A Geometric Characterization
of Partial Linearizability

PATRICK D. McSWIGGEN

A fundamental problem that arises in dynamical systems in the study of the local behavior of a diffeomorphism or flow is the linearizability of the system near a fixed point. This is the question of whether there exists a local change of variable converting the system to a linear one. Grobman [G1; G2] and Hartman [Ha2; Ha3] proved that there is always a linearizing change of variable that is a homeomorphism if the fixed point is hyperbolic. A fixed point of a diffeomorphism (resp. flow) is *hyperbolic* if the derivative at the point has no eigenvalues on the unit circle (resp. imaginary axis). Knowing that the change of variable is merely continuous, however, is not always satisfactory. It is natural, therefore, to ask how smooth this change of variable can be.

The question of whether there exists a linearizing change of variable of a certain smoothness goes back at least to Poincaré [P]; see [S1] for the analytic case. Sternberg [S1; S2] proved that, in general, the existence of a smooth linearization depends upon an absence of "resonance" in the eigenvalues of the derivative at the fixed point. This work has been extended by Belitskii [B1; B2], Sell [Se], and others.

If the fixed point is not hyperbolic, then except in trivial cases the system is not linearizable. However, Takens [T] proved that, in this case, if one decomposes the derivative at the fixed point into its center part (the part corresponding to eigenvalues with modulus 1) and its hyperbolic part then, if the eigenvalues of the hyperbolic part satisfy the same Sternberg nonresonance conditions, there will exist a smooth *partial linearization*. For a diffeomorphism this is a local change of variable in which the origin is the fixed point and the system has the form $(\xi, x) \mapsto [\varphi(\xi), B_\xi x]$, where all eigenvalues of $(D\varphi)_0$ have modulus 1 and B_ξ is linear and hyperbolic for each ξ.

In [M1], a characterization of smooth linearizability is given that is completely different in flavor from the Sternberg eigenvalue conditions. There it is shown that, for the special case of hyperbolic attracting or repelling fixed points, smooth linearizability is equivalent to the existence/smoothness of certain invariant submanifolds. In this paper we extend those results to the case where there are eigenvalues on the unit circle. That is, the derivative consists of a nontrivial center part, together with either an attracting or a repelling part. These will be called *partially attracting* or *partially repelling* fixed points, respectively. We show that in this

case the existence of a smooth partial linearization, in the manner of Takens, follows from the existence of smooth invariant submanifolds analogous to those of the hyperbolic case. A corollary is that any partially attracting or repelling fixed point of a C^3 diffeomorphism can be C^1 partially linearized. This is the analog for partially hyperbolic fixed points of a result of Hartman [Ha1], and is not implied by the work of Takens.

By the local nature of the question, we can take the fixed point to be the origin in $\mathbb{R}^n$. Therefore, let f be a local diffeomorphism fixing $0 \in \mathbb{R}^n$, which is at least C^r. Throughout this paper we assume the origin is a partially attracting fixed point. The repelling case can be treated by taking f^{-1}. Consequently, let $1 = \lambda_0 > \cdots > \lambda_m$ be the distinct moduli of the eigenvalues of $(Df)_0$. To each splitting of the spectrum, $\{\lambda_0, \ldots, \lambda_i\} > \{\lambda_{i+1}, \ldots, \lambda_m\}$, there exist complementary, generalized eigenspaces for $(Df)_0$. These are properly referred to as the pseudo-unstable and strong stable subspaces, respectively. However, as there actually are no unstable eigenvalues (those with modulus greater than 1), we will refer to the pseudo-unstable subspace for a given splitting as the center/weak-stable subspace, or center/weak subspace for short.

A general result is that, for the nonlinear function f, there are nonlinear analogs of these subspaces: the center/weak and strong stable manifolds. Locally these are embedded disks tangent to the corresponding subspaces at zero (and so can be given as graphs over these subspaces) that are germwise invariant under f. Moreover, under iteration by f, points in these invariant submanifolds exhibit the same asymptotic rates of expansion/contraction as vectors in the corresponding subspaces do under the linear map.

By the pseudo-stable manifold theorem [HPS], the strong stable manifold is unique and as smooth as the diffeomorphism. The center/weak manifolds, however, are not unique. In this case, the pseudo-stable manifold theorem guarantees the existence of center/weak manifolds of at least a minimum smoothness which depends upon the gap between the two pieces of the spectrum. For the splitting as above, the center/weak manifold will be at least C^s provided $\lambda_i^s > \lambda_{i+1}$. However, this does not imply that there might not exist center/weak manifolds that are more smooth than predicted. We say that f has a complete set of C^r center/weak manifolds if, to every such splitting of the spectrum of $(Df)_0$, there exists at least one center/weak manifold that is C^r. We will prove that the existence of a partial linearization is, up to a slight loss of differentiability, equivalent to the existence of a complete set of center/weak manifolds.

Throughout this paper we will follow the common practice whereby, for any real number $r \geq 0$, a function is said to be *of class C^r* if it possesses continuous derivatives up to order $[r]$ and its $[r]$th-order derivative is (r)-Hölder. Here $[r]$ and (r) denote the integer and fractional parts of r, respectively. The case when the top derivative is 1-Hölder or Lipschitz does not fit this well; this case is written as $C^{k+\mathrm{Lip}}$ ($k \in \mathbb{N}$), where $r = k + \mathrm{Lip}$ is considered distinct (and distinguishable) from $r = k + 1$.

THEOREM 1.   *Assume $0 \in \mathbb{R}^n$ is a partially attracting fixed point for a C^{r+2} ($1 < r < \infty$) germ of a diffeomorphism f. If f has a C^r partial linearization, then f*

has a complete set of C^r center/weak manifolds. Conversely, if f has a complete set of C^r center/weak manifolds, then f has a $C^{r-\varepsilon}$ partial linearization for any $\varepsilon > 0$.

This theorem has its analog for flows. A partial linearization for a flow is a local change of variable converting the original vector field to one of the form $[Y(\xi), A_\xi x]$, where $(DY)_0$ has only pure imaginary eigenvalues and all eigenvalues of A_ξ have nonzero real part. This is equivalent to the change of variable converting the original flow to one of the form $\Phi^t(\xi, x) = [\varphi^t(\xi), B_\xi^t x]$, where φ^t has a center at the origin and B_0^t is hyperbolic for $t \neq 0$. (Here $B_\xi^t = B(t, \xi)$ is not the power of a matrix, but rather the solution of the nonautonomous ODE: $B' = A_{\varphi^t(\xi)} B$ on $L(\mathbb{R}^n, \mathbb{R}^n)$.)

THEOREM 2. *Assume $0 \in \mathbb{R}^n$ is a partially attracting singularity for a C^{r+2} ($1 < r < \infty$) germ of a vector field X. If the local flow of X has a C^r partial linearization, then X has a complete set of C^r center/weak manifolds. Conversely, if X has a complete set of C^r center/weak manifolds, then the local flow of X has a $C^{r-\varepsilon}$ partial linearization for any $\varepsilon > 0$.*

COROLLARY 1. *A partially attracting fixed point of a $C^{2+\mathrm{Lip}}$ diffeomorphism or flow can be C^1 partially linearized.*

Two significant differences between the preceding results and those for purely attracting fixed points is an inherent loss of differentiability for this case, and a lack of uniqueness. For an attracting fixed point, it is shown in [M1] that the conjugacy is uniquely determined by where it sends the weak stable eigenspaces (provided it is at least $C^{1+\varepsilon}$, and assuming its derivative is the identity at the origin). The presence of a center direction in general destroys this uniqueness because the structure in a neighborhood of the fixed point may depend upon properties of the system far away from the point. However, the root source of the uniqueness is still there, and this is enough to obtain Theorem 2 from Theorem 1. (It is also possible to recover this uniqueness if one assumes that the fixed point is topologically attracting.)

REMARK. There is an omission in the statement of uniqueness in Theorem 1 of [M1]. This should read: "Moreover, if $r > 1$, *and the weak unstable manifolds are nested,* the given C^r conjugacy is the only conjugacy between f and its linear part which is ...." (This theorem is stated for a repelling instead of an attracting fixed point.) A similar correction must be made to Theorem 2 of [M1].

A more fundamental difference is that the differentiability obtained for the conjugacy here is arbitrarily close to, but not equal to, the differentiability of the center/weak manifolds. This appears to be unavoidable. Although the loss of differentiability, as measured by ε, can be chosen to be as small as desired, the neighborhood of the origin on which the partial linearization holds shrinks to zero as $\varepsilon \to 0$. (The loss of differentiability actually only pertains to the smoothness in the center direction. The smoothness of the conjugacy in the complimentary direction is the same as the center/weak manifolds. See the remark near the beginning of the proof of Proposition 2.)

Although they differ considerably in the details, the underlying structure of the proof of Theorem 1 is the same as in [M1]. We start by finding a strong stable foliation for each strong versus center/weak stable splitting of the spectrum of $(Df)_0$. A *strong stable foliation* for a particular splitting is an invariant foliation on a neighborhood of the fixed point having the strong stable manifold as a leaf (and hence is generally "parallel" to the strong stable subspace). This is handled by Proposition 1. The presence of the strong stable foliations will allow us to choose coordinates in which f is "triangular". This means that, relative to these coordinates, the ith component of f will depend only on the first i variables. Once f is in this form, we can inductively apply Proposition 2 to obtain the partial linearization.

PROPOSITION 1. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a C^{r+2} diffeomorphism with the origin as a partially attracting fixed point. To each strong versus center/weak stable splitting of the spectrum of $(Df)_0$, there exists a C^{r+1} strong stable foliation of a neighborhood of 0. Moreover, in each case the strong stable foliation can be chosen so that the collection of strong stable foliations is nested. That is, the leaves of a lower-dimensional foliation will be subordinate to (i.e., foliate) the leaves of the higher-dimensional foliations.*

If f is the time-1 map of the local flow of a C^{r+2} vector field with a singularity at the origin, then in addition the strong stable foliations can be chosen to be invariant under the local flow.

For Proposition 2 we assume that $(Df)_0$ has a three-way invariant splitting $E^c \oplus E^w \oplus E^s$. The labels are intended to indicate c̲enter, w̲eak stable, and s̲trong stable subspaces. These will correspond to blocks of $(Df)_0$ which have eigenvalues, respectively: on the unit circle; inside the unit circle but of intermediate size; and of only one modulus that is smaller than all others. Given a matrix L, we will use $|\mathrm{sp}(L)|$ to denote the set of moduli of the eigenvalues of L, and a comparison of that set to a real number means that the comparison holds for each element of the set.

PROPOSITION 2. *Let f be a germ of a $C^{r+1}(1 < r < \infty)$ diffeomorphism fixing $0 \in \mathbb{R}^n$. Suppose, relative to some splitting $E^c \oplus E^w \oplus E^s$, that f has the form $f(\xi, x, y) = [\varphi(\xi), f_1(\xi, x), f_2(\xi, x, y)]$ and $(Df)_0$ is block diagonal,*

$$(Df)_0 =: \begin{bmatrix} U & & \\ & A & \\ & & B \end{bmatrix},$$

such that $|\mathrm{sp}(U)| = 1$, $|\mathrm{sp}(B)| = \beta < 1$, and $\beta < |\mathrm{sp}(A)| < 1$. If f has a C^r center/weak stable manifold tangent to $E^c \oplus E^w$, then for any $s < r$ there exists a C^s local diffeomorphism G conjugating f to $(\xi, x, y) \mapsto [\varphi(\xi), f_1(\xi, x), B_\xi y]$ for some C^r function $\xi \mapsto B_\xi \in L(E^s, E^s)$.

Let ψ^t be a local flow of a C^{r+1} vector field that has a singularity at $0 \in \mathbb{R}^n$. Suppose $\psi^t(\xi, x, y) = [\varphi^t(\xi), \psi_1^t(\xi, x), \psi_2^t(\xi, x, y)]$, and that $f := \psi^1$ satisfies the foregoing hypotheses, where the center/weak stable manifold is invariant under ψ^t. Then, for any $s < r$, there is a C^s local diffeomorphism G such that

$(G \circ \psi^t \circ G^{-1})(\xi, x, y) = [\varphi^t(\xi), \psi_1^t(\xi, x), B_\xi^t y]$ *near* 0 *for some* C^r *function* $B_\xi^t \colon \mathbb{R} \times E^c \to L(E^s, E^s)$.

Moreover, in both cases G can be chosen so that it is the identity on the first two components; that is, $G(\xi, x, y) = [\xi, x, G_2(\xi, x, y)]$.

All of the work in proving Theorem 1 is contained in Proposition 2—which we leave for last. Proposition 1 is a relatively simple application of the C^r section theorem [HPS; Sh].

Proof of Proposition 1. This proof applies more naturally to a partially repelling fixed point, so replace f with its inverse. Let $T := (Df)_0$, and define $\tau := f - T$. Given any $\varepsilon > 0$, we can assume, by cutting τ off with a smooth bump function in a sufficiently small neighborhood of 0, that f is defined on all of $\mathbb{R}^n$, that τ has support in a small neighborhood about 0, and that $\|(D\tau)_p\| < \varepsilon$ for all p. If f is the time-1 map for the local flow of a vector field, we perform this modification instead on the vector field itself. In this way f will have the same properties, but in addition it will still be the time-1 map of a flow.

Let E^{cw} be the center/weak unstable plane and E^{uu} the strong unstable plane of T associated to some splitting $\sigma_{cw} \sqcup \sigma_{uu}$ of the spectrum of T, where $1 \le |\sigma_{cw}| < |\sigma_{uu}|$. Label the blocks of $(Df)_x$ relative to $E^{uu} \oplus E^{cw}$ as follows:

$$(Df)_x =: \begin{bmatrix} A_x & B_x \\ C_x & N_x \end{bmatrix}.$$

At $x = 0$, we have $T = A_0 \oplus N_0$. Since the eigenvalues of A_0 have moduli strictly greater than those of N_0, which are themselves at least 1 in modulus, we can choose $\alpha > \nu > 1$ as well as adapted norms on E^{uu}, E^{cw} such that $\|N_0\| < \nu < \alpha < \|A_0^{-1}\|^{-1}$; moreover, for any $\delta > 0$, we can choose these so that $\|N_0^{-1}\| < 1 + \delta$. Choose κ with $\nu/\alpha < \kappa < 1$, and fix δ so small that $\kappa(1 + \delta)^{r+1} < 1$.

For ε sufficiently small, we will have $\|A_x^{-1}\| < \alpha^{-1}$, $\|N_x\| < \nu$, $\|N_x^{-1}\| < 1 + \delta$, and $(Df^{-1})_x < 1 + \delta$ for all x. Let $L_1(E^{uu}, E^{cw})$ be the unit ball in $L(E^{uu}, E^{cw})$, and let $P \colon \mathbb{R}^n \to L_1(E^{uu}, E^{cw})$ be a plane field on $\mathbb{R}^n$. The image of this plane field under (Df) will be the plane field $\hat{P}$ defined by $\hat{P}_{f(x)} = (C_x + N_x P_x)(A_x + B_x P_x)^{-1}$. Therefore, an invariant plane field for (Df) corresponds to an invariant section of the function F, from $\mathbb{R}^n \times L_1(E^{uu}, E^{cw})$ to itself, defined by $F(x, P) := [f(x), (C_x + N_x P)(A_x + B_x P)^{-1}]$. By Proposition 4.5 of [HP], if ε is sufficiently small then F is well-defined and, moreover, it contracts fibers of the disk bundle $\Pi \colon \mathbb{R}^n \times L_1(E^{uu}, E^{cw}) \to \mathbb{R}^n$ by $\kappa < 1$.

Because f is C^{r+2}, (Df) is C^{r+1} and so F is C^{r+1}. By the C^r section theorem, F has a unique invariant section, which is continuous. Moreover, the invariant section will be C^{r+1} provided $\kappa \operatorname{Lip}(f^{-1})^{r+1} < 1$. However, $\operatorname{Lip}(f^{-1}) = \sup_x \|(Df^{-1})_x\| \le 1 + \delta$. Therefore, (Df) has a unique invariant plane field P such that $P_x \in L_1(E^{uu}, E^{cw})$ for all x, and it is C^{r+1}.

At this point we invoke general stable manifold theory. Since f is C^1 close to T and $\mathbb{R}^n$ is a ρ-pseudo hyperbolic set for T when $|\sigma_{cw}| < \rho < |\sigma_{uu}|$, it follows that if ε is sufficiently small then $\mathbb{R}^n$ will be a ρ-pseudo hyperbolic set for f.

Therefore, by the pseudo-stable manifold theorem, through each point there exists a unique strong unstable manifold that is C^{r+2}. Since these are unique, the collection of strong unstable manifolds forms a foliation of $\mathbb{R}^n$ by C^{r+2} leaves. The tangent plane field to this foliation is continuous, close to E^{uu}, and invariant under (Df). Since P is the only invariant plane field close to E^{uu}, the tangent plane field is P. However, P is then obviously integrable. Therefore, by Froebenius' theorem, the collection of strong unstable manifolds is in fact a C^{r+1} foliation.

Finally, this can be done for each weak unstable versus strong unstable splitting. In each case there will be a minimum ε needed to make it work, which will be achieved by cutting off τ in a sufficiently small neighborhood of the origin. If we decide in advance how small this neighborhood must be for each of the splittings, then we can do this once so that the same modified f can be used for each case. In this way, the collection of strong unstable foliations that we obtain are all invariant foliations for the same function. Because the modified function agrees with the original diffeomorphism near the origin, these foliations are locally invariant under the original f. Now suppose that the original diffeomorphism was the time-1 map of a flow. By construction, the modified f is the time-1 map of the flow φ^t of the modified vector field. The strong unstable manifolds for a flow are the same as for its time-1 map. Therefore, these foliations are invariant under φ^t. Since the modified vector field agrees with the original vector field near the origin, φ^t agrees with the original flow near the origin. Consequently, these foliations are locally invariant under the original flow.

We are left with showing that these foliations can be chosen so they are nested. In fact, as constructed, they already are. From the pseudo-stable manifold theorem, the strong unstable manifolds for a splitting $|\sigma_{cw}| < \rho < |\sigma_{uu}|$ are the equivalence classes of the relation $x_1 \sim x_2$ if and only if $\lim_{n\to\infty} \rho^n \| f^{-n}(x_1) - f^{-n}(x_2) \| = 0$. Given two splittings with $\rho_1 < \rho_2$, since the associated strong unstable manifolds are for the same f, the equivalence classes determined by ρ_2 will clearly be subsets of those for ρ_1. Consequently, the foliations are nested. $\square$

Proof of Theorem 1. The significant part of this proof is the second half—the construction of a partial linearization from a complete set of center/weak manifolds. The first half, that a C^r partial linearization implies a complete set of C^r center/weak manifolds, is just another simple application of the C^r section theorem. Since a partial linearization will take the invariant manifolds of one system to those of another, it is sufficient to show that f has a complete set of C^r center/weak manifolds after applying the partial linearization. Therefore, assume $f(\xi, x) = (\varphi(\xi), B_\xi x)$ is a function from $\mathbb{R}^c \times \mathbb{R}^m$ into itself, where the spectrum of $(D\varphi)_0$ is on the unit circle and B_0 is hyperbolic. Let $\mathbb{R}^w \times \mathbb{R}^s = \mathbb{R}^m$ be the weak stable and strong stable subspaces of B_0 for some splitting of its spectrum. We want to show that f has a C^r center/weak manifold tangent to $\mathbb{R}^c \times \mathbb{R}^w$.

If we view $\mathbb{R}^c \times \mathbb{R}^m$ as a vector bundle over $\mathbb{R}^c$, then f is a bundle map. Suppose, for each $\xi \in \mathbb{R}^c$, we have subspaces $V(\xi) \subset \mathbb{R}^m$ of dimension w such that $V(0) = \mathbb{R}^w \times \{0\}$ and $V(\varphi(\xi)) = B_\xi V(\xi)$. The union of these subspaces would then be a surface invariant under f, tangent to $\mathbb{R}^c \times \mathbb{R}^w$ at 0, and as smooth as V. Therefore,

it will be enough to show there is a C^r section of the bundle $\mathbb{R}^c \times L(\mathbb{R}^w, \mathbb{R}^s) \to \mathbb{R}^c$ that is invariant under the natural action of f. We now proceed as in the proof of Proposition 1. Because B contracts $\mathbb{R}^s$ more strongly than $\mathbb{R}^w$, the natural action of f is a fiber contraction on $\mathbb{R}^c \times L(\mathbb{R}^w, \mathbb{R}^s)$ by some $\kappa < 1$. The base map is φ, and since all eigenvalues of $(D\varphi)_0$ have modulus 1, we can make $\text{Lip}(\varphi^{-1})$ as close to unity as we wish by restricting to a sufficiently small neighborhood of the origin. In particular, since $r < \infty$, we can make $\kappa \, \text{Lip}(\varphi^{-1})^r < 1$. Consequently, we can apply the C^r section theorem and conclude that V exists and is as smooth as f.

We can now address the major part of the proof. Let $1 = \lambda_0 > \cdots > \lambda_m$ be the distinct moduli of eigenvalues of $(Df)_0$. Let E_i be the invariant subspace of $(Df)_0$ corresponding to eigenvalues with modulus λ_i. Relative to coordinates $(x_0, \ldots, x_m) \in E_0 \oplus \cdots \oplus E_m = \mathbb{R}^n$, $(Df)_0$ is block diagonal and λ_i is the modulus of the eigenvalues of block i, B_i. Let $F_i := E_{i+1} \oplus \cdots \oplus E_m$ $(i < m)$. By Proposition 1 there exist a total of m distinct C^{r+1} strong stable foliations of decreasing dimension, tangent to the planes $F_0, \ldots, F_{m-1}$ at 0, and these foliations are nested.

Since $f \colon E_1 \oplus F_1 \to E_1 \oplus F_1$ has an invariant foliation whose leaf through the origin is tangent to F_1, we can choose a foliation chart taking the leaves to the vertical planes, that is, the family of planes parallel to F_1. F_1 further decomposes as $F_1 = E_2 \oplus F_2$, and there is a second invariant foliation that is subordinate to the first and tangent to F_2. Thus, with respect to the first foliation chart, the second foliation will have leaves that are graphs over F_2 and are restricted to single $E_2 \oplus F_2$ planes. Consequently, the second foliation can be taken to the family of planes parallel to F_2 by a change of coordinates that moves points only in the E_2 direction. This means that the planes parallel to $F_1 = E_2 \oplus F_2$, the original foliation, are taken to themselves, and so this is a common foliation chart for both foliations. Continuing in this way from the highest-dimensional foliations to the lowest, we can construct a common foliation chart for all of the strong stable foliations, which is C^{r+1}.

Relative to this common foliation chart, f will be C^{r+1} and for each i will leave the family of planes parallel to F_i invariant. Since these are the sets $\{(x_0, \ldots, x_i) = \text{constant}\}$, this implies that, for each i, the components $f_0, \ldots, f_i$ cannot depend upon $(x_{i+1}, \ldots, x_m)$. This holds for every i, and so we have $f(x_0, \ldots, x_m) = [f_0(x_0), f_1(x_0, x_1), \ldots, f_m(x_0, \ldots, x_m)]$. Since the invariant foliations were already tangent to the subspaces F_i, this can be done without changing $(Df)_0$. Therefore, $(D_{x_i} f_i)_0 = B_i$.

Because E_0 is the generalized eigenspace for eigenvalues of modulus 1, which will be handled differently from the other subspaces, we distinguish this variable by writing $x_0 =: \xi$ and denoting f_0 by φ, so $f_0(x_0) = \varphi(\xi)$. By induction we prove that there is a $C^{r-\varepsilon}$ partial linearization for f. Suppose that for some $i \geq 1$ there is a $C^{r-\varepsilon}$ conjugacy $\hat{G}$ from f to $\hat{f}$, where

$$\hat{f}(\xi, x_1, \ldots, x_m)$$
$$= [\varphi(\xi), f_1(\xi, x_1), \ldots, f_i(\xi, x_1, \ldots, x_i), B_{i+1}(\xi)x_{i+1}, \ldots, B_m(\xi)x_m],$$

and where $B_j(\xi)$ is a C^r function from E_0 to $L(E_j, E_j)$ with $B_j(0) = B_j$. In addition, we assume $\hat{G}$ is the identity on the first $i + 1$ components, that is, $\hat{G}(x) = (x_0, \ldots, x_i, \hat{g}_{i+1}(x), \ldots, \hat{g}_m(x))$. The base case for the induction is $i = m$, where $f = \hat{f}$, and is vacuously true. We want to show there is a subsequent $C^{r-\varepsilon}$ conjugacy G that converts $f_i(\xi, x_1, \ldots, x_i)$ to $B_i(\xi)x_i$ while leaving the other components unchanged.

Let $x := (x_1, \ldots, x_{i-1})$, $y := x_i$, and $z := (x_{i+1}, \ldots, x_m)$. Depending on i, either x or z may span the zero subspace. Recollect the components of $\hat{f}$ so that, relative to these new variables, $\hat{f}(\xi, x, y, z) = [\varphi(\xi), \hat{f}_1(\xi, x), \hat{f}_2(\xi, x, y), \hat{B}(\xi)z]$. We then have that φ, $\hat{f}_1$, $\hat{f}_2$ are C^{r+1}, $(D\varphi)_0$ has eigenvalues on the unit circle, $(D_y \hat{f}_2)_0$ has eigenvalues of only a single modulus $\beta = \lambda_i < 1$, and the moduli of the eigenvalues of $(D_x \hat{f}_1)_0$ are in $\{\lambda_1, \ldots, \lambda_{i-1}\}$, which are strictly between 1 and β.

By assumption we have a complete set of C^r center/weak manifolds for f, including a C^r center/weak manifold tangent to $E_0 \oplus \cdots \oplus E_{i-1}$, that is, the span of (ξ, x). This is the graph of a C^r function $h \colon E_0 \oplus \cdots \oplus E_{i-1} \to F_i$. In other words, it has the form $\{(\xi, x, h_2(\xi, x), h_3(\xi, x))\}$, where h_2, h_3 are C^r. $\hat{G}$ takes this to a $C^{r-\varepsilon}$ center/weak manifold for $\hat{f}$. However, $\hat{G}$ is the identity on the first three terms. Consequently, the center/weak manifold for $\hat{f}$ can be written as $\{(\xi, x, h_2(\xi, x), \hat{h}_3(\xi, x))\}$, where h_2 is unchanged, and so is still C^r.

Because the first three components of $\hat{f}$ do not depend upon z, the graph of h_2 alone, $\{(\xi, x, h_2(\xi, x))\}$, must be invariant under $[\varphi(\xi), \hat{f}_1(\xi, x), \hat{f}_2(\xi, x, y)]$. Hence, the restricted system has a C^r center/weak manifold tangent to the span of (ξ, x) and so satisfies the hypotheses of Proposition 2. From this we can conclude that there exists a $C^{r-\varepsilon}$ conjugacy from the restricted system to $[\varphi(\xi), \hat{f}_1(\xi, x), B(\xi)y]$. Moreover, this conjugacy has the form $(\xi, x, g(\xi, x, y))$. Setting $G(\xi, x, y, z) = (\xi, x, g(\xi, x, y), z)$ gives us the desired conjugacy for the full system. $G \circ \hat{G}$ now satisfies the induction hypotheses for $i - 1$. Consequently, we can continue to linearize f term by term to get a $C^{r-\varepsilon}$ partial linearization. $\square$

Proof of Theorem 2. Let ψ^t be the local flow of X, and let $f = \psi^1$ be the time-1 map of ψ^t. The center/weak subspaces for $(DX)_0$ and $(Df)_0$ agree, and a submanifold that is invariant under ψ^t will be invariant under f. Therefore, f has a complete set of C^r center/weak manifolds. The proof now proceeds the same as Theorem 1, except upon applying Proposition 2, we use that the conjugacy holds for the full flow. $\square$

Proof of Corollary 1. As observed earlier, there will be guaranteed a center/weak manifold that is C^s provided $\lambda_i^s > \lambda_{i+1}$. Since $\lambda_i > \lambda_{i+1}$ for each i, there will exist some $\alpha > 0$ such that $\lambda_i^{1+\alpha} > \lambda_{i+1}$ for all i. Therefore, there is a complete set of $C^{1+\alpha}$ center/weak manifolds. Consequently, if f is $C^{3+\delta}$ for some δ ($0 < \delta \leq \alpha$), we can apply Theorem 1 with $r = 1 + \delta$ and conclude that there is a $C^{1+\delta-\varepsilon}$ partial linearization, and so in particular there is a C^1 partial linearization. Therefore, the corollary holds for any f that is C^s with $s > 3$. However, we can get

away with $s = 2 + \text{Lip}$ by using a slightly stronger version of Proposition 1. It is shown in [M2] that the loss in differentiability between the diffeomorphism and the strong stable foliation is strictly less than 1. Consequently, if f is $C^{2+\text{Lip}}$, we can find strong stable foliations that are $C^{2+\delta}$ for some $\delta > 0$. This means that in the proof of Theorem 1 we are applying Proposition 2 to a system that is C^{r+1}, where $r = 1 + \delta$. This is effectively the same as starting with a system that is $C^{3+\delta}$, and so there exists a C^1 partial linearization. $\qquad\square$

We are left with proving Proposition 2.

Proof of Proposition 2. We make the following notational conventions. Given a subspace $E \subset \mathbb{R}^n$, $E(t)$ will be the closed ball of radius t about $0 \in E$. When we wish to group the subspaces E^c, E^w, E^s, we will use $E^{cw} := E^c \oplus E^w$ and $E^{ws} := E^w \oplus E^s$. We denote points in these subspaces by $u := (\xi, x) \in E^{cw}$ and $v := (x, y) \in E^{ws}$, and points in $\mathbb{R}^n$ by p. Hence, $p = (u, y) = (\xi, v) = (\xi, x, y)$. The partial derivatives corresponding to these subspaces will be denoted $D_\xi, D_u, D_y, \ldots$. In the same vein, when regrouping the components of f we will use $\hat{f}_1(u) := [\varphi(\xi), f_1(\xi, x)]$ and $\mu(\xi, v) := [f_1(\xi, x), f_2(\xi, x, y)]$, so $f = \varphi \oplus \mu = \hat{f}_1 \oplus f_2$.

For $0 < \theta \leq \text{Lip}$, $H_\theta[\cdot]$ will denote the θ-Hölder constant of $(\cdot)$, and we will use H_θ^ξ and H_θ^v to denote the (uniform) θ-Hölder constant with respect to ξ and v, respectively. More precisely, for a function τ on $E^c \oplus E^{ws}$, taking values in some Banach space, we define

$$H_\theta^\xi[\tau] := \sup_{\xi_1 \neq \xi_2} \sup_v \frac{\|\tau(\xi_1, v) - \tau(\xi_2, v)\|}{\|\xi_1 - \xi_2\|^\theta},$$

where the supremums are over the domain of τ, and similarly for H_θ^v. Finally, we will use $\|\cdot\|_{C^r}$ to denote the C^r norm on a space of C^r functions:

$$\|f\|_{C^k} := \max\left\{ \sup_p \|(D^i f)_p\| : i = 0, \ldots, k \right\} \quad (k \in \mathbb{N});$$

$$\|f\|_{C^{k+\theta}} := \max\{\|f\|_{C^k}, H_\theta[(D^k f)]\} \quad (0 < \theta \leq \text{Lip}).$$

We need to show that, given a C^r center/weak manifold, there exists a C^s conjugacy for any $s < r$. Therefore, fix $s < r$. There are two possibilities. If s is sufficiently close to r, then either r is not an integer and $[r] < s < r$, or r is an integer and $r - 1 < s < r$. In either case, we can assume that s is not an integer. Therefore, set $s =: k + \theta$, where $k \in \mathbb{N}$ and $0 < \theta < \text{Lip}$, and choose δ with $0 < \delta \leq r - s$ and such that $\theta + \delta \leq \text{Lip}$.

Because $r < \infty$, there is a sufficiently small neighborhood of 0 on which f has a center manifold as smooth as f. We can therefore assume, without loss of generality, that E^c is invariant. The center/weak manifold of f is tangent to E^{cw} and so is locally the graph of a C^r function $h: E^{cw}(d) \to E^s$. If we make an initial change of variable by $H(u, y) = [u, y - h(u)]$, then $H \circ f \circ H^{-1}(u, y) = [\hat{f}_1(u), f_2(u, y + h(u)) - h \circ \hat{f}_1(u)]$ leaves E^{cw} invariant. Although $f \in C^{r+1}$, h is only C^r and so $H \circ f \circ H^{-1}$ is only C^r. On the other hand, $\hat{f}_1$ has not

changed and so is still C^{r+1}. The partial with respect to y of the second component is $(D_y f_2)_{H^{-1}(u,y)}$, which is the composition of C^r functions, and so is C^r. Therefore, the second component is C^{r+1} with respect to y. Moreover, since the graph of h is tangent to E^{cw} at 0, $(Dh)_0 = 0$ and so $(DH)_0 = I$. Therefore, $D(H \circ f \circ H^{-1})_0 = (Df)_0$.

A priori, E^c may no longer be invariant under $H \circ f \circ H^{-1}$. However, the family of planes parallel to E^s is an invariant foliation for f, and E^c is an invariant manifold. Therefore, the plane $E^c \oplus E^s$ is invariant under f. Since both $E^c \oplus E^s$ and the graph of h are invariant under f, their intersection is invariant, and their intersection is precisely the set that is taken to E^c by H. Therefore, even though H may not preserve E^c, it takes another invariant set to E^c and so E^c is still invariant under $H \circ f \circ H^{-1}$. Consequently, both E^c and E^{cw} are invariant under the new system.

Having made these preliminary changes, we will again use f to denote the modified diffeomorphism. Therefore, we now have

$$f(\xi, x, y) = [\varphi(\xi), f_1(\xi, x), f_2(\xi, x, y)],$$

where φ, f_1 are C^{r+1} and where f_2 is C^r and is C^{r+1} in y. Moreover, both E^c and E^{cw} are invariant under f, which implies $f_2(\xi, x, 0) \equiv 0 \equiv f_1(\xi, 0)$.

Let $\alpha := \max|\text{sp}(A)|$, and fix κ with $\alpha^\delta < \kappa < 1$. Fix $\varepsilon > 0$ sufficiently small that $(\beta + \varepsilon)(\beta^{-1} + \varepsilon)(1 + \varepsilon)^s(\alpha + \varepsilon)^\delta < \kappa < 1$. Since $|\text{sp}(B)| = \beta$ and $|\text{sp}(U)| = 1$, we can make a linear change of variable converting A, B, and U to real ε-Jordan form so that $\|B\| < \beta + \varepsilon$, $\|B^{-1}\| < \beta^{-1} + \varepsilon$, $\|A\| < \alpha + \varepsilon$, and $\|U\| < 1 + \varepsilon$. Take the norm on $E^c \oplus E^w \oplus E^s$ to be the box norm, $\|(\xi, x, y)\| := \max\{\|\xi\|, \|x\|, \|y\|\}$. Let O be the orthogonal matrix that is the real Jordan form of U minus the off-diagonal entries. Then we can assume $\|U - O\| < \varepsilon$.

Write $f = (Df)_0 + \tau$. By cutting τ off with a smooth bump function outside a small neighborhood of 0, we can assume that f is defined on all of $\mathbb{R}^n$ and is linear outside a small ball; given any $\varepsilon' > 0$, we can also assume $\sup_p \|(D\tau)_p\| < \varepsilon'$. Since E^c and E^{cw} are invariant under both $(Df)_0$ and τ, the new f will still leave these invariant and will be as smooth as before. If ε' is small enough, then $\|(D\varphi)_p\| \leq 1 + \varepsilon$, $\|(Df_1)_p\| \leq \alpha + \varepsilon$, and $\|(Df_2)_p\| \leq \beta + \varepsilon$ for all p. Moreover, the new f will still be invertible and $\|D(f^{-1})_2\| \leq \beta^{-1} + \varepsilon$.

REMARK. The size of the ball on which the new f agrees with the old is dependent upon ε, and shrinks as $\varepsilon \to 0$. Moreover, ε is a function of δ, the loss of differentiability, and $\varepsilon \to 0$ as $\delta \to 0$. Therefore, as observed in the comments after the statement of Corollary 1, in the absence of additional information (e.g., the fixed point is topologically attracting), the neighborhood on which the partial linearization is guaranteed to exist disappears as $\delta \to 0$.

It will be convenient to work on a compact invariant neighborhood of 0, so we make one last modification to f. On E^c we have $\varphi(\xi) = U\xi$ outside $E^c(t)$ for some small t. Between $E^c(t)$ and $E^c(1)$, smoothly interpolate between U and O so that $\varphi(\xi) = O\xi$ near $\|\xi\| = 1$. Since $\|U - O\|$ can be made as small as we wish, we can still assume $\|(D\tau)_p\| < \varepsilon'$ for all p. Because O is an isometry,

$O(E^c(1)) = E^c(1)$. Moreover, $\|(Df_1)_p\| \le \alpha + \varepsilon < 1$ and $\|(Df_2)_p\| \le \beta + \varepsilon < 1$. Therefore, f maps $\mathbb{R}^n(1)$ into itself.

With these preliminary modifications, we now have a function $f = (\varphi, f_1, f_2)$ that satisfies the following. It is equal (after a C^r change of variable) to our original f near the origin, $f[\mathbb{R}^n(1)] \subset \mathbb{R}^n(1)$, and E^c, E^{cw} are invariant. Both φ and f_1 are C^{r+1}, f_2 is C^r and is C^{r+1} in y, and $\|(Df)_p - (Df)_0\| < \varepsilon$ for all p.

We are finally ready to construct G. Define $B_\xi := (D_y f_2)_{(\xi, 0, 0)}$. Since f_2 is C^{r+1} in y, B_ξ is C^r. Define $F(\xi, x, y) := [\varphi(\xi), f_1(\xi, x), B_\xi y]$. We are looking for a conjugacy between f and F—that is, a diffeomorphism G such that $F \circ G = G \circ f$. Moreover, we want G to have the form $G(\xi, x, y) = [\xi, x, y + g(\xi, x, y)]$. Writing this out, we have $B_\xi(y + g) = f_2 + g \circ f$. If we define $\rho := f_2 - B_\xi y$, this can be written as $g = B_\xi^{-1} g \circ f + B_\xi^{-1} \rho =: \Phi(g)$. Therefore, a fixed point of Φ would give us the desired conjugacy.

Observe that Φ is an affine map, and therefore we look for a fixed point in a function space that is matched to ρ. From its definition, $\rho(\xi, x, y)$ is C^r, and C^{r+1} in y. Moreover, $\rho(\xi, x, 0) = f_2(\xi, x, 0) - B_\xi(0) = 0$, and $(D_y \rho)_{(\xi, 0, 0)} = (D_y f)_{(\xi, 0, 0)} - B_\xi = 0$. Let $\mathcal{G}$ be the set of C^s functions $g \colon \mathbb{R}^n(1) \to E^s$ such that $g(\xi, x, y)$ is C^{s+1} in y, and for which $g(\xi, x, 0) \equiv 0$ and $(D_y g)_{(\xi, 0, 0)} \equiv 0$. We claim Φ maps $\mathcal{G}$ into itself.

Because $s < r$, $\rho \in \mathcal{G}$. Moreover, for any $g \in \mathcal{G}$, $B_\xi^{-1} g$ will again be in $\mathcal{G}$. Therefore, to show that $\mathcal{G}$ is invariant under Φ we need only show $g \circ f \in \mathcal{G}$ for any $g \in \mathcal{G}$. Clearly, $g \circ f$ is C^s; since, in f, only f_2 depends on y, it follows that $D_y[g \circ f]_{(\xi, x, y)} = (D_y g)_{f(\xi, x, y)}(D_y f)_{(\xi, x, y)}$, where both terms are at least C^s. Hence, $g \circ f$ is C^{s+1} in y. Since $f(\xi, 0, 0) = (\varphi(\xi), 0, 0)$, we have $(D_y g)_{f(\xi, 0, 0)} = 0$ and so $D_y[g \circ f]_{(\xi, 0, 0)} = 0$. Likewise, $f(u, 0) = (\hat{f}_1(u), 0)$. Therefore, $(g \circ f)(u, 0) = g(\hat{f}_1(u), 0) = 0$. Consequently, $g \circ f \in \mathcal{G}$, from which it follows $\mathcal{G}$ is invariant under Φ.

We want a fixed point of Φ in $\mathcal{G}$, and to exhibit this fixed point we convert $\mathcal{G}$ to another form. Define $\partial_y \colon \mathcal{G} \to C^s[\mathbb{R}^n(1), L(E^s, E^s)]$ by $\partial_y(g) := (D_y g)$, and let $\mathcal{D}$ be the image of $\mathcal{G}$ under ∂_y. $C^s[\mathbb{R}^n(1), L(E^s, E^s)]$ is a Banach space under the C^s norm. We claim that ∂_y is a bijection onto a closed subspace.

LEMMA 1. $\partial_y \colon \mathcal{G} \to \mathcal{D}$ *is one-to-one, and* $\mathcal{D} \subset C^s[\mathbb{R}^n(1), L(E^s, E^s)]$ *is a closed subspace. Hence,* $\mathcal{D}$ *is a Banach space under the* C^s *norm.*

Proof. ∂_y is linear, so $\mathcal{D}$ is a subspace. We need to show ∂_y is one-to-one and its image is closed. $\mathbb{R}^n(1)$ is convex and so, given any $g \in \mathcal{G}$, we can integrate $(D_y g)$ along the line segment joining $(u, 0)$ and (u, y). Using $g(u, 0) \equiv 0$, we have

$$g(u, y) = g(u, y) - g(u, 0) = \int_0^1 (D_y g)_{(u, ty)}(y)\, dt.$$

Therefore, $(D_y g)$ uniquely determines g, and ∂_y is one-to-one. Differentiating both sides of this expression, under the integral on the right since all derivatives of $(D_y g)$ are continuous, we get the derivatives of g in terms of the derivatives of $(D_y g)$. Consequently, if $(g_n) \subset \mathcal{G}$ is a sequence such that $(D_y g_n) \to N \in \partial \mathcal{D}$,

then $(D_y g_n)$ is Cauchy and so (g_n) is a Cauchy sequence with respect to the usual C^s topology on $\mathcal{G}$. Therefore, it converges to some C^s function g. On the other hand, $(D_y g_n)$ is then converging uniformly to $(D_y g)$. Therefore, $(D_y g) = N$, and g is C^{s+1} in y. If $g_n(u, 0) \equiv 0$ and $(D_y g_n)_{(\xi, 0, 0)} \equiv 0$ for all n, then the same holds for g and so $g \in \mathcal{G}$. Therefore, $\partial_y(g) = N$ is in $\mathcal{D}$, and $\mathcal{D}$ is closed. $\qquad\square$

Because Φ maps $\mathcal{G}$ into itself, the identification between $\mathcal{G}$ and $\mathcal{D}$ given by ∂_y induces a new map $\Theta \colon \mathcal{D} \to \mathcal{D}$ determined by $\Theta \circ \partial_y = \partial_y \circ \Phi$. This is just the map $(D_y g) \mapsto D_y[\Phi(g)] = B_\xi^{-1}(D_y g)_{f(u, y)}(D_y f)_{(u, y)} + B_\xi^{-1}(D_y \rho)_{(u, y)}$. Consequently, rather than looking for a fixed point of Φ in $\mathcal{G}$, we can look for a fixed point of the induced map Θ in $\mathcal{D}$.

Having translated the problem to one on $\mathcal{D}$, we will now work almost exclusively in terms of $(\xi, v) \in E^c \oplus E^{ws}$. The proof proceeds as follows. We produce yet another subspace $\mathcal{N} \subset \mathcal{D}$, which is not closed but does have a fixed point of Θ. Moreover, this fixed point is globally attracting in $\mathcal{N}$ relative to the C^s topology and so is unique in $\mathcal{N}$.

The subspace $\mathcal{N}$ is defined in terms of a modification of the C^s norm. Toward that end, we make the following observations. Continuous higher-order derivatives are symmetric and so, for $N \in \mathcal{D}$, $(D^i N)_{(\xi, v)}$ can be reconstructed from the derivatives $(D_v D^{i-1} N)_{(\xi, v)}$ and $(D_\xi^i N)_{(\xi, v)}$. In fact, if π_ξ and π_v are the projections onto E^c and E^{ws} respectively, and if $w_1, \ldots, w_i \in \mathbb{R}^n$, then $w_j = \pi_\xi w_j + \pi_v w_j$ and we can expand $(D^i N)_{(\xi, v)}[w_1, \ldots, w_i]$ as

$$
\begin{aligned}
(D^i N)_{(\xi, v)}&[w_1, \ldots, w_i] \\
&= (D_\xi^i N)_{(\xi, v)}[\pi_\xi w_1, \ldots, \pi_\xi w_i] \\
&\quad + \sum_{j=1}^{i}(D_v D^{i-1} N)_{(\xi, v)}[\pi_v w_j, \pi_\xi w_1, \ldots, \pi_\xi w_{j-1}, w_{j+1}, \ldots, w_i].
\end{aligned}
$$

Therefore,

$$
\|(D^i N)_{(\xi, v)}\| \leq (i + 1) \max\big\{\|(D_v D^{i-1} N)_{(\xi, v)}\|, \|(D_\xi^i N)_{(\xi, v)}\|\big\}.
$$

On the other hand, one also has

$$
\max\big\{\|(D_v D^{i-1} N)_{(\xi, v)}\|, \|(D_\xi^i N)_{(\xi, v)}\|\big\} \leq \|(D^i N)_{(\xi, v)}\|.
$$

Since these hold for all (ξ, v), we obtain both of the following for $i = 1, \ldots, k$:

$$
\frac{1}{i + 1}\|(D^i N)\|_{C^0} \leq \max\big\{\|(D_v D^{i-1} N)\|_{C^0}, \|(D_\xi^i N)\|_{C^0}\big\} \leq \|(D^i N)\|_{C^0};
$$

$$
\frac{1}{k + 1}H_\theta[(D^k N)] \leq \max\big\{H_\theta[(D_v D^{k-1} N)], H_\theta[(D_\xi^k N)]\big\} \leq H_\theta[(D^k N)].
$$

Similarly, we can decompose the θ-Hölder constant of $(D_\xi^k N)$ into its ξ- and v-parts. For any function we have

$$
H_\theta[\tau] \leq H_\theta^\xi[\tau] + H_\theta^v[\tau] \leq 2\max\{H_\theta^\xi[\tau], H_\theta^v[\tau]\}.
$$

Therefore,

$$\frac{1}{2} H_\theta [(D_\xi^k N)] \leq \max \big\{ H_\theta^\xi [(D_\xi^k N)], H_\theta^\nu [(D_\xi^k N)] \big\} \leq H_\theta [(D_\xi^k N)].$$

Putting these together yields the following lemma.

LEMMA 2. *For $N \in \mathcal{D}$, define $\|N\|_*$ by*

$$\|N\|_* = \max \big\{ \|N\|_{C^0}, \|(D_\nu N)\|_{C^0}, \|(D_\xi N)\|_{C^0},$$
$$\dots, \|(D_\nu D^{k-1} N)\|_{C^0}, \|(D_\xi^k N)\|_{C^0},$$
$$H_\theta [(D_\nu D^{k-1} N)], H_\theta^\nu [(D_\xi^k N)], H_\theta^\xi [(D_\xi^k N)] \big\}.$$

Then $\| \cdot \|_*$ and $\| \cdot \|_{C^s}$ are equivalent norms on $\mathcal{D}$.

For a function τ on $E^c \oplus E^{ws}$, we will use $\|\tau\|_\delta$ to denote the following modification of the C^0 norm of τ: $\|\tau\|_\delta := \sup \{ \|\tau(\xi, \nu)\| / \|\nu\|^\delta : \|\xi\| \leq 1, \, 0 < \|\nu\| \leq 1 \}$. We will use the same notation regardless of the range of τ. Since $\|\nu\| \leq 1$, we have $\|\tau\|_{C^0} \leq \|\tau\|_\delta \leq \infty$. We will need to make a similar change to $H_\theta^\xi [(D_\xi^k N)]$. In this case, let $Q_{\theta,\delta}[\cdot]$ denote the following:

$$Q_{\theta,\delta}[\tau] := \sup_{\substack{\xi_1 \neq \xi_2 \\ \nu \neq 0}} \frac{\|\tau(\xi_1, \nu) - \tau(\xi_2, \nu)\|}{\|\xi_1 - \xi_2\|^\theta \|\nu\|^\delta}.$$

It follows that $H_\theta^\xi [\tau] \leq Q_{\theta,\delta}[\tau] \leq \infty$.

LEMMA 3. *For $N \in \mathcal{D}$, define $\|\|N\|\|$ by*

$$\|\|N\|\| = \max \big\{ \|N\|_\delta, \|(D_\nu N)\|_{C^0}, \|(D_\xi N)\|_\delta,$$
$$\dots, \|(D_\nu D^{k-1} N)\|_{C^0}, \|(D_\xi^k N)\|_\delta,$$
$$H_{\theta+\delta}[(D_\nu D^{k-1} N)], H_{\theta+\delta}^\nu [(D_\xi^k N)], Q_{\theta,\delta}[(D_\xi^k N)] \big\}.$$

Let $\mathcal{N} := \{ N \in \mathcal{D} : \|\|N\|\| < \infty \}$. Then $(\mathcal{N}, \|\| \cdot \|\|)$ is a Banach space. Moreover, convergence in $\mathcal{N}$ with respect to $\|\| \cdot \|\|$ implies convergence in the C^s norm.

Proof. It is clear that $\|\| \cdot \|\|$ is a norm, and term-by-term $\| \cdot \|_* \leq \|\| \cdot \|\|$. Therefore, convergence in $\|\| \cdot \|\|$ implies convergence in $\| \cdot \|_*$, which in turn implies convergence in $\| \cdot \|_{C^s}$. Consequently, we need only show that $\mathcal{N}$ is complete under $\|\| \cdot \|\|$. However, again since $\| \cdot \|_* \leq \|\| \cdot \|\|$, if (N_j) is Cauchy with respect to $\|\| \cdot \|\|$ then it is Cauchy with respect to $\| \cdot \|_{C^s}$. Hence, N_j converges to some $N \in \mathcal{D}$ in the C^s topology. On the other hand, since (N_j) is Cauchy with respect to $\|\| \cdot \|\|$, $\|\|N_j\|\|$ is bounded $\leq K$ for some K, and if $N_j \to N$ in the C^s topology, it follows that $\|\|N\|\| \leq K$. Therefore, $N \in \mathcal{N}$, and $\mathcal{N}$ is complete. $\square$

We want to show that Θ maps $\mathcal{N}$ into itself and has a fixed point which attracts all other elements of $\mathcal{N}$ under iteration by Θ. Since Θ is an affine map, this will

follow if we show $B_\xi^{-1}(D_y\rho) \in \mathcal{N}$ and $N \mapsto B_\xi^{-1}(N \circ f)(D_y f)$ is a bounded linear operator on $\mathcal{N}$ with spectrum strictly inside the unit circle. The first half is handled by the following lemma.

LEMMA 4.   *If $N \in \mathcal{D}$ is $C^{s+\delta}$, then $N \in \mathcal{N}$.*

Proof. We need to show that the terms of $\|\|N\|\|$ which differ from $\|N\|_*$ are finite if $N \in \mathcal{D}$ is $C^{s+\delta}$. These are the terms $\|(D_\xi^i N)\|_\delta$ ($i = 0, \ldots, k$) as well as the last terms. If $N \in \mathcal{D}$, then $N = (D_y g)$ for some $g \in \mathcal{G}$; hence $N_{(\xi,0)} = (D_y g)_{(\xi,0)} \equiv 0$. Therefore, $(D_\xi^i N)_{(\xi,0)} = 0$ for all $i = 0, \ldots, k$ and all ξ. Consequently, $\|(D_\xi^i N)_{(\xi,v)}\| = \|(D_\xi^i N)_{(\xi,v)} - (D_\xi^i N)_{(\xi,0)}\| \le H_\delta[(D_\xi^i N)]\|v\|^\delta$. Since $i < s$, $(D_\xi^i N)$ is at least C^δ, and $\|(D_\xi^i N)\|_\delta \le H_\delta[(D_\xi^i N)] < \infty$.

Because $(D^k N)$ is $C^{\theta+\delta}$, the first two of the last three terms are bounded by $K := H_{\theta+\delta}[(D^k N)]$. Likewise, $\|(D_\xi^k N)_{(\xi_1,v)} - (D_\xi^k N)_{(\xi_2,v)}\| \le K \|\xi_1 - \xi_2\|^{\theta+\delta}$. However, we also have $\|(D_\xi^k N)_{(\xi,v)}\| = \|(D_\xi^k N)_{(\xi,v)} - (D_\xi^k N)_{(\xi,0)}\| \le K \|v\|^{\theta+\delta}$. Therefore,

$$\|(D_\xi^k N)_{(\xi_1,v)} - (D_\xi^k N)_{(\xi_2,v)}\| \le 2K \min\{\|\xi_1 - \xi_2\|^{\theta+\delta}, \|v\|^{\theta+\delta}\}$$
$$\le 2K \|\xi_1 - \xi_2\|^\theta \|v\|^\delta,$$

and so $Q_{\theta,\delta}[(D_\xi^k N)] < \infty$. $\square$

Let χ be the linear part of Θ, that is, $\chi(N)_{(\xi,v)} := B_\xi^{-1} N_{f(\xi,v)}(D_y f)_{(\xi,v)}$ for $N \in \mathcal{D}$. We want to show $\chi \in L(\mathcal{N}, \mathcal{N})$ and $|\mathrm{sp}(\mathcal{N})| < 1$, or equivalently, that there is a new norm on $\mathcal{N}$ with respect to which $\|\chi\| < 1$.

LEMMA 5.   *There exists a norm $\|\|\cdot\|\|'$ on $\mathcal{N}$, equivalent to $\|\|\cdot\|\|$, for which $\|\chi\| < 1$.*

The proof of this lemma is very technical, and we defer it for the moment. We first use Lemma 5 to complete the proof of Proposition 2. If there exists a norm for which $\|\chi\| < 1$, then χ (and hence Θ) is a contraction of $\mathcal{N}$ into itself with respect to this norm. Therefore, Θ has a unique fixed point in $\mathcal{N}$. Since $\mathcal{N} \subset \mathcal{D}$, this corresponds to a (not necessarily unique) fixed point g of Φ in $\mathcal{G}$. By design, $G = \mathrm{id} + (0, 0, g)$ is then a C^s function that linearizes the last component of f; that is, it solves $F \circ G = G \circ f$. Moreover, $g(\xi, x, 0) \equiv 0$ and so $(D_u g)_{(\xi,x,0)} \equiv 0$ and, by construction, $(D_y g)_{(\xi,0,0)} \equiv 0$. Therefore, $(Dg)_0 = 0$. Consequently, $(DG)_0 = I + [0, 0, (Dg)_0] = I$, which implies that G is a diffeomorphism near 0.

Recall that this f differs from the original diffeomorphism. If we return to using f to denote the original diffeomorphism, then G is a conjugacy between F and a diffeomorphism that agrees with $H \circ f \circ H^{-1}$ on a neighborhood of the origin. Therefore, $G \circ H$ is a conjugacy between f and F near 0. Finally, both G and H are the identity on (ξ, x), so the same is true for $G \circ H$, which completes the proof for diffeomorphisms.

Now suppose that ψ^t is a local flow satisfying the hypotheses of the proposition, and that $f = \psi^1$. We can apply the previous case to obtain a conjugacy G for f. However, in order for G to be a conjugacy for the full flow, we will need the

modified f used in the proof to still be the time-1 map of a flow agreeing with the original flow near 0. This can be accomplished by making the modifications to the vector field itself. As before, we make a C^r change of variable so that E^c and E^{cw} are invariant. However, in general, a C^r change of variable produces only a C^{r-1} vector field. Therefore, we need to verify that we do not lose any differentiability.

Let X denote the original vector field. Since ψ^t is "triangular", X has the form $X(\xi, x, y) = [X_0(\xi), X_1(\xi, x), X_2(\xi, x, y)] = [\hat{X}_1(u), X_2(u, y)]$. Let the center/weak manifold be the graph of $h \colon E^{cw} \to E^s$, and set $H(u, y) = [u, y + h(u)]$. A change of variable by H produces the vector field $Y(p) = (DH)_p^{-1}(X \circ H)(p) = [\hat{X}_1(u), X_2 \circ H(u, y) - (Dh)_u \hat{X}_1(u)]$. This involves (Dh), which is only C^{r-1}. On the other hand, Y leaves E^{cw} invariant. Hence the last component of Y must be zero when $y = 0$. Therefore, $(Dh)_u \hat{X}_1(u) = X_2 \circ H(u, 0)$, and $Y(p) = [\hat{X}_1(u), X_2 \circ H(u, y) - X_2 \circ H(u, 0)]$, which is clearly C^r.

We can now proceed as we did before. We first make a change of variable that takes a C^{r+1} center manifold to E^c. By the foregoing argument, this does not affect the differentiability of X. We then apply the change of variable to obtain the vector field Y. As with the discrete case, both E^c and E^{cw} are necessarily invariant under Y. Moreover, $\hat{Y}_1 = \hat{X}_1$ is C^{r+1}, and $(D_y Y_2)$ is C^r, so Y_2 is C^r and it is C^{r+1} in y. Continuing as before, we construct a new vector field that is as smooth as before, is as close as we wish to a linear vector field, leaves the unit ball invariant, and agrees with the original vector field (up to a C^r change of variable) near 0. The flow of this new vector field is then as close as we wish to the linear flow, and agrees with the original flow (up to a C^r change of variable) near 0.

Use X now to denote the new vector field, and ψ^t its flow. X is still triangular, $X(\xi, x, y) = [X_0(\xi), X_1(\xi, x), X_2(\xi, x, y)]$, and so

$$\psi^t(\xi, x, y) = [\varphi^t(\xi), \psi_1^t(\xi, x), \psi_2^t(\xi, x, y)].$$

We see that $[\varphi^t(\xi), \psi_1^t(\xi, x)]$ is the flow of $[X_0(\xi), X_1(\xi, x)]$, and is therefore C^{r+1}. Moreover, $(D_y \psi_2^t)$ solves the differential equation $A' = [(D_y X_2) \circ \psi_2^t]A$, which is C^r. Therefore, $(D_y \psi_2^t)$ is C^r, and ψ_2^t is C^{r+1} in y. Let f be the time-1 map of ψ^t. Then f agrees with the original time-1 map (up to a C^r change of variable) near 0, has been modified to fit the requirements of the proof, and is still the time-1 map of a flow. In particular, both ψ^t and f leave E^c and E^{cw} invariant.

Now repeat the proof for diffeomorphisms to get a diffeomorphism G conjugating f to $[\varphi(\xi), f_1(\xi, x), B_\xi y]$. Define $B_\xi^t := (D_y \psi_2^t)_{(\xi, 0, 0)}$ and $\Psi^t(\xi, x, y) := [\varphi^t(\xi), \psi_1^t(\xi, x), B_\xi^t y]$. B_ξ^t is C^r, and $B_\xi^1 = (D_y \psi_2^1)_{(\xi, 0, 0)} = B_\xi$. Hence

$$\Psi^1(\xi, x, y) = [\varphi(\xi), f_1(\xi, x), B_\xi y]$$

is the function to which f is conjugate. Therefore, $\Psi^1 \circ G = G \circ f = G \circ \psi^1$. We want to show that $\Psi^t \circ G = G \circ \psi^t$ for all t. For any fixed t, define $G^t := (\Psi^t)^{-1} \circ G \circ \psi^t$. If $G^t = G$ for each t, then we will have shown what we want.

Since $\psi^{t+s} = \psi^t \circ \psi^s$, we have $B_\xi^{t+s} := (D_y \psi_2^{t+s})_{(\xi, 0, 0)} = B_{\varphi^s(\xi)}^t B_\xi^s$. Consequently, $\Psi^{t+s} = \Psi^t \circ \Psi^s$, and Ψ^t is a flow. Therefore, $(\Psi^t)^{-1} = \Psi^{-t}$. Moreover, $G^t \circ f = (\Psi^{-t} \circ G \circ \psi^t) \circ \psi^1 = \Psi^{-t} \circ (G \circ \psi^1) \circ \psi^t = \Psi^{-t} \circ (\Psi^1 \circ G) \circ \psi^t =$

$\Psi^1 \circ (\Psi^{-t} \circ G \circ \psi^t) = \Psi^1 \circ G^t$. Hence G^t also conjugates f to Ψ^1, and $G^t(\xi, x, y) = (\Psi^{-t} \circ G \circ \psi^t)(\xi, x, y) = [\xi, x, B_{\varphi^t(\xi)}^{-t} G_2 \circ \psi^t(\xi, x, y)]$. If we define g_t by $g_t(\xi, x, y) = B_{\varphi^t(\xi)}^{-t} G_2 \circ \psi^t(\xi, x, y) - y$, then g_t is C^s and $G^t = \mathrm{id} + [0, 0, g_t]$.

All three of G, f, and Ψ^1 leave E^{cw} invariant. Therefore, G^t does also, and so $g_t(\xi, x, 0) \equiv 0$. We have $(D_y g_t)_p = B_{\varphi^t(\xi)}^{-t}(D_y G_2)_{\psi^t(p)}(D_y \psi_2^t)_p - I$. Therefore, $(D_y g_t)$ is C^s, and g_t is C^{s+1} in y. Moreover,

$$(D_y g_t)_{(\xi, 0, 0)} = B_{\varphi^t(\xi)}^{-t}(D_y G_2)_{(\varphi^t(\xi), 0, 0)}(D_y \psi_2^t)_{(\xi, 0, 0)} - I.$$

Since $(D_y G_2)_{(\xi, 0, 0)} = I$ for all ξ and since $(D_y \psi_2^t)_{(\xi, 0, 0)} = B_\xi^t$, we have

$$(D_y g_t)_{(\xi, 0, 0)} = B_{\varphi^t(\xi)}^{-t} B_\xi^t - I = B_\xi^0 - I = 0.$$

Consequently, g_t is in $\mathcal{G}$. Since G^t conjugates f to Ψ^1, g_t is a fixed point of Φ. Hence, if we can show that $(D_y g_t)$ is in $\mathcal{N}$, it will follow that $(D_y g_t)$ is a fixed point of Θ. However, we showed that Θ has a unique fixed point in $\mathcal{N}$. Consequently, we will be able to conclude that $(D_y g_t) = (D_y g)$, and hence $g_t = g$ and $G^t = G$. By definition $(D_y g_t)$ is in $\mathcal{N}$ provided $\|\|(D_y g_t)\|\| < \infty$. However, we can avoid calculating $\|\|(D_y g_t)\|\|$ and still derive our desired conclusion.

Because $s = k + \theta$ and $\theta > 0$, we can choose $\hat{\theta}$, $\delta > 0$ such that $s = k + \hat{\theta} + \delta$. Let $\hat{r} := s$ and $\hat{s} := k + \hat{\theta}$. Since $s < r$, the hypotheses of the proposition remain valid with $\hat{r}$ replacing r. Consequently, there is a unique $\hat{N} \in \mathcal{N}(\hat{s}, \delta)$ such that $\hat{N} = (D_y \hat{g})$ for some $\hat{g} \in \mathcal{G}(\hat{s})$, and $\hat{G} = \mathrm{id} + [0, 0, \hat{g}]$ conjugates f to Ψ^1. However, $\mathcal{G}(\hat{s}) \supset \mathcal{G}(s)$, and so both g and g_t are in $\mathcal{G}(\hat{s})$, from which it follows that $(D_y g)$ and $(D_y g_t)$ are in $\mathcal{D}(\hat{s})$. But by Lemma 4, since both $(D_y g_t)$ and $(D_y g)$ are C^s ($s = \hat{s} + \delta$), both are then in $\mathcal{N}(\hat{s}, \delta)$. Consequently, $(D_y g_t) = (D_y g) = \hat{N}$. Therefore, $G^t = G$, and G conjugates to the full flow, which completes the proof. □

REMARK. The fixed point g of Φ has $(D_y g) \in \mathcal{N}$. In particular, $\|\|(D_y g)\|\| < \infty$. From the definition of $\|\| \cdot \|\|$, this implies that all the mixed partials in $(D^k D_y g)$, and hence in $(D^k g)$, are $C^{\theta + \delta}$ except $(D_\xi^k D_y g)$, which is $C^{\theta + \delta}$ in $\nu = (x, y)$ but only C^θ in ξ. Consequently, although g is as a whole only $C^{r - \delta}$, it is actually C^r as a function of (x, y). This fact persists through each of the applications of Proposition 2. Therefore, the final partial linearization is $C^{r - \delta}$, but it is C^r as a function of the noncenter variables.

We are left with proving Lemma 5. For this we will need the higher-order chain rule and associated Leibnitz' rule.

THEOREM 3 [Pu] (Higher order chain rule). *Let $\mathbb{E}$, $\mathbb{F}$, $\mathbb{G}$ be Banach spaces, let $U \subset \mathbb{E}$ and $V \subset \mathbb{F}$ be open sets, and let $f: U \to V$ and $g: V \to \mathbb{G}$ be maps of class C^k. Then $g \circ f$ is of class C^k and, for $x \in U$ and $v_1, \ldots, v_k \in \mathbb{E}$, we have*

$$D^k(g \circ f)_x(v_1, \ldots, v_k) = \sum_{s=1}^{k} \sum_{\beta \in \mathcal{B}_{s,k}} (D^s g)_{f(x)} (D^\beta f)_x(v_\beta),$$

where $\mathcal{B}_{s,k}$ is the set of all partitions of $\{1, \ldots, k\}$ into s disjoint, nonempty subsets and where, for $\beta \in \mathcal{B}_{s,k}$ ($\beta = \{\beta_1, \ldots, \beta_s\}$ with $\beta_1 \sqcup \cdots \sqcup \beta_s = \{1, \ldots, k\}$), we

define the following: $|\beta_i| = order\ of\ \beta_i$; $v_{\beta_i} := \{v_j : j \in \beta_i\}$ *(so* $v_{\beta_1} \sqcup \cdots \sqcup v_{\beta_s} = \{v_1, \ldots, v_k\}$); *and*

$$(D^\beta f)_x(v_\beta) := [(D^{|\beta_1|}f)_x(v_{\beta_1}), \ldots, (D^{|\beta_s|}f)_x(v_{\beta_s})].$$

COROLLARY 2 (Leibnitz' rule). *Let* $\mathbb{E}, \mathbb{F}_1, \mathbb{F}_2, \mathbb{F}_3$ *be Banach spaces, let* $U \subset \mathbb{E}$ *be an open set, and let* $A\colon U \to L(\mathbb{F}_1, \mathbb{F}_2)$ *and* $B\colon U \to L(\mathbb{F}_2, \mathbb{F}_3)$ *be maps of class* C^k. *The function* $AB\colon U \to L(\mathbb{F}_1, \mathbb{F}_3)$ *defined by* $(AB)_x = A_x B_x$ *is of class* C^k *and*

$$D^k(AB)_x(v_1, \ldots, v_k) = A_x[(D^k B)_x(v_1, \ldots, v_k)] + [(D^k A)_x(v_1, \ldots, v_k)]B_x$$
$$+ \sum_{\beta \in \mathcal{B}_{2,k}} \big([(D^{|\beta_1|}A)_x(v_{\beta_1})][(D^{|\beta_2|}B)_x(v_{\beta_2})]$$
$$+ [(D^{|\beta_2|}A)_x(v_{\beta_2})][(D^{|\beta_1|}B)_x(v_{\beta_1})]\big).$$

Moreover,

$$\|D^k(AB)_x\| \le \sum_{s=0}^{k} \binom{k}{s} \|(D^s A)_x\| \cdot \|(D^{k-s}B)_x\|.$$

Proof of Lemma 5. We begin by stating the problem more abstractly; this is primarily a notational convenience. We have $f\colon E^c \oplus E^{ws} \to E^c \oplus E^{ws}$ and $f(\xi, v) = [\varphi(\xi), \mu(\xi, v)]$, where $\|(D\varphi)_\xi\| \le 1 + \varepsilon$, $\|(D\mu)_{(\xi,v)}\| \le \alpha + \varepsilon =: a < 1$, and $\mu(\xi, 0) \equiv 0$.

Define $R_\xi := B_\xi^{-1}$ and $S_{(\xi,v)} := (D_y f)_{(\xi,v)}$. Both R and S are C^r functions into $L(E^s, E^s)$, $\|R_\xi\| \cdot \|S_{(\xi,v)}\| \le (\beta + \varepsilon)(\beta^{-1} + \varepsilon) =: b$, and a, b, ε satisfy $b(1 + \varepsilon)^r a^\delta < \kappa < 1$. With these we have $\chi(N) = R(N \circ f)S$ for $N \in \mathcal{N}$.

Define $\||N|\|_i$ $(i = 0, \ldots, k)$ to be the maximum of the leading $2i + 1$ terms of $\||N|\|$:

$$\||N|\|_i := \max\{\|N\|_\delta, \|(D_v N)\|_{C^0}, \|(D_\xi N)\|_\delta,$$
$$\ldots, \|(D_v D^{i-1}N)\|_{C^0}, \|(D_\xi^i N)\|_\delta\}.$$

Because $(D_\xi^i N)_{(\xi,0)} = 0$, if $i < k$ then

$$\|(D_\xi^i N)_{(\xi,v)}\| \le \left[\sup_t \|(D_v D_\xi^i N)_{(\xi,tv)}\|\right]\|v\|.$$

Therefore, $\|(D_\xi^i N)\|_\delta \le \|(D_v D_\xi^i N)\|_{C^0} \le \|(D_v D^i N)\|_{C^0}$. Consequently, taking just the first $2i$ terms:

$$\max\{\|N\|_\delta, \|(D_v N)\|_{C^0}, \|(D_\xi N)\|_\delta, \ldots, \|D_\xi^{i-1}N\|_\delta, \|(D_v D^{i-1}N)\|_{C^0}\}$$
$$= \max\{\|(D_v N)\|_{C^0}, \|(D_v DN)\|_{C^0}, \ldots, \|(D_v D^{i-1}N)\|_{C^0}\}$$
$$= \|(D_v N)\|_{C^{i-1}}.$$

Therefore, $\||N|\|_{i-1} \le \|(D_v N)\|_{C^{i-1}} \le \||N|\|_i$ for all $i = 1, \ldots, k$.

In order to calculate $\|\chi\|$ we need to express the derivatives of $\chi(N)$ in terms of the derivatives of N. We do this in steps.

LEMMA 6. *If $A, B\colon E^c \oplus E^{ws} \to L(E^s, E^s)$, where $B \in C^r$, $A \in C^s$, and $\||A\|| < \infty$, then both $\||AB\||, \||BA\|| < \infty$, and in each of the following cases there exists a $K > 0$, independent of A and B, such that the given inequality holds:*

(i) $\|AB\|_\delta \leq \|A\|_\delta \|B\|_{C^0}$;

(ii) $\|D_\nu D^{i-1}(AB)\|_{C^0} \leq \|(D_\nu D^{i-1}A)\|_{C^0}\|B\|_{C^0} + K\|A\|_{C^{i-1}}\|B\|_{C^i}$
     $(i = 1, \ldots, k)$;

(iii) $\|D_\xi^i(AB)\|_\delta \leq \|(D_\xi^i A)\|_\delta \|B\|_{C^0} + K\||A\||_{i-1}\|B\|_{C^i}$ $(i = 1, \ldots, k)$;

(iv) $H_{\theta+\delta}[D_\nu D^{k-1}(AB)] \leq H_{\theta+\delta}[(D_\nu D^{k-1}A)]\|B\|_{C^0} + K\|A\|_{C^k}\|B\|_{C^r}$;

(v) $H_{\theta+\delta}^\nu[D_\xi^k(AB)] \leq H_{\theta+\delta}^\nu[(D_\xi^k A)]\|B\|_{C^0} + K\|A\|_{C^k}\|B\|_{C^r}$;

(vi) $Q_{\theta,\delta}[D_\xi^k(AB)] \leq Q_{\theta,\delta}[(D_\xi^k A)]\|B\|_{C^0} + K\||A\||_k\|B\|_{C^r}$.

Moreover, the same inequalities hold if AB is replaced by BA on the left-hand side.

Proof. A partial derivative is obtained by restricting the full derivative to the appropriate subspace. Hence, $D_\nu D^{i-1} = D^i|_{E^{ws} \times (\mathbb{R}^n)^{i-1}}$ and $D_\xi^i = D^i|_{(E^c)^i}$. Moreover, the norm of a partial derivative is no more than the norm of the full derivative. Therefore, from Leibnitz' rule:

$$\|D_\nu D^{i-1}(AB)_p\| \leq \|(D_\nu D^{i-1}A)_p\|\|B_p\| + \sum_{t=0}^{i-1}\binom{i}{t}\|(D^tA)_p\|\|(D^{i-t}B)_p\|.$$

Since $t \leq i - 1$ in the sum, $\|(D^tA)_p\| \leq \|A\|_{C^{i-1}}$ and $\|(D^{i-t}B)_p\| \leq \|B\|_{C^i}$ for all t, from which (ii) follows. Similarly, $\|D_\xi^i(AB)_p\|$ decomposes into a sum of terms $\|(D_\xi^t A)_p\|\|(D_\xi^{i-t}B)_p\|$. For $p = (\xi, \nu)$, $\|(D_\xi^t A)_{(\xi,\nu)}\| \leq \|(D_\xi^t A)\|_\delta \|\nu\|^\delta$. For each $t \leq i - 1$, $\|(D_\xi^t A)\|_\delta \leq \||A\||_{i-1}$ and $\|(D_\xi^{i-t}B)_p\| \leq \|B\|_{C^i}$. Therefore, $\|D_\xi^i(AB)_{(\xi,\nu)}\|/\|\nu\|^\delta \leq [\|(D_\xi^i A)_{(\xi,\nu)}\|/\|\nu\|^\delta]\|B_p\| + K\||A\||_{i-1}\|B\|_{C^i}$ for some $K > 0$, from which (iii) follows. (i) is just a special case of (iii).

To obtain (iv)–(vi), expand $D^k(AB)_{p_1} - D^k(AB)_{p_2}$ into a sum of terms of the form $(D^tA)_{p_1}(D^{k-t}B)_{p_1} - (D^tA)_{p_2}(D^{k-t}B)_{p_2}$. (There are, in general, several terms for each t, each corresponding to a different permutation of the input vectors.) Expanding further:

$$(D^tA)_{p_1}(D^{k-t}B)_{p_1} - (D^tA)_{p_2}(D^{k-t}B)_{p_2}$$
$$= [(D^tA)_{p_1} - (D^tA)_{p_2}](D^{k-t}B)_{p_1} + (D^tA)_{p_2}[(D^{k-t}B)_{p_1} - (D^{k-t}B)_{p_2}].$$

Consequently,

$$\|D_\nu D^{k-1}(AB)_{p_1} - D_\nu D^{k-1}(AB)_{p_2}\|$$
$$\leq \|(D_\nu D^{k-1}A)_{p_1} - (D_\nu D^{k-1}A)_{p_2}\|\|B_{p_1}\| + \|(D_\nu D^{k-1}A)_{p_2}\|\|B_{p_1} - B_{p_2}\|$$
$$+ \sum_{t=0}^{k-1}\binom{k}{t}\Big(\|(D^tA)_{p_1} - (D^tA)_{p_2}\|\|(D^{k-t}B)_{p_1}\|$$
$$+ \|(D^tA)_{p_2}\|\|(D^{k-t}B)_{p_1} - (D^{k-t}B)_{p_2}\|\Big).$$

For $t < k$, $(D^tA)_{p_2}$ is at least C^1. Therefore,

$$\|(D^t A)_{p_1} - (D^t A)_{p_2}\| \leq \|(D^{t+1}A)\|_{C^0}\|p_1 - p_2\|.$$

Moreover, $\|(D^{t+1}A)\|_{C^0} \leq \|A\|_{C^k}$. Likewise,

$$\|(D^{k-t}B)_{p_1} - (D^{k-t}B)_{p_2}\| \leq \|B\|_{C^r}\|p_1 - p_2\|^{\theta+\delta},$$

and we also have

$$\|(D_v D^{k-1}A)_{p_2}\|\|B_{p_1} - B_{p_2}\| \leq \|A\|_{C^k}\|B\|_{C^r}\|p_1 - p_2\|^{\theta+\delta}.$$

Because $\|p_1 - p_2\| \leq 2$ and $\theta + \delta \leq 1$, (iv) follows. Since this argument did not use anything specific to $D_v D^{k-1}$ or the form of the points p_i, if we instead use D_ξ^k, and consider points only of the form $p_i = (\xi, v_i)$, in which case $\|p_1 - p_2\| = \|v_1 - v_2\|$, we obtain (v).

For the last case, expand $D_\xi^k(AB)$ as in the previous two cases but now *not* replacing the lower-order partials by their full derivatives. This yields

$$\|D_\xi^k(AB)_{(\xi_1,v)} - D_\xi^k(AB)_{(\xi_2,v)}\|$$

$$\leq \|(D_\xi^k A)_{(\xi_1,v)} - (D_\xi^k A)_{(\xi_2,v)}\|\|B_{(\xi_1,v)}\| + \|(D_\xi^k A)_{(\xi_2,v)}\|\|B_{(\xi_1,v)} - B_{(\xi_2,v)}\|$$

$$+ \sum_{t=0}^{k-1}\binom{k}{t}\Big(\|(D_\xi^t A)_{(\xi_1,v)} - (D_\xi^t A)_{(\xi_2,v)}\|\|(D_\xi^{k-t}B)_{(\xi_1,v)}\|$$

$$+ \|(D_\xi^t A)_{(\xi_2,v)}\|\|(D_\xi^{k-t}B)_{(\xi_1,v)} - (D_\xi^{k-t}B)_{(\xi_2,v)}\|\Big).$$

For $t < k$, $(D_\xi^t A)$ is C^1, and so we have

$$\|(D_\xi^t A)_{(\xi_1,v)} - (D_\xi^t A)_{(\xi_2,v)}\| \leq \Big(\sup_\xi\|(D_\xi^{t+1}A)_{(\xi,v)}\|\Big)\|\xi_1 - \xi_2\|$$

and

$$\|(D_\xi^{t+1}A)_{(\xi,v)}\| \leq \|(D_\xi^{t+1}A)\|_\delta\|v\|^\delta \leq \|A\|_k\|v\|^\delta.$$

We also have $\|(D_\xi^{k-t}B)_{(\xi,v)}\| \leq \|B\|_{C^r}$. Therefore,

$$\|(D_\xi^t A)_{(\xi_1,v)} - (D_\xi^t A)_{(\xi_2,v)}\|\|(D_\xi^{k-t}B)_{(\xi_1,v)}\|$$

$$\leq \Big(\|A\|_k\|B\|_{C^r}\|\xi_1 - \xi_2\|^{1-\theta}\Big)\|\xi_1 - \xi_2\|^\theta\|v\|^\delta.$$

Since $\xi_i \in E^c(1)$, $\|\xi_1 - \xi_2\|^{1-\theta} \leq 2$. Similarly, for $t \leq k$,

$$\|(D_\xi^t A)_{(\xi_2,v)}\|\|(D_\xi^{k-t}B)_{(\xi_1,v)} - (D_\xi^{k-t}B)_{(\xi_2,v)}\|$$

$$\leq \Big(\|A\|_k\|v\|^\delta\Big)\Big(\|B\|_{C^r}\|\xi_1 - \xi_2\|\Big)$$

$$\leq \Big(2\|A\|_k\|B\|_{C^r}\Big)\|\xi_1 - \xi_2\|^\theta\|v\|^\delta.$$

Because

$$Q_{\theta,\delta}[D_\xi^k(AB)] := \sup_{\xi_1\neq\xi_2,\,v\neq 0}\Big(\|D_\xi^k(AB)_{(\xi_1,v)} - D_\xi^k(AB)_{(\xi_2,v)}\|/[\|\xi_1 - \xi_2\|^\theta\|v\|^\delta]\Big),$$

(vi) follows. Finally, the previous estimates did not depend on the order of A and B. Therefore, they can be repeated on BA to obtain the exact same results. □

LEMMA 7.   *In each of the following cases there exist constants $K, K', K'' > 0$, depending on f, such that for any $N \in \mathcal{N}$ the given inequality holds*:

(i) $\|N \circ f\|_\delta \le a^\delta \|N\|_\delta$;

(ii) $\|D_\nu D^{i-1}(N \circ f)\|_{C^0} \le a(1 + \varepsilon)^{i-1} \|(D_\nu D^{i-1}N)\|_{C^0} + K\|N\|_{C^{i-1}}$
     $(i = 1, \dots, k)$;

(iii) $\|D_\xi^i(N \circ f)\|_\delta \le a^\delta(1 + \varepsilon)^i \|(D_\xi^i N)\|_\delta + K\|(D_\nu N)\|_{C^{i-1}}$ $(i = 1, \dots, k)$;

(iv) $H_{\theta+\delta}[D_\nu D^{k-1}(N \circ f)] \le a(1 + \varepsilon)^r H_{\theta+\delta}[(D_\nu D^{k-1}N)] + K\|N\|_{C^k}$;

(v) $H_{\theta+\delta}^\nu[D_\xi^k(N \circ f)] \le a^{\theta+\delta}(1 + \varepsilon)^k H_{\theta+\delta}^\nu[(D_\xi^k N)] + K' H_{\theta+\delta}[(D_\nu D^{k-1}N)] + K\|N\|_{C^k}$;

(vi) $Q_{\theta,\delta}[D_\xi^k(N \circ f)] \le a^\delta(1 + \varepsilon)^r Q_{\theta,\delta}[(D_\xi^k N)] + K'' H_{\theta+\delta}^\nu[(D_\xi^k N)] + K' H_{\theta+\delta}[(D_\nu D^{k-1}N) + K\|\|N\|\|_k$.

Proof. As in the previous lemma, we will use that (a) a partial derivative is the restriction of the full derivative to a subspace and (b) the norm of the full derivative is always at least as large as that of the partial derivative. Applying these facts to $D_\nu D^{i-1}(N \circ f)$, we have

$$\|D_\nu D^{i-1}(N \circ f)_p\| = \|D^i(N \circ f)_p|_{E^{ws} \times (\mathbb{R}^n)^{i-1}}\|$$

$$\le \|(D^iN)_{f(p)}[(Df)_p^{(i)}]|_{E^{ws} \times (\mathbb{R}^n)^{i-1}}\|$$

$$+ \sum_{t=1}^{i-1} \sum_{\beta \in \mathcal{B}_{t,i}} \|(D^tN)_{f(p)}(D^\beta f)_p\|.$$

Here $(Df)_p^{(i)}$ means i copies of $(Df)_p$. We have

$$(Df)_p|_{E^{ws}} = (D_\nu f)_p = [\,0\quad (D_\nu \mu)_p\,]^{\mathrm{T}},$$

and so

$$(D^iN)_{f(p)}[(Df)_p^{(i)}]|_{E^{ws} \times (\mathbb{R}^n)^{i-1}} = (D_\nu D^{i-1}N)_{f(p)}[(D_\nu \mu)_p, (Df)_p^{(i-1)}].$$

Because $\|(D_\nu \mu)_p\| \le a$ and $\|(Df)_p\| \le (1 + \varepsilon)$,

$$\|(D^iN)_{f(p)}[(Df)_p^{(i)}]|_{E^{ws} \times (\mathbb{R}^n)^{i-1}}\| \le a(1 + \varepsilon)^{i-1} \|D_\nu D^{i-1}N\|_{C^0}.$$

For the terms inside the sum,

$$\|(D^tN)_{f(p)}(D^\beta f)_p\| \le \|(D^tN)_{f(p)}\| \|(D^{|\beta_1|}f)_p\| \cdots \|(D^{|\beta_t|}f)_p\|.$$

There are $t < i$ derivatives $\|(D^{|\beta_j|}f)_p\| \le \|f\|_{C^i}$, and $\|(D^tN)_{f(p)}\| \le \|N\|_{C^{i-1}}$. Therefore, $\|(D^tN)_{f(p)}(D^\beta f)_p\| \le (\|f\|_{C^i})^{i-1} \|N\|_{C^{i-1}}$. This holds for each term in the sum, yielding (ii).

For $\tau = D_\xi^i N$ (including $i = 0$), $\|(\tau \circ f)(\xi, \nu)\| = \|\tau(\varphi(\xi), \mu(\xi, \nu))\| \le \|\tau\|_\delta \|\mu(\xi, \nu)\|^\delta$. Since $\mu(\xi, 0) \equiv 0$, $\|\mu(\xi, \nu)\| \le [\sup_p \|(D_\nu \mu)_p\|] \|\nu\| \le a\|\nu\|$. Therefore, $\|\tau \circ f\|_\delta \le a^\delta \|\tau\|_\delta$. In particular, we have case (i). For case (iii) $(i > 0)$,

$$\|D_\xi^i(N \circ f)_p\| = \|D^i(N \circ f)_p|_{(E^c)^i}\|$$

$$\le \|(D^iN)_{f(p)}[(D_\xi f)_p^{(i)}]\| + \sum_{t=1}^{i-1} \sum_{\beta \in \mathcal{B}_{t,i}} \|(D^tN)_{f(p)}(D_\xi^\beta f)_p\|.$$

Here we use $(D_\xi^l f)_p = [\,(D_\xi^l \varphi)_p \;\; (D_\xi^l \mu)_p\,]^\mathrm{T}$, and therefore

$$(D^t N)_{f(p)}[(D_\xi^l f)_p, \cdot\,]$$
$$= (D_\xi D^{t-1} N)_{f(p)}[(D_\xi^l \varphi)_p, \cdot\,] + (D_\nu D^{t-1} N)_{f(p)}[(D_\xi^l \mu)_p, \cdot\,].$$

We can expand $(D^i N)_{f(p)}[(D_\xi f)_p^{(i)}]$ completely in this manner to get

$$(D^i N)_{f(p)}[(D_\xi f)_p^{(i)}]$$
$$= (D_\xi^i N)_{f(p)}[(D_\xi \varphi)_p^{(i)}]$$
$$+ \sum_{j=1}^{i} (D_\xi^{j-1} D_\nu D^{i-j} N)_{f(p)}[(D_\xi \varphi)_p^{(j-1)}, (D_\xi \mu)_p, (D_\xi f)_p^{(i-j)}].$$

Note that

$$\|(D_\xi^{j-1} D_\nu D^{i-j} N)_{f(p)}\| \leq \|(D_\nu D^{i-1} N)_{f(p)}\| \leq \|(D_\nu N)\|_{C^{i-1}}$$

and

$$\|(D_\xi \varphi)_p\| \leq \|(D_\xi f)_p\| \leq (1 + \varepsilon).$$

Set $p = (\xi, \nu)$. Since $\mu(\xi, 0) \equiv 0$, $(D_\xi \mu)_{(\xi, 0)} \equiv 0$. Moreover, $r \geq 1 + \delta$ and so $(D_\xi \mu)$ is at least C^δ. Consequently $\|(D_\xi \mu)_{(\xi, \nu)}\| \leq H_\delta[(D_\xi \mu)]\|\nu\|^\delta \leq \|f\|_{C^r}\|\nu\|^\delta$. From the foregoing we have $\|(D_\xi^i N)_{f(\xi, \nu)}\| \leq \|(D_\xi^i N) \circ f\|_\delta \|\nu\|^\delta \leq a^\delta \|(D_\xi^i N)\|_\delta \|\nu\|^\delta$. Therefore,

$$\|(D^i N)_f[(D_\xi f)^{(i)}]\|_\delta \leq a^\delta (1 + \varepsilon)^i \|(D_\xi^i N)\|_\delta + i(1 + \varepsilon)^{i-1}\|f\|_{C^r}\|(D_\nu N)\|_{C^{i-1}}.$$

Each of the terms $(D^t N)_{f(p)}(D_\xi^\beta f)_p$ can be expanded in a similar manner. In this case we use $\|(D_\xi^l \varphi)_p\| \leq \|(D_\xi^l f)_p\| \leq \|f\|_{C^r}$ and $(D_\xi^l \mu)_{(\xi, 0)} \equiv 0$, so $\|(D_\xi^l \mu)_{(\xi, \nu)}\| \leq \|f\|_{C^r}\|\nu\|^\delta$. Therefore,

$$\|(D^t N)_f(D_\xi^\beta f)\|_\delta \leq a^\delta (\|f\|_{C^r})^t \|(D_\xi^t N)\|_\delta + t(\|f\|_{C^r})^t \|(D_\nu N)\|_{C^{t-1}}.$$

Because $t < i$, $\|(D_\xi^t N)\|_\delta \leq \|\|N\|\|_{i-1} \leq \|(D_\nu N)\|_{C^{i-1}}$. Hence

$$\|(D^t N)_f(D_\xi^\beta f)\|_\delta \leq K'\|(D_\nu N)\|_{C^{i-1}},$$

which—when combined with the estimate for the first term—yields (iii).

For cases (iv)–(vi) we need to work with $D^k(N \circ f)_{p_1} - D^k(N \circ f)_{p_2}$. This reduces to a sum of terms $(D^t N)_{f(p_1)}(D^\beta f)_{p_1} - (D^t N)_{f(p_2)}(D^\beta f)_{p_2}$. Let $\Delta[\tau]$ be short for $\tau(p_1) - \tau(p_2)$; then, using the multilinearity of $(D^t N)$, this can be rewritten as

$$\Delta[(D^t N) \circ f](D^\beta f)_{p_1}$$
$$+ \sum_{j=1}^{t} (D^t N)_{f(p_2)}[(D^{|\beta_1|} f)_{p_2}, \ldots, \Delta(D^{|\beta_j|} f), \ldots, (D^{|\beta_t|} f)_{p_1}].$$

If $t < k$ then

$$\|\Delta[(D^t N) \circ f]\| \le \|(D^{t+1}N)\|_{C^0} \|f(p_1) - f(p_2)\| \le (1+\varepsilon)\|N\|_{C^k} \|p_1 - p_2\|.$$

If $t \le k$ then

$$\|(D^t N)_{f(p_2)}\| \le \|N\|_{C^k}, \qquad \|(D^{|\beta_j|}f)_{p_1}\| \le \|f\|_{C^r},$$

and

$$\|\Delta(D^{|\beta_j|}f)\| \le \|f\|_{C^r} \|p_1 - p_2\|^{\theta+\delta}.$$

Therefore, every term in the expansion of $D^k(N \circ f)_{p_1} - D^k(N \circ f)_{p_2}$ except $\Delta[(D^k N) \circ f][(Df)_{p_1}^{(k)}]$ can be bounded by a constant times $\|N\|_{C^k} \|p_1 - p_2\|^{\theta+\delta}$. Consequently, if we let V be either of the subspaces $E^{ws} \times (\mathbb{R}^n)^{k-1}$ or $(E^c)^k$, we have

$$\|[D^k(N \circ f)_{p_1} - D^k(N \circ f)_{p_2}]|_V\|$$

$$\le \|[(D^k N)_{f(p_1)} - (D^k N)_{f(p_2)}][(Df)_{p_1}^{(k)}]|_V\| + K\|N\|_{C^k} \|p_1 - p_2\|^{\theta+\delta}.$$

If $V = E^{ws} \times (\mathbb{R}^n)^{k-1}$, then $D^k|_V = D_v D^{k-1}$ and

$$[(D^k N)_{f(p_1)} - (D^k N)_{f(p_2)}][(Df)_{p_1}^{(k)}]|_V$$

$$= [(D^k N)_{f(p_1)} - (D^k N)_{f(p_2)}][(D_v f)_{p_1}, (Df)_{p_1}^{(k-1)}]$$

$$= [(D_v D^{k-1} N)_{f(p_1)} - (D_v D^{k-1} N)_{f(p_2)}][(D_v \mu)_{p_1}, (Df)_{p_1}^{(k-1)}].$$

Since

$$\|(D_v D^{k-1} N)_{f(p_1)} - (D_v D^{k-1} N)_{f(p_2)}\|$$

$$\le H_{\theta+\delta}[(D_v D^{k-1} N)]\|f(p_1) - f(p_2)\|^{\theta+\delta},$$

$\|(D_v \mu)_{p_1}\| \le a$, $\|(Df)_{p_1}\| \le (1+\varepsilon)$, and $\|f(p_1) - f(p_2)\| \le (1+\varepsilon)\|p_1 - p_2\|$, we get (iv).

Now let $V = (E^c)^k$ and restrict to $p_i = (\xi, v_i)$. This case can be broken down like (iii):

$$[(D^k N)_{f(p_1)} - (D^k N)_{f(p_2)}][(Df)_{p_1}^{(k)}]|_V$$

$$= [(D^k N)_{f(p_1)} - (D^k N)_{f(p_2)}][(D_\xi f)_{p_1}^{(k)}]$$

$$= [(D_\xi^k N)_{f(p_1)} - (D_\xi^k N)_{f(p_2)}][(D_\xi \varphi)_{p_1}^{(k)}]$$

$$+ \sum_{j=1}^{i} \Big([(D_v D^{k-1} N)_{f(p_1)} - (D_v D^{k-1} N)_{f(p_1)}]$$

$$\times [(D_\xi \mu)_p, (D_\xi \varphi)_p^{(j-1)}, (D_\xi f)_p^{(i-j)}] \Big).$$

Inside the sum,

$$\|(D_v D^{k-1} N)_{f(p_1)} - (D_v D^{k-1} N)_{f(p_1)}\|$$

$$\le H_{\theta+\delta}[(D_v D^{k-1} N)]\|f(p_1) - f(p_2)\|^{\theta+\delta},$$

and each of the arguments is bounded by $\|(Df)\|_{C^0}$. Therefore, inside the summation is bounded by

$$K'H_{\theta+\delta}[(D_\nu D^{k-1}N)]\|p_1 - p_2\|^{\theta+\delta} = K'H_{\theta+\delta}[(D_\nu D^{k-1}N)]\|\nu_1 - \nu_2\|^{\theta+\delta}$$

for some $K' > 0$.

Because φ does not depend on ν,

$$\|f(\xi, \nu_1) - f(\xi, \nu_2)\| = \|\mu(\xi, \nu_1) - \mu(\xi, \nu_2)\| \le a\|\nu_1 - \nu_2\|.$$

Consequently, $\|(D_\xi^k N)_{f(\xi, \nu_1)} - (D_\xi^k N)_{f(\xi, \nu_2)}\| \le H_{\theta+\delta}^\nu[(D_\xi^k N)](a\|\nu_1 - \nu_2\|)^{\theta+\delta}$.
Combining this with the previous results and $\|(D_\xi \varphi)_{p_1}\| \le (1 + \varepsilon)$, we have (v).

Finally we are left with the task of showing (vi). For this case we need to return to the expansion of $D^k(N \circ f)_{p_1} - D^k(N \circ f)_{p_2}$, but must restrict to $V = (E^c)^k$ before making any estimates:

$$\|D_\xi^k(N \circ f)_{p_1} - D_\xi^k(N \circ f)_{p_2}\|$$

$$\le \sum_{t=1}^k \sum_{\beta \in \mathcal{B}_{t,k}} \|\Delta[(D^t N) \circ f](D_\xi^\beta f)_{p_1}\|$$

$$+ \sum_{t=1}^k \sum_{\beta \in \mathcal{B}_{t,k}} \sum_j \|(D^t N)_{f(p_2)}[(D_\xi^{|\beta_1|} f)_{p_2},$$

$$\dots, \Delta(D_\xi^{|\beta_j|} f), \dots, (D_\xi^{|\beta_t|} f)_{p_1}]\|.$$

These terms are analogous to those in case (iii). Expanding in the same manner, and using $p_i = (\xi_i, \nu)$ and $\|(D_\xi^l \mu)_{(\xi, \nu)}\| \le \|f\|_{C^r}\|\nu\|^\delta$, for the first sum we have

$$\|\Delta[(D^t N) \circ f](D_\xi^\beta f)_{p_1}\|$$

$$\le \|\Delta[(D_\xi^t N) \circ f]\|\|(D_\xi^{|\beta_1|}\varphi)_{p_1}\| \cdots \|(D_\xi^{|\beta_t|}\varphi)_{p_1}\|$$

$$+ \sum_j \left(\|\Delta[(D_\nu D^{t-1}N) \circ f]\| \right.$$

$$\left. \times \|(D_\xi^{|\beta_1|}\varphi)_{p_1}\| \cdots \|(D_\xi^{|\beta_j|}\mu)_{p_1}\| \cdots \|(D_\xi^{|\beta_t|} f)_{p_1}\| \right)$$

$$\le \|\Delta[(D_\xi^t N) \circ f]\|\|(D_\xi^{|\beta_1|}\varphi)_{p_1}\| \cdots \|(D_\xi^{|\beta_t|}\varphi)_{p_1}\|$$

$$+ t \cdot H_{\theta+\delta}[(D_\nu D^{t-1}N)](a\|\xi_1 - \xi_2\|)^{\theta+\delta}(\|f\|_{C^r})^t\|\nu\|^\delta.$$

Dividing $\|\Delta[(D_\xi^t N) \circ f]\| := \|(D_\xi^t N)_{[\varphi(\xi_1), \mu(\xi_1, \nu)]} - (D_\xi^t N)_{[\varphi(\xi_2), \mu(\xi_2, \nu)]}\|$ into parts:

$$\|(D_\xi^t N)_{[\varphi(\xi_1), \mu(\xi_1, \nu)]} - (D_\xi^t N)_{[\varphi(\xi_2), \mu(\xi_2, \nu)]}\|$$

$$\le \|(D_\xi^t N)_{[\varphi(\xi_1), \mu(\xi_1, \nu)]} - (D_\xi^t N)_{[\varphi(\xi_2), \mu(\xi_1, \nu)]}\|$$

$$+ \|(D_\xi^t N)_{[\varphi(\xi_2), \mu(\xi_1, \nu)]} - (D_\xi^t N)_{[\varphi(\xi_2), \mu(\xi_2, \nu)]}\|$$

$$\le Q_{\theta,\delta}[(D_\xi^t N)]\|\varphi(\xi_1) - \varphi(\xi_2)\|^\theta \|\mu(\xi_1, \nu)\|^\delta$$

$$+ H_{\theta+\delta}^\nu[(D_\xi^t N)]\|\mu(\xi_1, \nu) - \mu(\xi_2, \nu)\|^{\theta+\delta}.$$

We have $\|\varphi(\xi_1) - \varphi(\xi_2)\| \le (1 + \varepsilon)\|\xi_1 - \xi_2\|$ and $\|\mu(\xi_1, \nu)\| \le \|(D_\nu\mu)\|_{C^0}\|\nu\| \le a\|\nu\|$. On the one hand, $\|\mu(\xi_1, \nu) - \mu(\xi_2, \nu)\| \le \|(D_\xi\mu)\|_{C^0}\|\xi_1 - \xi_2\|$, but also $\|\mu(\xi_1, \nu) - \mu(\xi_2, \nu)\| \le 2\|(D_\nu\mu)\|_{C^0}\|\nu\|$. Therefore, $\|\mu(\xi_1, \nu) - \mu(\xi_2, \nu)\|^{\theta+\delta} \le K' \min\{\|\xi_1 - \xi_2\|^{\theta+\delta}, \|\nu\|^{\theta+\delta}\} \le K'\|\xi_1 - \xi_2\|^\theta\|\nu\|^\delta$. Hence $\|\Delta[(D_\xi^t N) \circ f]\| \le \left\{a^\delta(1 + \varepsilon)^\theta Q_{\theta,\delta}[(D_\xi^t N)] + K' H_{\theta+\delta}^\nu[(D_\xi^t N)]\right\}\|\xi_1 - \xi_2\|^\theta\|\nu\|^\delta$.

If $t < k$ then

$$\|(D_\xi^t N)_{(\xi_1, \nu)} - (D_\xi^t N)_{(\xi_2, \nu)}\| \le (\sup_\xi \|(D_\xi^{t+1} N)_{(\xi, \nu)}\|)\|\xi_1 - \xi_2\|.$$

Since $\|(D_\xi^{t+1} N)_{(\xi, \nu)}\| \le \|(D_\xi^{t+1} N)\|_\delta\|\nu\|^\delta \le \|\|N\|\|_k\|\nu\|^\delta$, it follows that

$$\|(D_\xi^t N)_{(\xi_1, \nu)} - (D_\xi^t N)_{(\xi_2, \nu)}\| \le \|\|N\|\|_k\|\xi_1 - \xi_2\|\|\nu\|^\delta.$$

Therefore, $Q_{\theta,\delta}[(D_\xi^t N)] \le \|\|N\|\|_k$. Moreover,

$$H_{\theta+\delta}^\nu[(D_\xi^t N)] \le \|(D_\nu D^t N)\|_{C^0} \le \|\|N\|\|_k,$$

and

$$H_{\theta+\delta}[(D_\nu D^{t-1} N)] \le \|(D_\nu D^t N)\|_{C^0} \le \|\|N\|\|_k.$$

Consequently, $\|\Delta[(D^t N) \circ f](D_\xi^\beta f)_{p_1}\| \le K\|\|N\|\|_k\|\xi_1 - \xi_2\|^\theta\|\nu\|^\delta$ for $t < k$ (some $K > 0$).

If $t = k$, then $(D_\xi^\beta f)_{p_1} = (D_\xi f)_{p_1}^{(k)}$ and $\|(D_\xi\varphi)_{p_1}\| \le (1 + \varepsilon)$. Consequently, we have

$$\|\Delta[(D^k N) \circ f][(D_\xi f)_{p_1}^{(k)}]\|/(\|\xi_1 - \xi_2\|^\theta\|\nu\|^\delta)$$

$$\le a^\delta(1 + \varepsilon)^r Q_{\theta,\delta}[(D_\xi^k N)] + K'' H_{\theta+\delta}^\nu[(D_\xi^k N)] + K' H_{\theta+\delta}[(D_\nu D^{k-1} N)].$$

The second sum can now be estimated in a similar fashion:

$$\|(D^t N)_{f(p_2)}[(D_\xi^{|\beta_1|} f)_{p_2}, \dots, \Delta(D_\xi^{|\beta_j|} f), \dots, (D_\xi^{|\beta_l|} f)_{p_1}]\|$$

$$\le \|(D_\xi^t N)_{f(p_2)}\|(\|f\|_{C^r})^{t-1}\|\Delta(D_\xi^{|\beta_j|}\varphi)\|$$

$$+ \sum_{l<j} \|(D_\nu D^{t-1} N)_{f(p_2)}\|(\|f\|_{C^r})^{t-2}\|(D_\xi^{|\beta_l|}\mu)_{p_2}\| \cdot \|\Delta(D_\xi^{|\beta_j|} f)\|$$

$$+ \|(D_\nu D^{t-1} N)_{f(p_2)}\|(\|f\|_{C^r})^{t-1}\|\Delta(D_\xi^{|\beta_j|}\mu)\|$$

$$+ \sum_{l>j} \|(D_\nu D^{t-1} N)_{f(p_2)}\|(\|f\|_{C^r})^{t-2}\|(D_\xi^{|\beta_l|}\mu)_{p_1}\| \cdot \|\Delta(D_\xi^{|\beta_j|}\varphi)\|.$$

To this we can apply the following estimates:

$$\|\Delta(D_\xi^{|\beta_j|}\varphi)\| \le \|\Delta(D_\xi^{|\beta_j|} f)\| \le \|f\|_{C^r}\|\xi_1 - \xi_2\|^{\theta+\delta};$$

$$\|(D_\xi^t N)_{f(p_2)}\| \le \|\|N\|\|_k\|\mu(\xi_2, \nu)\|^\delta \le a^\delta\|\|N\|\|_k\|\nu\|^\delta;$$

$$\|(D_\nu D^{t-1} N)_{f(p_2)}\| \le \|\|N\|\|_k;$$

$$\|(D_\xi^{|\beta_l|}\mu)_p\| \le \|f\|_{C^r}\|\nu\|^\delta.$$

This leaves only $\|\Delta(D_\xi^{|\beta_j|}\mu)\|$ to estimate. Since $(D_\xi^l\mu)$ is $C^{\theta+\delta}$,

$$\|\Delta(D_\xi^l\mu)\| := \|(D_\xi^l\mu)_{(\xi_1,\nu)} - (D_\xi^l\mu)_{(\xi_2,\nu)}\| \le \|f\|_{C^r}\|\xi_1 - \xi_2\|^{\theta+\delta},$$

but also $\|\Delta(D_\xi^l\mu)\| \le \|(D_\xi^l\mu)_{(\xi_1,\nu)}\| + \|(D_\xi^l\mu)_{(\xi_2,\nu)}\| \le 2\|f\|_{C^r}\|\nu\|^{\theta+\delta}$. Consequently, we again get $\|\Delta(D_\xi^l\mu)\| \le K\|\xi_1 - \xi_2\|^{\theta}\|\nu\|^\delta$ for some $K > 0$. Therefore, for some (different) $K > 0$,

$$\|(D^t N)_{f(p_2)}[(D_\xi^{|\beta_1|}f)_{p_2}, \ldots, \Delta(D_\xi^{|\beta_j|}f), \ldots, (D_\xi^{|\beta_t|}f)_{p_1}]\|$$
$$\le K\|\|N\|\|_k\|\xi_1 - \xi_2\|^\theta\|\nu\|^\delta.$$

Combining the first and second sums yields (vi), completing the proof. □

With these two lemmas we can now complete our proof of Lemma 5. The definition of $\|\|\cdot\|\|$ is the maximum of a set of $2k + 4$ elements. The order in which these are listed in the definition is significant. We have already observed that the max of the first $2i$ ($i = 1, \ldots, k$) elements for $\|\|N\|\|$ is $\|(D_\nu N)\|_{C^{i-1}}$, and the max of the first $2i + 1$ ($i = 1, \ldots, k$) elements is $\|\|N\|\|_i$ (by definition). By the same argument as in the proof of Lemma 2, $\|N\|_{C^i} \le (i+1)\|\|N\|\|_i$. Therefore, the cases in Lemma 6 estimate the size of the jth element in the set determining $\|\|AB\|\|$ in terms of the first j elements for $\|\|A\|\|$. In particular, for each i, $\|\|AB\|\|_i \le K\|\|A\|\|_i$ for some K not depending on A (but depending on i and B). Therefore, if we apply Lemma 6 twice, and then Lemma 7, to $\chi(N) = R[(N \circ f)S]$, and use $\|R\|_{C^0}\|S\|_{C^0} \le b$, we have:

(i) $\|\chi(N)\|_\delta \le a^\delta b\|N\|_\delta$;

(ii) $\|D_\nu D^{i-1}(\chi(N))\|_{C^0} \le a(1+\varepsilon)^{i-1}b\|(D_\nu D^{i-1}N)\|_{C^0} + K\|\|N\|\|_{i-1}$
$(i = 1, \ldots, k)$;

(iii) $\|D_\xi^i(\chi(N))\|_\delta \le a^\delta(1+\varepsilon)^i b\|(D_\xi^i N)\|_\delta + K\|(D_\nu N)\|_{C^{i-1}}$ $(i = 1, \ldots, k)$;

(iv) $H_{\theta+\delta}[(D_\nu D^{k-1}(\chi(N)))] \le a(1+\varepsilon)^r b H_{\theta+\delta}[(D_\nu D^{k-1}N)] + K\|\|N\|\|_k$;

(v) $H_{\theta+\delta}^\nu[(D_\xi^k(\chi(N)))] \le a^{\theta+\delta}(1+\varepsilon)^k b H_{\theta+\delta}^\nu[(D_\xi^k N)] +$
$K'H_{\theta+\delta}[(D_\nu D^{k-1}N)] + K\|\|N\|\|_k$;

(vi) $Q_{\theta,\delta}[D_\xi^k(\chi(N))] \le a^\delta(1+\varepsilon)^r b Q_{\theta,\delta}[(D_\xi^k N)] + K''H_{\theta+\delta}^\nu[(D_\xi^k N)] +$
$K'H_{\theta+\delta}[(D_\nu D^{k-1}N) + K\|\|N\|\|_k$.

Since $b(1+\varepsilon)^r a^\delta < \kappa$, the leading term in each case has a coefficient less than $\kappa < 1$.

Now define the new norm $\|\|\cdot\|\|'$ by

$$\|\|N\|\|' = \max\{\|N\|_\delta, b_1\|(D_\nu N)\|_{C^0}, c_1\|(D_\xi N)\|_\delta,$$

$$\ldots, b_k\|(D_\nu D^{k-1}N)\|_{C^0}, c_k\|(D_\xi^k N)\|_\delta,$$

$$d_1 H_{\theta+\delta}[(D_\nu D^{k-1}N)], d_2 H_{\theta+\delta}^\nu[(D_\xi^k N)], d_3 Q_{\theta,\delta}[(D_\xi^k N)]\},$$

where $b_1 \ge c_1 \ge \cdots \ge b_k \ge c_k \ge d_1 \ge d_2 \ge d_3 > 0$ are some collection of positive constants. $\|\|\cdot\|\|'$ is clearly equivalent to $\|\|\cdot\|\|$ for any positive constants. Moreover, with this relationship between the constants we have $c_i\|\|N\|\|_i \le \|\|N\|\|'$

and $b_i\|(D_\nu N)\|_{C^i} \leq \||N\||'$ for all i. Therefore, we can use this in the preceding (i)–(vi) to obtain:

(i) $\|\chi(N)\|_\delta \leq \kappa\||N\||'$;

(ii) $b_i\|D_\nu D^{i-1}(\chi(N))\|_{C^0} \leq (\kappa + Kb_i/c_i)\||N\||'$ $(i = 1, \ldots, k)$;

(iii) $c_i\|D_\xi^i(\chi(N))\|_\delta \leq (\kappa + Kc_{i-1}/b_i)\||N\||'$ $(i = 1, \ldots, k)$;

(iv) $d_l H_{\theta+\delta}[D_\nu D^{k-1}(\chi(N))] \leq (\kappa + Kd_1/c_k)\||N\||'$;

(v) $d_2 H_{\theta+\delta}^\nu[D_\xi^k(\chi(N))] \leq (\kappa + K'd_2/d_1 + Kd_2/c_k)\||N\||'$;

(vi) $d_3 Q_{\theta,\delta}[D_\xi^k(\chi(N))] \leq (\kappa + K''d_3/d_2 + K'd_3/d_1 + Kd_3/c_k)\||N\||'$.

Given any κ' with $\kappa < \kappa' < 1$, we can choose the constants so that each successive one is sufficiently smaller than the previous that all of the coefficients on the right are less than κ'. As $\||\chi(N)\||'$ is the maximum of the left-hand sides, we have $\||\chi(N)\||' \leq \kappa'\||N\||'$ for all $N \in \mathcal{N}$. Consequently, there exists a norm equivalent to $\||\cdot\||$ for which $\|\chi\| \leq \kappa' < 1$, which completes the proof of Lemma 5. $\qquad\square$

References

[B1] G. R. Belickiĭ, *Functional equations, and conjugacy of local diffeomorphisms of finite smoothness class* (Russian), Funktsional. Anal. Appl. i Prilozhen. 7 (1973), 17–28; translation in Functional Anal. Appl. 7 (1973), 268–277.

[B2] ———, *Equivalence and normal forms of germs of smooth mappings* (Russian), Uspekhi Mat. Nauk 33 (1978), 95–155; translation in Russian Math. Surveys 33 (1978), 107–177.

[G1] D. Grobman, *Homeomorphisms of systems of differential equations,* Dokl. Akad. Nauk SSSR 128 (1959), 880–881.

[G2] ———, *Topological classification of neighborhoods of a singular point in n-space,* Mat. Sb. (N.S.) 56(98) (1962), 77–94.

[Ha1] P. Hartman, *On local homeomorphisms of Euclidean space,* Bol. Soc. Mat. Mexicana (2) 5 (1960), 220–241.

[Ha2] ———, *A lemma in the theory of structural stability of differential equations,* Proc. Amer. Math. Soc. 11 (1960), 610–620.

[Ha3] ———, *On the local linearization of differential equations,* Proc. Amer. Math. Soc. 14 (1963), 568–573.

[HP] M. Hirsch and C. Pugh, *Stable manifolds and hyperbolic sets,* Proc. Sympos. Pure Math., 14, pp. 133–163, Amer. Math. Soc., Providence, RI, 1970.

[HPS] M. Hirsch, C. Pugh, and M. Shub, *Invariant manifolds,* Lecture Notes in Math., 583, Springer, New York, 1977.

[M1] P. McSwiggen, *A geometric characterization of smooth linearizability,* Michigan Math. J. 43 (1996), 321–335.

[M2] ———, *On the smoothness of strong unstable foliations,* preprint.

[P] H. Poincaré, *Oeuvres,* vol. 1, XCIX–CV.

[Pu] C. Pugh, unpublished lecture notes.

[Se] G. Sell, *Smooth linearization near a fixed point,* Amer. J. Math. 107 (1985), 1035–1091.

[Sh] M. Shub, *Global stability of dynamical systems,* Springer, New York, 1987.

[S1] S. Sternberg, *Local contractions and a theorem of Poincaré,* Amer. J. Math. 79 (1957), 809–824.

[S2] ———, *On the structure of local homeomorphisms of Euclidean n-space,*
      Amer. J. Math. 80 (1958), 623–631.

 [T] F. Takens, *Partially hyperbolic fixed points,* Topology 10 (1971), 133–147.

Department of Mathematical Sciences
University of Cincinnati
Cincinnati, OH 45221-0025

Pat.McSwiggen@math.uc.edu