either of these cases, so we have to work harder to show that the embeddings are controlled in C^s or in $C^{s-1,1}$ if h is.

The first step is to observe that we can still salvage something from the previous argument, even without Theorem 5.1: For fixed $r_1 < s < r_2$, we define a Banach space $\tilde{C}^s$ to be the space of all functions f which can be expressed in the form $f = \sum_{j=0}^{\infty} f_j$ where

$$\|f_j\|_r \le Me^{j(r-s)} \tag{5.19}$$

for all $r \in [r_1, r_2]$. For a norm $\|.\|_{\tilde{s}}$ we take the infimum of M over all such decompositions of f. The properties of the operators R_j imply that $C^s \subset C^{s-1,1} \subseteq \tilde{C}^s$, and $\|f\|_{\tilde{s}} \le C\|f\|_{s-1,1}$ for $f \in C^{s-1,1}$, since we can take the decomposition $f = \sum_j R_j f$.

We also note that the space $\tilde{C}^s$ is independent of the choice of r_1 and r_2, and can be characterised as the space of functions f for which

$$\|R_j f\|_r \le C(r)e^{j(r-s)}$$

for every j and every $r \ge 0$. If we take a different choice of r_1 and r_2 then the norm may change, but remains equivalent to the previous one. To see this, suppose we have any decomposition $f = \sum_j f_j$ satisfying (5.19), and consider the operators R_i applied to each piece:

$$\|R_i f_j\|_{C^r} \le \tilde{C}_{r,r'} e^{i(r-r')} \|f_j\|_{C^{r'}} \le \tilde{C}_{r,r'} M e^{i(r-r')} e^{j(r'-s)},$$

provided $r_1 \le r' \le r_2$.

We will sum the above estimates to get an estimate for $R_i f$: Note that $\sum_j f_j$ converges to f in C^r for any r: For $r < s$ the sum is absolutely convergent, so $R_i(\sum f_j)$ converges to $R_i f$ in C^r. But for each i, $\|R_i(\sum f_j)\|_{C^p} \le \tilde{C}_{p,r_1} e^{i(p-r_1)} \|\sum f_j\|_{C^{r_1}}$ where p is some large number. Thus the partial sums are bounded in C^p and convergent in C^{r_1}. By interpolation, the sum is convergent in C^r for any $r < p$.

Therefore we have for any r

$$\begin{aligned}\|R_i f\|_{C^r} &= \|R_i(\textstyle\sum f_j)\|_{C^r} \\ &\le \sum_j \|R_i f_j\|_{C^r} \\ &\le CM \sum_j e^{i(r-r'(j))+j(r'(j)-s)} \\ &= CMe^{i(r-s)} \sum_j e^{(i-j)(s-r'(j))}.\end{aligned}$$

where we can choose $r'(j)$ in $[r_1, r_2]$ for each j. We now choose $r'(j) = r_1$ for $j > i$, and $r'(j) = r_2$ for $j \le i$. This gives

$$\begin{aligned}\|R_i f\|_{C^r} &\le CMe^{i(r-s)} \left(\sum_{j<i} e^{-(i-j)(r_2-s)} + \sum_{j\ge i} e^{-(j-i)(s-r_1)} \right) \\ &\le CMe^{i(r-s)} \left(\sum_{l=1}^{\infty} e^{-l(r_2-s)} + \sum_{l=0}^{\infty} e^{-l(s-r_1)} \right) \\ &\le \tilde{C}Me^{i(r-s)}\end{aligned}$$

where

$$\tilde{C} = C \left(\frac{1}{e^{r_2-s}-1} + \frac{1}{1-e^{-(s-r_1)}} \right).$$

This shows that we can replace the original decomposition of f by the decomposition $f = \sum R_j f$, without substantially changing the constant M on the range $[r_1, r_2]$, and that the estimates extend (with increased constants) to any larger range of r.

Remark 5.6. The same construction could be used for non-integer s, but then the space $\tilde{C}^s$ is the same as C^s, with equivalent norm.

Having defined the space $\tilde{C}^s$, we observe that in applying the smoothing operators we can often replace C^s norms with $\tilde{C}^s$ norms in the estimates: In particular, if $f \in \tilde{C}^s$ and $r < s$ (we will assume r is not too close to s, so we might as well take $r < r_1$) , then we can estimate the approximation of $T_{e^i}f$ to f in C^r as follows:

$$\begin{aligned}
\|T_{e^i}f - f\|_r &= \|\sum_j (T_{e^i}f_j - f_j)\|_r \\
&\le C\sum_j e^{i(r-r'(j))}\|f_j\|_{r'(j)} \\
&\le CM\sum_j e^{i(r-r'(j)}e^{j(r'(j)-s)} \\
&\le CMe^{i(r-s)}\left(\sum_{j<i} e^{(i-j)(s-r_2)} + \sum_{j\ge i} e^{(i-j)(s-r_1)}\right) \\
&\le CMe^{i(r-s)} \qquad (5.20)
\end{aligned}$$

where the constant depends on $r_2 - s$ and $s - r_1$.

Now we are ready to begin the proof that the embeddings converge in C^s or $C^{s-1,1}$. As a first step, we show that if h is in small in $\tilde{C}^s$ then the embedding converges in $\tilde{C}^s$, and the error $E(h)$ in the metric satisfies

$$\|E(h)\|_{2s-2} \le C\|h\|_{\tilde{s}}^2.$$

To see this, we just observe that the four estimates (5.8)–(5.11) can be obtained, with the same proof, with $\|h\|_s$ replaced by $\|h\|_{\tilde{s}}$ wherever it appears: The only differences are that we use the definition of $\tilde{C}^s$ instead of Theorem 5.1 to get the bound on u_{k+1}, and we use (5.20) instead of the usual approximation estimate to obtain (5.11). The convergence of the embeddings then follows again from the definition pf $\tilde{C}^s$, and the estimates on the metric errors go through unchanged.

Remark 5.7. Thus we can prove that we can do a $\tilde{C}^s$ perturbation to get any sufficiently small $\tilde{C}^s$ metric change. While interesting, this is not so useful without further understanding of the space $\tilde{C}^s$.

Now we consider the case where h is in $C^{s-1,1}$. Then in particular it is in $\tilde{C}^s$, so we know the embeddings converge in $\tilde{C}^s$, and we have the estimates (5.8)–(5.11). We will prove that u_j remains bounded in $C^{s-1,1}$, from which it follows that the limiting embedding F_∞ has the same bound. This will follow from the fact that the partial sums $\sum_{j=0}^k R_j h = T_{e^{k+1}}h$ remain bounded in $C^{s-1,1}$ (this is just the bound for the smoothing operator).

To use this, we have to express u_k in a way that involves the partial sums. This is analogous to an integration by parts:

$$
\begin{aligned}
u_{k+1} &= \sum_{j=0}^{k} \dot{F}_j \\
&= \sum_{j=0}^{k} L_{F_0+v_j} h_j \\
&= L_{F_0}\left(\sum_{j=0}^{k} h_j\right) + \sum_{l=1}^{k}\left(L_{F_0+v_l} - L_{F_0+v_{l-1}}\right)\left(\sum_{j=l}^{k} h_j\right). \qquad (5.21)
\end{aligned}
$$

We can estimate this in $C^{s-1,1}$ as follows:

$$
\begin{aligned}
\|u_{k+1}\|_s &\le C\left(\left\|\sum_{j=0}^{k} h_j\right\|_0 + \left\|\sum_{j=0}^{k} h_j\right\|_s\right) \\
&\quad + \sum_{l=1}^{k}\left(\|v_l - v_{l-1}\|_{s+2}\left\|\sum_{j=l}^{k} h_j\right\|_0 + \|v_l - v_{l-1}\|_2\left\|\sum_{j=l}^{k} h_j\right\|_s\right) \\
&\le C\|h\|_s + \sum_{l=1}^{k} e^{2l}e^{-sl}\|h\|_s^2 \\
&\quad + \sum_{l=1}^{k} e^{(2-s)l}\|h\|_s^2 \\
&\le C\|h\|_s.
\end{aligned}
$$

This completes the proof for the $C^{s-1,1}$ case.

Next we turn to the proof for the C^s case for s an integer. If h is C^s then it is certainly $C^{s-1,1}$, so we have convergence of the embeddings in $C^{s-1,1}$. To show that the limiting embedding is C^s we will show that the embeddings u_k form a Cauchy sequence in C^s. This will use the fact that the partial sums $T_{e^k}h = \sum_{j=0}^{k-1} h$ converge to h in C^s. From (5.21) we have the following expression for the difference between u_{k+1} and u_{l+1} for $l < k$:

$$
u_{k+1} - u_{l+1} = L_{F_0+v_{l+1}}\left(\sum_{j=l+1}^{k} h_j\right) + \sum_{i=l+2}^{k}\left(L_{F_0+v_i} - L_{F_0+v_{i-1}}\right)\left(\sum_{j=i}^{k} h_j\right) \qquad (5.22)
$$

For convenience we define

$$\Delta_l = \sup_{m\geq n\geq l}\left\|\sum_{i=n}^{m} h_i\right\|_s,$$

and note that $\Delta_l \to 0$ as $l\to\infty$. Estimating this as before, we find

$$\begin{aligned}
&\|u_{k+1}-u_{l+1}\|_s\\
&\quad\leq C\left(\|F_0+v_{l+1}\|_{s+2}\left\|\sum_{j=l+1}^{k} h_j\right\|_0+\|F_0+v_{l+1}\|_2\left\|\sum_{j=l+1}^{k} h_j\right\|_s\right)\\
&\quad\quad+C\sum_{i=l+2}^{k}\left(\|v_i-v_{i-1}\|_{s+2}\left\|\sum_{j=i}^{k} h_j\right\|_0+\|v_i-v_{i-1}\|_2\left\|\sum_{j=i}^{k} h_j\right\|_s\right)\\
&\quad\leq C\left(1+e^{2(l+1)}\|h\|_s\right)\sum_{j=l+1}^{k} e^{-js}\|h\|_s+C\left\|\sum_{j=l+1}^{k} h_j\right\|_s\\
&\quad\quad+C\sum_{i=l+2}^{k}\left(e^{2i}\|h\|_s\sum_{j=i}^{k} e^{-js}\|h\|_s+e^{i(2-s)}\|h\|_s\left\|\sum_{j=i}^{k} h_j\right\|_s\right)\\
&\quad\leq C\left(e^{-(l+1)s}+e^{-(l+1)(s-2)}\|h\|_s\right)\|h\|_s+C\Delta_{l+1}\\
&\quad\quad+C\sum_{i=l+2}^{k}\left(e^{-(s-2)i}\|h\|_s^2+e^{-i(s-2)}\|h\|_s\Delta_{l+1}\right)\\
&\quad\leq Ce^{-(s-2)(l+1)}(1+\Delta_{l+1})\|h\|_s.
\end{aligned}$$

This can be made arbitrarily small by choosing l sufficiently large, so the sequence $\{u_k\}$ is Cauchy in C^s, and the limiting embedding F_∞ is C^s.

The argument using the Leray-Schauder fixed-point theorem to remove the errors goes through unchanged.

Remark 5.8. This works for integers $s>2$. It fails for $s=2$, in several points: First, we no longer have that the error $E(h)$ is compact in C^s — this is not so crucial, since we could work a bit harder and deduce that the error is differentiable in C^2 near 0, with derivative zero, then apply the classical implicit function theorem instead to deduce $h+E(h)$ covers a neighbourhood of the origin in C^2. More crucial is the fact that we can no longer show convergence in C^2, since we rely on the exponential decay at rate $e^{-(s-2)}$ to bound various terms. It is not known whether C^2 metrics can be C^2-isometrically embedded.

5.10. **Higher regularity.** To complete the proof of the perturbation result, we show that the embeddings converge in $C^{s'}$ for non-integer $s' > s$ if h is $C^{s'}$ and F_0 is sufficiently regular. Here we do not want to assume that h is small in $C^{s'}$ or to decrease δ any further. Integer cases can also be treated by methods analogous to those in the previous section.

Given s', we choose some $r_3 > s'+1$, and assume that F_0 is bounded in C^{r_3+2}, with norm A'. The first step is to observe that the estimate (5.10) on v_k obtained in the proof of convergence in C^s extends (possibly with a larger constant C_3' instead of C_3) to hold with r_3 replacing r_2: The bound on $\|v_k\|_r$ follows from the bound on $\|u_k\|_s$ together with the properties of the smoothing operator.

We will prove that the estimates (5.8)–(5.11) holds (for some new constants $\tilde{C}_1, \ldots, \tilde{C}_4$) with r_2 replaced by r_3 and s replaced by s'. This follows by induction as before: In the proof of (5.8), we obtain from (5.12)

$$\begin{aligned}\|\dot{F}_k\|_r &\le C\left(\|F_0 + v_k\|_{r+2}\|h_k\|_0 + \|F_0 + v\|_2\|h_k\|_r\right)\\ &\le C\left((A' + C_3'\delta e^{k(r+2-s)_+})e^{-ks'}\|h\|_{s'} + e^{k(r-s')}\|h\|_{s'}\right)\\ &\le C(1 + A' + C_3'\delta)e^{k(r-s')}\|h\|_{s'},\end{aligned}$$

since $s' > 2$. This proves the estimate if we choose $\tilde{C}_1 = C(1+A'+C_3'\delta)$. Here we do not need to choose δ small as we did before, because the estimate does not involve $\tilde{C}_3$, only C_3. The remaining estimates now follow without change.

It follows from the new version of (5.8) and Theorem 5.1 that the limiting embedding is $C^{s'}$. The error in the metric can also be bounded: In equation (5.13) we can estimate

$$\begin{aligned}\|E_k\|_r &\le C\|\dot{F}_k\|_{C^1}\|\dot{F}_k\|_{C^{r+1}}\\ &\le CC_1 e^{k(1-s)}\tilde{C}_1 e^{k(1+r-s')}\|h\|_s\|h\|_{s'}\\ &\le CC_1\tilde{C}_1\delta e^{k(r-(s'+s-2))}\|h\|_{s'},\end{aligned}$$

and

$$\begin{aligned}\|E_k'\|_r &\le C\left(\|u_k - v_k\|_1\|\dot{F}_k\|_{r+1} + \|u_k - v_k\|_{r+1}\|\dot{F}_k\|_1\right)\\ &\le CC_4 e^{k(1-s)}\|h\|_s\tilde{C}_1 e^{k(r+1-s')}\|h\|_{s'}\\ &\le CC_4\tilde{C}_1\delta e^{k(r-(s'+s-2))}\|h\|_{s'},\end{aligned}$$

provided $r < r_3 - 1$. Theorem 5.1 then implies the estimate

$$\|E\|_{s'+s-2} \le C\delta\|h\|_{s'}.$$

As before, we can also show that the limit metric and the limit embedding vary $C^{s'}$-continuously as a function of h, and that the error E is continuous in $C^{s'+s-2}$.

We now want to apply the Schauder fixed point theorem to show that if $\|\varphi\|_s < \delta'$ (with the same δ' as before) and $\|\varphi\|_{s'} < \infty$, then there is some $h \in C_{s'}$ such that $h + E(h) = \varphi$.

Consider the closed convex set $A = \{h : \ \|h\|_s \leq \delta', \ \|h\|_{s'} \leq M\}$ for some constant M yet to be chosen. The same estimates as before show that if $\|\varphi\|_s \leq \delta'$ and $h \in A$ then $\|E(\varphi + h)\|_s < \delta'$. To estimate the $C^{s'}$ norm we note that by interpolation, if $\|h\|_{s'} \leq M$ then

$$\begin{aligned}\|E(\varphi+h)\|_{s'} &\leq C\|E(\varphi+h)\|_{s'+s-2}^{\frac{s'-s}{s'+s-2}}\|E(\varphi+h)\|_s^{\frac{s-2}{s'+s-2}} \\ &\leq C\delta\left(\|\varphi\|_{s'}+M\right)^{1-\frac{s-2}{s'+s-2}} \\ &< M,\end{aligned}$$

provided M is sufficiently large compared to $\|\varphi\|_{s'}$. Therefore the map $-E(\varphi + .)$ is compact and continuous, and maps A strictly inside itself. Therefore we have a fixed point, which is a $C^{s'}$ symmetric bilinear form v such that $h + E(h) = \varphi$ for $h = \varphi + v$.

This also gives the C^∞ case: If $\|h\|_s < \delta'$ and h is C^∞, then we get for any $s' > s$ a $C^{s'}$ embedding achieving the metric $g_0 + h$, with bounds in C^r depending only on $\|h\|_r$ for each $r \in [s, s']$. Taking a limit as $s' \to \infty$, and using a diagonal subsequence construction, we obtain a limit which is C^∞.

5.11. **Further remarks.** It is useful to note that the result we obtain is somewhat stronger than the one stated by Nash: To obtain a C^s embedding of a C^s metric g it suffices to start from a $C^{s+3+\varepsilon}$ free embedding F_0, with metric g_0 satisfying $\|g_0 - g\|_s < \delta'$, where δ' depends on s, $\|F_0\|_{s+3+\varepsilon}$, and a bound on $G_{F_0}^{-1}$ (the latter controls the freeness of F_0).

The reason why Nash did not bother to state this is probably that the stronger result still doesn't seem to be enough to get around the need for Nash's elaborate construction using the y and z embeddings: If we approximately isometrically embed a C^s metric in the C^s sense, the $C^{s+3+\varepsilon}$ norms of the embedding necessarily become large if the metric is not this regular, and so we have no control over the required δ'. In fact this can be circumvented using better methods for approximations — I'll make some more remarks on this point after discussing Günther's argument, since it is also his work which provide the better approximation results.

I should also remark that the proof I have given here can easily be adapted to other settings where there is loss of differentiability, or to prove a general implicit function theorem of Nash-Moser type. See the survey of Hamilton [16] for many examples and applications of this kind of result. There are many approaches to the proof of this kind of result, ranging from the Newton method employed by Moser [33]–[34] and by Schwartz [45]–[46] — which yields a relatively simple proof but one which is not optimal as regards differentiability assumptions — to methods of Jacobowitz [24]— which use extension of real-analytic functions to complex-analytic ones, applying complex-analytic methods and then employing classical approximation techniques to get results for lower differentiability classes — to the arguments of Sergeraert [47] and Hamilton [16], which are aimed at producing results in the setting of suitable Frechet spaces (see also [31] for further extensions), and the argument of Nash himself [37]–[38] (see also Hörmander [21] for a similar argument with a little further motivation and explanation) which is beautiful and delicate but decidedly non-obvious ("like lightning striking" according to Gromov). I like Hörmander's argument because it is significantly simpler and more transparent than Nash's, but still gives good results for natural graded sequences of Banach spaces (such as C^k as I had here) as well as for the Frechét space limit.

6. Günther's argument

Next we will work through the argument of Günther [13]–[14] which gets around the loss of differentiability, and so allows the isometric embedding theorem to be deduced from a Banach space fixed point theorem.

6.1. **Loss of differentiability.** Recall the problem which forced us to use the Nash-Moser argument: Given a general C^s variation (let us assume for simplicity that s is not an integer), the change in the metric is bounded in C^{s-1}, but not in C^s, so one might expect to apply the inverse function theorem by showing that the derivative of this map is invertible. But the 'inverse' we construct, given by taking normal variations, is only C^{s-1} if the metric change is C^{s-1}, so this is not actually an inverse for the derivative as a map between C^s and C^{s-1} — that is, our inverse for the derivative is unbounded.

The source of this unboundedness is easy to identify: If we take an actual variation in a normal direction, rather than an infinitesimal variation, then we have an extra term which is quadratic in the derivatives of the variation. This drops out for an infinitesimal variation (so the derivative maps C^s to C^s), but ruins the regularity for an actual variation.

One way to think about the problem is like this: If we consider a variation $F_t = F_0 + tV$ where V is normal to F_0, the rate of variation V is no longer normal to F_t for positive times t. One should instead modify the variation to keep it normal to the moving submanifold. That is hard to do — one runs into problems in showing the existence of such a continuous deformation — but we are left with the feeling that there should be non-trivial variations which do not result in loss of regularity.

If the problem is that our variations are not normal to the deformed submanifolds, it should be possible to correct for this by including some suitable component of the variation which is tangential to the submanifold. Günther managed to do this, by showing that the quadratic error term can be counteracted by a suitable tangential variation.

6.2. **Constructing good variations: The torus case.** For simplicity let us first consider the case of the torus, so that we have a flat metric on our manifold. This simplifies things slightly, as we can commute derivatives without generating curvature terms, and we don't have to worry about covariant derivatives.

If we want to achieve a metric variation h_{ij}, then the equation we must satisfy is the following:

$$D_iF \cdot D_jV + D_jF \cdot D_iV + D_iV \cdot D_jV = h_{ij}.$$

As we have already observed, this looks simpler if the variations are normal. This is encapsulated in the following reformulation of the above equation:

$$D_j(D_iF \cdot V) + D_i(D_jF \cdot V) - 2D_iD_jF \cdot V + D_iV \cdot D_jV = h_{ij}. \tag{6.1}$$

We are free to choose the tangential part $V \cdot D_jF$ as well as the component in the direction of the second derivatives $V \cdot D_iD_jF$, since the first and second derivatives of a free embedding are independent. The idea is to try to move the bad term, the one where the quadratic term where the derivatives are lost, into the first brackets to allow cancellation by the tangential part. This seems to make sense: If V is a C^s map, then the quadratic term is C^{s-1}, so one might hope it could be written as the derivative of something in C^s. It seems unlikely this can be done by purely algebraic manipulations, but Günther observed that something nice happens if you apply the Laplacian (with respect to the flat metric) to the quadratic term:

$$\begin{aligned}\Delta(D_iV\cdot D_jV) &= D_i\Delta V\cdot D_jV + D_j\Delta V\cdot D_iV + 2D_kD_iV\cdot D_kD_jV\\ &= D_i\left(\Delta V\cdot D_jV\right) + D_j\left(\Delta V\cdot D_iV\right)\\ &\quad + 2\left(D_kD_jV\cdot D_kD_iV - \Delta V\cdot D_iD_jV\right).\end{aligned} \tag{6.2}$$

The crucial point is that the top order terms in the Laplacian are the ones where both derivatives fall on the same factor, none on the other. But this allows us to write the top order term as a derivative plus a more regular error: Observe that if V is C^s, this is $D_i(C^{s-2}) + D_j(C^{s-2}) + C^{s-2}$. But $\Delta - 1$ commutes with differentiation and has an inverse $T : C^{s-2} \to C^s$, (bounded provided s is not an integer) so we can write

$$\begin{aligned}D_iV\cdot D_jV &= (\Delta-1)^{-1}(\Delta-1)(D_iV\cdot D_jV)\\ &= D_i\left(T(\Delta V\cdot D_jV)\right) + D_j\left(T(\Delta V\cdot D_iV)\right)\\ &\quad + T\left(2D_kD_iV\cdot D_kD_jV - 2\Delta V\cdot D_iD_jV - D_iV\cdot D_jV\right).\end{aligned} \tag{6.3}$$

This is exactly what we need: Plugging this into Equation (6.1) we obtain

$$\begin{aligned}h_{ij} &= D_i\left(D_jF\cdot V + T(\Delta V\cdot D_jV)\right) + D_j\left(D_iF\cdot V + T(\Delta V\cdot D_iV)\right)\\ &-2D_iD_jF\cdot V + T\left(2D_kD_iV\cdot D_kD_jV - 2\Delta V\cdot D_iD_jV - D_iV\cdot D_jV\right).\end{aligned}$$

To solve this we require that

$$V\cdot D_iF = -T(\Delta V\cdot D_iV) \tag{6.4}$$

and

$$V\cdot D_iD_jF = -\frac{h_{ij}}{2} + T\left(D_kD_iV\cdot D_kD_jV - \Delta V\cdot D_iD_jV - \frac{1}{2}D_iV\cdot D_jV\right). \tag{6.5}$$

Since the embedding is free, any system of the form

$$\begin{aligned}V\cdot D_iF &= A_i\\ V\cdot D_iD_jF &= B_{ij}\end{aligned} \tag{6.6}$$

has a unique solution in the span of the first and second derivatives of F, which we can denote by $L(A,B)$. If F is free and C^{s+2}, then L is a bounded linear map from $C^s \times C^s$ to C^s. We have to solve an equation of the form

$$V = L(Q_1(V), -\frac{1}{2}h_{ij} + Q_2(V)),$$

where Q_1 and Q_2 are continuous maps from C^s to C^s satisfying $\|Q_i(V)\|_s \leq C\|V\|_s\|V\|_2$. This can be tackled using a fixed point theorem in the

Banach space C^s, or by observing that

$$V - L(Q_1(V), Q_2(V))$$

is a smooth function of V in C^s, with derivative at $V = 0$ equal to the identity. Thus by the (Banach space) inverse function theorem, it covers a neighbourhood of zero. It follows that we can solve the perturbation problem about C^{s+2} free embeddings for sufficiently small C^s perturbations.

I'll say more about this later. First I will consider the general case, where essentially the same method works, though the non-flatness of the background metric introduces some extra terms.

6.3. **Constructing good variations: The general case.** Now we consider an arbitrary freely embedded submanifold M^n in $\mathbb{R}^N$. We equip M with a metric g. The problem we need to solve is the same as before, in any local coordinates: If h is some (small) C^s section of the bundle of symmetric 2-tensors on M, then we need

$$D_jF \cdot D_jV + D_jF \cdot D_iV + D_iV \cdot D_jV = h_{ij}.$$

This can be written as

$$\nabla_j(D_iF \cdot V) + \nabla_i(D_jF \cdot V) - 2\nabla_i\nabla_jF \cdot V + D_iV \cdot D_iV = h_{ij}.$$

As before the main point is to split up the quadratic term in a good way. The covariant Laplacian $\Delta = g^{kl}\nabla_k\nabla_l$ is again a self-adjoint elliptic operator from C^s to C^{s-2}, and $\Delta - 1$ is invertible. There is some difference arising from the fact that the Laplacian does not quite commute with derivatives: We need to satisfy

$$\begin{aligned}
0 &= (\Delta-1)\left(\nabla_j(D_iF\cdot V)+\nabla_i(D_jF\cdot V)-2\nabla_i\nabla_jF\cdot V+D_iV\cdot D_iV-h_{ij}\right)\\
&= \nabla_j\left((\Delta-1)(D_iF\cdot V)\right)+\nabla_j\left((\Delta-1)(D_iF\cdot V)\right)-2(\Delta-1)(\nabla_i\nabla_jF\cdot V)\\
&\quad + R_j^p\nabla_p(D_iF\cdot V)+R_i^p\nabla_p(D_jF\cdot V)+(\Delta-1)(D_iV\cdot D_jV)
\end{aligned}$$

We expand the last term as follows:

$$\begin{aligned}
(\Delta-1)(D_iV\cdot D_jV) &= \Delta\nabla_iV\cdot D_jV+\Delta\nabla_jV\cdot D_iV\\
&\qquad + 2\nabla_k\nabla_iV\cdot\nabla_k\nabla_jV - D_iV\cdot D_jV\\
&= \nabla_i\Delta V\cdot D_jV+\nabla_j\Delta V\cdot D_iV\\
&\quad + R_i^pD_pV\cdot D_jV+R_j^pD_pV\cdot D_iV\\
&\quad + 2\nabla_k\nabla_iV\cdot\nabla_k\nabla_jV-D_iV\cdot D_jV\\
&= \nabla_i(\Delta V\cdot D_jV)+\nabla_j(\Delta V\cdot D_iV)\\
&\quad + R_i^pD_pV\cdot D_jV+R_j^pD_pV\cdot D_iV\\
&\quad + 2\nabla_k\nabla_iV\cdot\nabla_k\nabla_jV-2\Delta V\cdot\nabla_i\nabla_jV-D_iV\cdot D_jV.
\end{aligned}$$

Substituting in the previous equation, we get

$$\begin{aligned}0 &= \nabla_j\left((\Delta-1)(D_iF\cdot V)+\Delta V\cdot D_iV\right)+\nabla_i\left((\Delta-1)(D_jF\cdot V)+\Delta V\cdot D_jV\right)\\ &\quad -2(\Delta-1)(\nabla_i\nabla_jF\cdot V+\frac{1}{2}h_{ij})+R_j^p\nabla_p(D_iF\cdot V)+R_i^p\nabla_p(D_jF\cdot V)\\ &\quad +R_i^pD_pV\cdot D_jV+R_j^pD_pV\cdot D_iV-D_iV\cdot D_iV\\ &\quad +2\nabla_k\nabla_iV\cdot\nabla_k\nabla_jV-2\Delta V\cdot\nabla_i\nabla_jV\end{aligned}$$

Now we are in business: We require that the tangential part of V be such that the first two terms vanish, so that

$$V\cdot D_iF=-N_i(V)$$

where

$$(\Delta-1)N_i=\Delta V\cdot D_iV,$$

and then we require that the components in the direction of the second derivatives satisfy the remaining identity:

$$V\cdot\nabla_i\nabla_jF=-\frac{1}{2}h_{ij}+M_{ij}(V),$$

where

$$\begin{aligned}(\Delta-1)M_{ij} &= \nabla_k\nabla_iV\cdot\nabla_k\nabla_jV-\Delta V\cdot\nabla_i\nabla_jV-\frac{R_i^p\nabla_pN_j}{2}-\frac{R_j^p\nabla_pN_i}{2}\\ &\quad +\frac{1}{2}R_i^pD_pV\cdot D_jV+\frac{1}{2}R_j^pD_pV\cdot D_iV-\frac{1}{2}D_iV\cdot D_jV.\end{aligned}$$

If V is C^s, then $\Delta V\cdot D_iV$ is C^{s-2}, so N_i is bounded in C^s, with norm $\|N(V)\|_s\leq C\|V\|_s\|V\|_1$ (for $s\geq 2$). Thus M is also C^s, with $\|M\|_s\leq C\|V\|_s\|V\|_2$.

As before, the system can be solved for sufficiently small h in C^s by an implicit function theorem argument, or with slightly stronger results by a fixed point or successive approximations argument. Note that we require the embedding F to be C^{s+2} so that the solution $L(A,B)$ of the system

$$\begin{aligned}V\cdot D_iF&=A_i;\\ V\cdot\nabla_i\nabla_jF&=B_{ij}\end{aligned}$$

(determined uniquely in the span of the first and second derivatives of F) is bounded from $C^s(M,T^*M)\times C^s(M,S^{(2)}_*M)$ to $C^s(M,\mathbb{R}^N)$.

6.4. **The perturbation result.** A careful successive approximations argument yields the following result:

Theorem 6.1. *Let F be a free C^{s+2} embedding, and $h \in C^s(M, S_*^{(2)}M)$ with $s > 2$. There is a positive number θ independent of F, h and s such that if*

$$\|L\|_{B(C^2(M,T^*M)\times C^2(M,S_*^{(2)}M),C^2(M,\mathbb{R}^N))}\|L(0,h)\|_{C^2} \leq \theta,$$

then there exists $V \in C^s(M,\mathbb{R}^N)$ (small in C^2) such that

$$D_i(F+V)\cdot D_j(F+V) = D_iF\cdot D_jF + h_{ij}.$$

This is very nice: The smallness condition on the perturbation is in C^2, with the smallness determined essentially by the freeness of the embedding (roughly the size of the operator G^{-1} defined in the previous sections) together with the fourth derivatives of the embedding. In particular, if $s > 4$, then this means that effectively we can perturb about any free C^s embedding F to get nearby C^s metrics: First take a C^{s+2} (or C^∞!) embedding F' which is close to F (we can keep the C^4 norm comparable while making the C^3 difference as small as desired). Then the freeness of F' is not much worse than that of F, and the C^4 norms of F' are also comparable, so we can perturb about F' to get any metric change which is small in C^2. Any C^s metric which is close to g_F in C^2 is close to $g_{F'}$ in C^2, so can be obtained by perturbing about F', and the resulting embedding will be close to F in C^2. Some care is needed here, because we do not claim that the resulting embedding is close to F in C^s.

Note that an argument like that just outlined also shows that any C^s metric ($s > 4$) which can be realized by a free C^r embedding in $\mathbb{R}^N$ with $r > 4$, can also be realised by a free C^s embedding in $\mathbb{R}^N$: First take a smooth approximation of the initial embedding, then perturb about this to get a C^r embedding with the same metric.

6.5. **More on approximations.** Günther observed that the perturbation result can be applied to improve the results about approximate isometric embeddings, reducing the dimension required for the isometric embedding theorem.

The basic tool is the following variant on the above perturbation result:

Theorem 6.2. *Let $B \subset \mathbb{R}^n$ be the open unit ball and B_1 and B_2 open sets with $\bar{B}_1 \subset B_2$ and $\bar{B}_2 \subset B$. Let $F \in C^{s+2}(\bar{B},\mathbb{R}^N)$ be a free mapping, and $h \in C^s(B,\mathbb{R}^{\frac{n(n+2)}{2}})$ with $s > 2$. There exists $\theta > 0$ (independent of F, s and h) such that if* $\operatorname{supp} h \subset B_1$ *and*

$\|L\|_{B(C^2,C^2)}\|L(0,h)\|_{C^2} \leq \theta$, *then there exists* $V \in C^s(B,\mathbb{R}^N)$ *with* $\mathrm{supp}V \subset B_2$ *and*

$$D_i(F+V)\cdot D_j(F+V) = D_iF\cdot D_jF + h_{ij}.$$

Thus we can do compactly supported variations of the metric with compactly supported variations of the embedding. The proof is very similar to that for the previous case: First choose a smooth cut-off function ρ with support in B_2 and with $\rho = 1$ in B_1. As before, we want to solve the equation

$$D_i(V\cdot D_jF)+D_j(V\cdot D_iF)-2V\cdot D_iD_jF+D_iVD_jV-h_{ij}=0. \tag{6.7}$$

To ensure that V has compact support we will insist that it has the form $V=\rho^2W$ with W bounded. Substituting this into equation (6.7), we obtain

$$\begin{aligned}0 &= \rho^3D_i\left(\frac{W}{\rho}\cdot D_jF\right)+\rho^3D_i\left(\frac{W}{\rho}\cdot D_jF\right)\\ &\quad+3\rho D_i\rho W\cdot D_jF+3\rho D_i\rho W\cdot D_jF\\ &\quad-2\rho^2W\cdot D_iD_jF+\rho^4D_iW\cdot D_jW-h_{ij}\\ &\quad+2\rho^3D_i\rho W\cdot D_jW+2\rho^3D_j\rho W\cdot D_iW+4\rho^2D_i\rho D_j\rho|W|^2.\end{aligned}$$

The strategy will be as before to absorb the highest order part of the quadratic error term into the derivatives where they can be cancelled by the tangential part of V, and then prescribe the component of V in direction D_iD_jF by setting the remaining terms equal to zero. The key term is $\rho^4D_iW\cdot D_jW$, which we rewrite using the following:

$$\begin{aligned}&\Delta\left(\rho D_iW\cdot D_jW\right)\\ &\quad=\rho D_i\Delta W\cdot D_jW+\rho D_j\Delta W\cdot D_iW\\ &\qquad+\Delta\rho D_iW\cdot D_jW+2\rho D_kD_iW\cdot D_kD_jW\\ &\qquad+2D_k\rho D_kD_iW\cdot D_jW+2D_k\rho D_kD_jW\cdot D_iW\\ &\quad=D_i\left(\rho\Delta W\cdot D_jW\right)+D_j\left(\rho\Delta W\cdot D_iW\right)\\ &\qquad-D_i\rho\Delta W\cdot D_jW-D_j\rho\Delta W\cdot D_iW-2\rho\Delta W\cdot D_iD_jW\\ &\qquad+\Delta\rho D_iW\cdot D_jW+2\rho D_kD_iW\cdot D_kD_jW\\ &\qquad+2D_k\rho D_kD_iW\cdot D_jW+2D_k\rho D_kD_jW\cdot D_iW\end{aligned}$$

We also note that $D_j\rho W\cdot D_iW=\frac{1}{2}D_i\left(D_j\rho|W|^2\right)-\frac{1}{2}|W|^2D_iD_j\rho$. Equation (6.7) then becomes

$$\begin{aligned}
0=\rho^3D_i&\left(\frac{W}{\rho}\cdot D_jF+\Delta^{-1}\left(\rho\Delta W\cdot D_jW\right)+D_j\rho|W|^2\right)\\
&+\rho^3D_j\left(\frac{W}{\rho}\cdot D_iF+\Delta^{-1}\left(\rho\Delta W\cdot D_iW\right)+D_i\rho|W|^2\right)\\
&+3\rho D_i\rho W\cdot D_jF+3\rho D_i\rho W\cdot D_jF-2\rho^2W\cdot D_iD_jF-h_{ij}\\
&-\rho^3D_iD_j\rho|W|^2+4\rho^2D_i\rho D_j\rho|W|^2\\
&-\rho^3\Delta^{-1}\left(D_i\rho\Delta W\cdot D_jW+D_j\rho\Delta W\cdot D_iW+2\rho\Delta W\cdot D_iD_jW\right)\\
&+\rho^3\Delta^{-1}\left(\Delta\rho D_iW\cdot D_jW+2\rho D_kD_iW\cdot D_kD_jW\right)\\
&+2\rho^3\Delta^{-1}\left(D_k\rho D_kD_iW\cdot D_jW+D_k\rho D_kD_jW\cdot D_iW\right)
\end{aligned}$$

To simplify this we write

$$A_j(W)=\Delta^{-1}\left(\rho\Delta W\cdot D_jW\right)+|W|^2D_j\rho,$$

and observe that if W is bounded in C^s, then $A_j(W)$ is also in C^s, with norm at most $C\|W\|_s\|W\|_1$. Similarly we write

$$\begin{aligned}
B_{ij}(W)=&\left(2D_i\rho D_j\rho-\frac{1}{2}\rho D_iD_j\rho\right)|W|^2\\
&\quad+\rho\Delta^{-1}\left(\rho D_kD_iW\cdot D_kD_jW-\rho\Delta W\cdot D_iD_jW\right)\\
&\quad+\rho\Delta^{-1}\left(D_k\rho D_kD_iW\cdot D_jW+D_k\rho D_kD_jW\cdot D_iW\right)\\
&+\frac{1}{2}\rho\Delta^{-1}\left(\Delta\rho D_iW\cdot D_jW-D_i\rho\Delta W\cdot D_jW-D_j\rho\Delta W\cdot D_iW\right).
\end{aligned}$$

Again, $\|B_{ij}(W)\|_s\le C\|W\|_s\|W\|_2$. Then we can solve the perturbation system by setting

$$\begin{aligned}
W\cdot D_jF&=-\rho A_j(W)\\
W\cdot D_iD_jF&=-\frac{3}{2}D_i\rho A_j(W)-\frac{3}{2}D_j\rho A_i(W)+B_{ij}(W)-\frac{1}{2}h_{ij}.
\end{aligned}$$

Here I used the fact that $\rho=1$ on the support of h. This is now a very nice system, and we can apply a fixed-point theorem to get a solution if h is sufficiently small.

To apply this result, Günther takes any strictly short free embedding F_0 of the manifold, and takes a decomposition of the difference metric $g-g_{F_0}$ as in (3.1) from Lecture 3. Each of the terms $a_k^2df_k^2$ is defined on some coordinate patch, and one can even choose local coordinates such that $f_k=x^1$ on this patch, so the term has the form $a^2(dx^1)^2$. Then he modifies F_0 on the coordinate patch to produce an approximately isometric embedding which satisfies the conditions of Theorem 6.2. This allows him to deduce that the embedding can be perturbed to get metric exactly $g_{F_0} + (a^2 dx^1)^2$. This can now be repeated for each of the remaining terms to get the desired isometric embedding. The idea is rather similar to Nash's argument for the C^1 isometric embedding, except that more care must be taken in the approximation, and the compactly supported perturbation result means that we only have to do a finite number of steps instead of an infinite sequence of them.

Günther's method to get approximate isometric embeddings requires five extra dimensions, beyond the span of the first and second derivatives of the embedding — thus to obtain an isometric embedding we must have a free embedding and we must be in dimension at least $\frac{n(n+3)}{2} + 5$. In particular, we are guaranteed to have an isometric embedding into dimension $\max\{\frac{n(n+3)}{2} + 5, \frac{n(n+5)}{2}\}$. For high dimensions this is just $\frac{n(n+5)}{2}$, so in some dimensions this is sharp: There are examples of manifolds which cannot be freely embedded into any dimension less than $\frac{n(n+5)}{2}$ (see the remarks at the end of Lecture 2). For $n = 2$ it gives dimension 10, which is probably far from sharp.

There is a free immersion of S^n into $\mathbb{R}^{\frac{n(n+3)}{2}}$ given by the map

$$(z_1, \ldots, z_{n+1}) \mapsto \left(\frac{z_1^2}{\sqrt{2}}, \ldots, \frac{z_{n+1}^2}{\sqrt{2}}, z_1 z_2, \ldots, z_n z_{n+1} \right).$$

This is in fact a free isometric immersion for the standard metric (or a free isometric embedding for the standard metric on projective space). It follows that we have a free isometric immersion of any metric on S^n or $\mathbb{RP}^n$ into $\mathbb{R}^{\frac{n(n+3)}{2}+5}$. It is an interesting question whether this could be reduced to $\frac{n(n+3)}{2}$.

References

[1] A. D. Aleksandrov, "The intrinsic geometry of convex surfaces", Gostekhizdat, Moscow-Leningrad 1948.

[2] S. Bochner *Analytic mapping of compact Riemann spaces into Euclidean spaces*, Duke Math. J. **3** (1937), 339–354.

[3] C. Burstin, *Ein Beitrag zum problem der Einbettung der Riemannschen Räume euklidischen Räumen*, Mat. Sb. **38** (1931), 74–85.

[4] E. Cartan, *Sur la possibilité de plonger un espace riemannian donné dans un espace euclidien*, Annal. Soc. Polon. Math. **6** (1927), 1–7.

[5] Ja. M. Èliašberg, *Singularities of folding type* (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **34** (1970), 1110–1126.

[6] M. L. Gromov and Ja. M. Èliašberg, *Elimination of singularities of smooth mappings*, Izv. Akad. Nauk SSSR Ser. Mat. **35** (1971), 600–626.

[7] D. Gilbarg and N. Trudinger, "Elliptic Partial Differential Equations of Second Order" Second edition, Springer-Verlag 1983.

[8] H. Grauert, *On Levi's problem and the embedding of real analytic manifolds*, Ann. of Math. **68**, (1958), 460–472.

[9] M. Gromov, *Isometric embeddings and immersions*, Soviet Math. Dokl. **11** (1970), 794–797.

[10] M. Gromov, "Partial Differential Relations", Springer, 1986.

[11] M. Gromov and V. Rokhlin, *Embeddings and immersions in Riemannian geometry*, Usp. Mat. Nauk **25** (1970), 3–62; trans. in Russ. Math. Surv. **25** (1970), 1–57.

[12] M. Günther, *On the perturbation problem associated to isometric embeddings of Riemannian manifolds*, Ann. Global Anal. Geom. **7** (1989), 69–77.

[13] M. Günther, *Zum einbettungsatz von J. Nash*, Math. Nachr. **144** (1989), 165–187.

[14] M. Günther, *Isometric embeddings of Riemannian manifolds*, Proc. ICM Kyoto (1990), 1137–1143.

[15] P.-F. Guan and Y.-Y. Li, *The Weyl problem with nonnegative Gauss curvature*, J. Differential Geom. **39** (1994), 331–342.

[16] R. Hamilton, *The inverse function theorem of Nash and Moser*, Bull. Amer. Math. Soc. **7** (1982), 65–222.

[17] E. Heinz, *On elliptic Monge-Ampre equations and Weyl's embedding problem*, J. Analyse Math. **7** (1959), 1–52.

[18] E. Heinz, *On Weyl's embedding problem*, J. Math. Mech. **11** (1962), 421–454.

[19] M. Hirsch, *On embedding differentiable manifolds in Euclidean space*, Ann. of Math. **73** (1961), 566–571.

[20] M. Hirsch "Differential topology", Springer 1976.

[21] L. Hörmander *The boundary problems of physical geodesy*, Arch. Rat. Mech. Anal. **62** (1976), 1–52.

[22] L. Hörmander, *On the Nash-Moser implicit function theorem*, Lecture Notes, Stanford 1977.

[23] L. Hörmander *On the Nash-Moser implicit function theorem*, Ann. Acad. Sci. Fenn. **10**, 255–259.

[24] H. Jacobowitz, *Implicit function theorems and isometric embeddings*, Ann. Math. **95** (1972), 191–225.

[25] M. Janet, *Sur la possibilité de plonger un espace riemannian donné dans un espace euclidien*, Annal. Soc. Polon. Math. **5** (1926), 422–430.

[26] N. Kuiper, *On C^1 isometric imbeddings I*, Proc. Kon. Acad. Wet. Amsterdam A (Indagationes Math.) **58** (1955), 545–556.

[27] N. Kuiper, *On C^1 isometric imbeddings II*, Proc. Kon. Acad. Wet. Amsterdam A (Indagationes Math.) **58** (1955), 683–689.

[28] H. Lewy *On the existence of a closed convex surface realising a given Riemannian metric*, Proc. Nat. Acad. Sci. USA **24** (1938), 104–106.

[29] C. S. Lin, *The local isometric embedding in R^3 of 2-dimensional Riemannian manifolds with nonnegative curvature*, J. Differential Geom. **21** (1985), 213–230.

[30] C. S. Lin, *The local isometric embedding in R^3 of two-dimensional Riemannian manifolds with Gaussian curvature changing sign cleanly*, Comm. Pure Appl. Math. **39** (1986), 867–887.

[31] S. Lojasiewicz and E. Zehnder, *An inverse function theorem in Fréchet spaces*, J. Funct. Anal. **33** (1979), 165–174.
[32] C. B. Morrey, *The analytic embedding of abstract real analytic manifolds*, Ann. Math. **68** (1958), 159–201.
[33] J. Moser, *A new technique for the construction of solutions of nonlinear differential equations*, Proc. Nat. Acad. Sci. USA **47** (1961), 1824–1831.
[34] J. Moser, *A rapidly convergent iteration method and nonlinear differential equations I*, Ann. Scuola Norm. Sup. Pisa **20** (1966), 265–315; *II*, 499–535
[35] J. Nash *Real algebraic manifolds*, Ann. of Math. **56** (1952),
[36] J. Nash C^1 *isometric embeddings*, Ann. of Math. **60** (1954), 383–396.
[37] J. Nash *The imbedding problem for Riemannian manifolds*, Ann. of Math. **63** (1956), 20–63.
[38] J. Nash *Analyticity of the solutions of implicit function problems with analytic data*, Ann. Math. **84** (1966), 345–355.
[39] L. Nirenberg *The Weyl and Minkowski problems in differential geometry in the large*, Comm. Pure Appl. Math. **6** (1953), 337–394.
[40] L. Nirenberg *An abstract form of the Cauchy-Kowalekski theorem*, J. Differential Geometry **6** (1972), 561–576.
[41] L. Nirenberg, *Variational and topological methods in nonlinear problems*, Bull. Amer. Math. Soc. **4** (1981), 267–302.
[42] A. V. Pogorelov, "Izgibanie vypuklyh poverhnoste". (Russian) [Deformation of convex surfaces] Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow-Leningrad, (1951), 184 pp.
[43] A. V. Pogorelov, *An example of a two-dimensional Riemannian metric that does not admit a local realization in E_3*, Dokl. Akad. Nauk SSSR **198** (1971), 42–43.
[44] L. Schläfli, *Nota alla memoria del. Sig. Beltrami, sugli spazii di curvatura constante*, Ann. di mat., second series, **5** (1871–1873), 170–193.
[45] J. T. Schwarz *On Nash's implicit functional theorem*, Comm. Pure Appl. Math. **13** (1960), 509–530.
[46] J. T. Schwarz, "Nonlinear Functional Analysis", Gordon and Breach, 1969.
[47] F. Sergeraert, *Une généralisation du théorème des fonctions implicites de Nash*, C. R. Acad. Sci. Paris Ser. A **270** (1970), 861–863.
[48] H. Weyl "Die Idee der Riemannschen Flache", Göttingen Lecture notes, 1911–12.
[49] H. Weyl, *Über die bestimmung einer geschlossenen konvexen flache durch ihr linienelement*, Vierteljschr. Naturforsch. Ges. Zürich **61** (1916), 40–72.
[50] H. Whitney *Differentiable manifolds*, Ann. of Math. **37** (1936), 645–680.
[51] H. Whitney *The self-intersection of a smooth n-manifold in $2n$-space*, Ann. of Math. **45** (1944), 220–246.
[52] H. Whitney *The singularities of a smooth n-manifold in $(2n-1)$-space*, Amm. of Math. **45**, (1944), 247–293.
[53] H. Whitney *Analytic extensions of differentiable functions defined in closed sets*, Trans. Amer. Math. Soc. **36** (1934), 63–89.
[54] E. Zehnder, *Generalized implicit function theorems with applications to small divisor problems I*, Comm. Pure Appl. Math. **28** (1975), 91–140.
[55] E. Zehnder, *Generalized implicit function theorems with applications to small divisor problems II*, Comm. Pure Appl. Math. **29** (1976), 49–111.

CMA, ANU, ACT 0200, Australia
E-mail address: andrews@maths.anu.edu.au
URL: http://www.maths.anu.edu.au/~andrews