

Multiphase averaging for generalized flows on manifolds

H. S. DUMAS

*Department of Mathematical Sciences, University of Cincinnati,
Cincinnati OH 45221–0025, USA*

F. GOLSE

*UFR de Mathématiques, Université Paris VII, Tour 45–55, 5^e étage, 4 place Jussieu,
75251 Paris Cedex 05, France*

P. LOCHAK

DMI, Ecole Normale Supérieure, 45 rue d’Ulm, 75230 Paris Cedex 05, France

(Received 5 October 1992)

Abstract. We present a new proof of a strengthened version of Anosov’s multiphase averaging theorem, originally stated for systems of ODEs with slow variables evolving in \mathbf{R}^m and fast variables evolving on a smooth immersed manifold. Our result allows the fast variables to belong to an arbitrary smooth compact Riemannian manifold, and the vector field to have only Sobolev regularity. This is accomplished using normal form techniques adapted to a slightly generalized version of the DiPerna-Lions theory of generalized flows for ODEs.

1. *Introduction*

We present a new proof of a strengthened version of Anosov’s multiphase averaging theorem [2], which was originally stated for systems of ODEs with slow variables evolving in \mathbf{R}^m and fast variables evolving on a smooth immersed manifold (cf. [3], [8] and [10] for discussions in English of Anosov’s theorem). Our version allows the fast variables to belong to an arbitrary smooth compact Riemannian manifold, and the vector field to have only Sobolev regularity. We make use of the recent work of DiPerna and Lions [6] to weaken the smoothness hypotheses, yet our proof is fundamentally different from Anosov’s. In [2], Anosov carries out a trajectory-by-trajectory analysis of the rate of ‘ergodization’ of the fast variables. We present a proof more along dynamical systems lines, following Neishtadt [11], in making use of functional analysis methods descended from Kasuga’s proof of the adiabatic theorem [9]. In doing so we have probably found the most general situation to which normal form methods (i.e., solving a so-called ‘homological equation’, even existentially) may be applied.

It is worth pointing out that although explicit (and for the most part optimal) error estimates are obtained for multiphase averaging theorems under varying hypotheses in [4], [7], and [11], the proofs of those theorems make use of Fourier series methods, requiring in turn that the fast variables belong to the n -torus $\mathbf{T}^n \equiv \mathbf{R}^n/\mathbf{Z}^n$ and that the unperturbed frequencies be independent of the fast variables. In this case (the so-called standard form for multiphase averaging), the unperturbed system gives rise to linear flow on \mathbf{T}^n , which, under mild assumptions, is ergodic for almost all frequency vectors (those with mutually incommensurate components), and nonergodic (or resonant) otherwise. Upon perturbation, slow variation of the frequency vector allows control of resonances (small divisors) appearing in the transformation to normal form; this in turn permits explicit error estimates. By contrast, we treat the more general case where the fast variables belong to an arbitrary smooth manifold and the frequency depends on both the fast and slow variables, so that the notion of resonance is lost. This requires us to extend the DiPerna–Lions theory of generalized flows to smooth compact Riemannian manifolds. In this case one also operates under the more general assumption (as in [2]) that the unperturbed flow is almost everywhere ergodic. In our proof however, we replace Anosov’s intricate analysis of the behavior of trajectories by a kind of ‘Koopmanism’, so that much of the work is done via standard theorems in functional analysis. In some respects, the proof presented here bears a relationship with Anosov’s proof that is analogous to the relationship between Birkhoff’s and von Neumann’s proofs of their respective ergodic theorems.

The remainder of the article is organized as follows. In §2, we set up and state the averaging theorem, while in §3 we summarize and generalize those parts of the DiPerna–Lions theory of generalized flows to be used in the sequel. In §4 we prove the averaging theorem in two steps: first by transforming to normal form, then by approximately solving the linear (or ‘homological’) equation for the auxiliary function of the transformation.

2. Set-up and main result

Let \mathcal{M} be an arbitrary smooth compact Riemannian manifold of dimension n . For small $\epsilon \geq 0$, we consider the system of ODEs on $\mathbf{R}^m \times \mathcal{M}$

$$\frac{dX}{dt} = \epsilon f(X, Y) \tag{2.1a}$$

$$\frac{dY}{dt} = \omega(X, Y) + \epsilon g(X, Y). \tag{2.1b}$$

associated with the perturbation $(\epsilon f, \omega + \epsilon g)$ of the vector field $(0, \omega)$. Here X and Y denote elements of \mathbf{R}^m and \mathcal{M} , respectively, and $|dY|$ denotes the Riemannian n -density normalized to satisfy the relation $\int_{\mathcal{M}} |dY| = 1$. We assume

- (H1) $f, g \in L_{loc}^\infty \cap W_{loc}^{1,1}(\mathbf{R}^m \times \mathcal{M})$ and $\omega \in L_{loc}^\infty \cap W_{loc}^{1,2}(\mathbf{R}^m \times \mathcal{M})$
- (H2) $\operatorname{div}_Y \omega = \operatorname{div}(0, \omega) = 0$ and $\operatorname{div}_X f + \operatorname{div}_Y g = \operatorname{div}(f, g) \in L_{loc}^\infty(\mathbf{R}^m \times \mathcal{M})$
- (H3) $\bar{f}(X, Y) = 0$ for $|X| \geq R$ for some fixed $R > 0$
- (H4) $\bar{f} \in W_{loc}^{1,\infty}(\mathbf{R}^m)$ where $\bar{f}(X) = \int_{\mathcal{M}} f(X, Y) |dY|$.

In (H1) and (H4) we use standard notation for Sobolev spaces (i.e., $f \in W_{loc}^{1,p}$ means that f together with all its first-order derivatives belong to L_{loc}^p ; cf [1] or [5] for more details).

The goal of averaging theory is to approximate the X -components of solutions of (2.1) by solutions of the so-called averaged system associated to (2.1):

$$\frac{dW}{dt} = \epsilon \bar{f}(W). \quad (2.2)$$

Roughly speaking, if the vector field $(\epsilon f, \omega + \epsilon g)$ is well behaved, this approximation method works for small ϵ , provided the projection onto \mathcal{M} of the unperturbed flow (i.e., the flow of (2.1) in the case $\epsilon = 0$) is ergodic. We introduce this final hypothesis of ergodicity below, after establishing the relevant notation for flows.

As described in the next section, the theory of DiPerna and Lions on transport equations with coefficients in Sobolev spaces shows that, under assumptions (H1) through (H4), the systems (2.1) and (2.2) generate global generalized flows which leave the respective domains $D_R = \overline{B_R(0)} \times \mathcal{M}$ and $\overline{B_R(0)}$ invariant (here $\overline{B_R(0)}$ is the closed ball of radius R centered on the origin in \mathbf{R}^m). We shall denote these flows by $(X_t^\epsilon, Y_t^\epsilon) \equiv (X_t^\epsilon, Y_t^\epsilon)(x, y)$ and $W_t^\epsilon \equiv W_t^\epsilon(x)$ respectively. This notation for flows means, in the classical case, that $t \mapsto (X_t^\epsilon, Y_t^\epsilon) \equiv (X_t^\epsilon, Y_t^\epsilon)(x, y)$ is the integral curve of the vector field $(\epsilon f, \omega + \epsilon g)$ passing through (x, y) at time $t = 0$. We shall use this notation for flows in the remainder of the article, and we point out that the unperturbed flow is consistently denoted (X_t^0, Y_t^0) .

With this notation, we introduce the last assumption on the ergodicity of the unperturbed flow of (2.1):

(H5) Any function $F \in L^2(D_R)$ such that $F \circ (X_t^0, Y_t^0) = F$ for all $t \in \mathbf{R}$ is $dx|dy$ -almost everywhere equal to a function of X only.

Finally, we introduce the pointwise error E_t^ϵ between solutions of the perturbed system (2.1) and its associated averaged system (2.2), defined as

$$E_t^\epsilon(x, y) = X_t^\epsilon(x, y) - W_t^\epsilon(x).$$

(E_t^ϵ is Lipschitz in t and well defined for almost all (x, y) ; see the discussion of generalized flows in §3.)

We may now state our main result as the following

AVERAGING THEOREM. *Assume hypotheses (H1) through (H5) pertaining to systems (2.1) and (2.2), and fix $T > 0$. Then for any p , $1 \leq p < \infty$, the average error*

$$\int_{D_R} \sup_{0 \leq t \leq T/\epsilon} |E_t^\epsilon(x, y)|^p dx|dy| \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0.$$

In particular, for any $r > 0$, we have

$$\text{meas}\{(x, y) \in D_R \mid \sup_{0 \leq t \leq T/\epsilon} |E_t^\epsilon(x, y)| > r\} \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0.$$

We prove this theorem following a brief discussion of generalized flows in the next section. For a short discussion of how a simplified version of this proof may be applied

to Anosov's original theorem (i.e., under stronger smoothness assumptions and in a less general geometric setting), see [8].

3. Review of DiPerna-Lions generalized flows.

For the convenience of the reader, in the present section we gather together some definitions and basic facts concerning generalized flows to which we will refer several times in the course of the proof below.

Let B be an open ball in \mathbf{R}^d and b a vector field on \mathbf{R}^d such that

$$b \in L_{loc}^\infty \cap W_{loc}^{1,1}(\mathbf{R}^d), \quad \operatorname{div} b \in L_{loc}^\infty(\mathbf{R}^d), \quad \operatorname{supp} b \subset \bar{B}. \quad (3.1)$$

The two objects of interest in DiPerna and Lions' theory of generalized flows are

— the ODE

$$\frac{dX}{dt} = b(X), \quad X|_{t=0} = x; \quad (3.2)$$

— the PDE associated to (3.2)

$$\partial_t f = b(x) \cdot \operatorname{grad}_x f, \quad f|_{t=0} = \phi(x) \quad (3.3)$$

for the unknown function $f \equiv f(t, x)$, where $x \in \mathbf{R}^d$ and $t > 0$.

Usually, (3.2) is referred to as the 'equation of characteristics' of the PDE (3.3), whereas (3.3) is referred to as the Liouville equation associated to (3.2). In the classical theory of first-order PDEs (i.e., when the vector field b is smooth), (3.3) and (3.2) are related by the so-called 'method of characteristics' for solving (3.3), which is abstracted by the formula

$$f(t, x) = \phi(X_t(x)), \quad (3.4)$$

where X_t is the flow associated to (3.2).

DiPerna and Lions consider the two objects mentioned above and use PDE techniques to solve (3.3) when the vector field b has only the regularity assumed in (3.1). They prove that (3.3) has a unique solution $f \in L_{loc}^\infty(\mathbf{R}; L^p(\mathbf{R}^d))$ under the assumption that the initial data $\phi \in L^p(\mathbf{R}^d)$, for $1 \leq p \leq \infty$.

A slight modification of their argument shows that the same result is true if one replaces $L^p(\mathbf{R}^d)$ by $L_{loc}^p(\mathbf{R}^d)$ in the above statement; indeed, our assumption on the support of b indicates that the flow is classical and trivial outside \bar{B} . In this case, the generalized flow of b , denoted X_t as in the classical case, is defined as the solution of (3.3) in the sense of distributions for the initial data $\phi = \operatorname{Id}_{\mathbf{R}^d}$.

DiPerna and Lions then prove three important facts concerning the generalized flow X_t :

- (a) Let λ denote Lebesgue measure. Then for all $t \in \mathbf{R}$, the image of λ under X_t , denoted $(X_t)_*(\lambda)$, satisfies the estimate

$$\exp\{-|t| \|\operatorname{div} b\|_{L^\infty(\bar{B})}\} \lambda \leq (X_t)_*(\lambda) \leq \exp\{|t| \|\operatorname{div} b\|_{L^\infty(\bar{B})}\} \lambda. \quad (3.5)$$

This is easily recognized as a weak form of Liouville's theorem, which, in the case of a smooth vector field b , states that

$$\partial_t \det(DX_t) = \operatorname{div} b(X_t) \det(DX_t),$$

DX_t being the Jacobian matrix of X_t .

- (b) When the initial data ϕ belongs to $L^p(\mathbf{R}^d)$, the unique solution of (3.3) is given by (3.4), recast more simply as $f = \phi \circ X$.
- (c) The ODE (3.2) is satisfied in the following sense: for all $\beta \in C^1(\mathbf{R}^d)$, $\beta \circ X \in L^\infty(\mathbf{R} \times \mathbf{R}^d)$ and

$$\partial_t \beta(X) = (D\beta)(X) \cdot b(X), \quad \beta(X)|_{t=0} = \beta, \quad (3.6)$$

where equation (3.6) holds in the distribution sense on $\mathbf{R} \times \mathbf{R}^d$.

We note that in view of a) above, in the case where $\operatorname{div} b = 0$, the Lebesgue measure λ is invariant under the generalized flow:

$$(X_t)_*(\lambda) = \lambda$$

for all $t \in \mathbf{R}$. We shall need a few other simple consequences of results a), b) and c) in the case $\operatorname{div} b = 0$, and we summarize these in the following

LEMMA 1. *Assume (3.1) and $\operatorname{div} b = 0$. Denote by U_t the map defined on $L^2(B, \lambda)$ by*

$$U_t(f) = f \circ X_t,$$

where X_t is the generalized flow of b . Then

- (1) U_t defines a strongly continuous group of isometries on $L^2(B, \lambda)$.
- (2) The infinitesimal generator of U_t is the unbounded operator on $L^2(B, \lambda)$ defined by $(Af)(x) = b(x) \cdot \operatorname{grad} f$ with domain $\mathcal{D}(A) = \{f \in L^2(B, \lambda) / b(x) \cdot \operatorname{grad} f \in L^2(B, \lambda)\}$.

Proof. (1) U_t is an isometry of $L^2(B, \lambda)$ for all $t \in \mathbf{R}$ since the measure λ is invariant under the flow X_t . The group property is a consequence of b) above and the uniqueness of the solution of (3.3) in $L^\infty(\mathbf{R}; L^2(\mathbf{R}^d))$. Finally, the strong continuity property means that

$$\|f \circ X_t - f\|_{L^2} \rightarrow 0 \quad \text{as } t \rightarrow 0 \quad (3.7)$$

for any $f \in L^2(B, \lambda)$. Using the fact that U_t is an isometry of $L^2(B, \lambda)$ for all $t \in \mathbf{R}$ and approximating f by smooth functions in the L^2 -topology reduces the proof to verifying (3.7) when $f \in C^\infty(\bar{B})$. But then, using result c) above,

$$\partial_t f(X_t(x)) = (Df)(X_t(x)) \cdot b((X_t(x))).$$

Multiplying both sides of the above equality by $f(X_t(x)) - f(x)$ leads to

$$\frac{1}{2} \partial_t (f(X_t(x)) - f(x))^2 = (f(X_t(x)) - f(x))(Df)(X_t(x)) \cdot b((X_t(x))).$$

Observe now that the functions f and $(Df) \cdot b$ belong to $L^\infty(B, \lambda)$. Therefore their compositions with X_t , $f(X_t(x))$ and $(Df)(X_t(x)) \cdot b((X_t(x)))$ belong to $L^\infty(\mathbf{R}; L^\infty(B, \lambda))$, since the measure λ is preserved under X_t for all $t \in \mathbf{R}$. Therefore, the right side of (3.8) belongs to $L^\infty(\mathbf{R}; L^\infty(B, \lambda))$. Integrating over $(0, t) \times B$ and letting $t \rightarrow 0$ leads to the convergence stated in (3.7).

- (2) Let $f \in L^2(B, \lambda)$ be such that

$$\frac{1}{t} (f \circ X_t - f) \rightarrow l \quad \text{as } t \rightarrow 0, \quad (3.9)$$

in $L^2(B, \lambda)$. Then for all $\phi \in C_0^\infty(B)$,

$$\int_B \frac{1}{t} (f \circ X_t - f) \phi d\lambda \rightarrow \int_B l \phi d\lambda. \quad (3.10)$$

Equation (3.10) is transformed into

$$\int_B \frac{1}{t} (\phi \circ X_{-t} - \phi) f d\lambda \rightarrow \int_B l \phi d\lambda$$

using the invariance of λ under X_t . But we claim that, for all $\phi \in C_0^\infty(B)$,

$$\frac{1}{t} (\phi \circ X_{-t} - \phi) \rightarrow -b \cdot \text{grad} \phi \quad \text{as } t \rightarrow 0 \quad (3.11)$$

weakly in $L^2(B, \lambda)$. Therefore (3.11) shows that

$$\int_B l \phi d\lambda = - \int_B f b \cdot \text{grad} \phi, \quad (3.12)$$

and hence $l = \text{div}(bf) = b \cdot \text{grad} f \in L^2(B, \lambda)$, since $\text{div} b = 0$. This shows that $\mathcal{D}(A) = \{f \in L^2(B, \lambda) / b(x) \cdot \nabla_x f \in L^2(B, \lambda)\}$ and $Af = b(x) \cdot \text{grad} f$ as announced.

To complete the proof of part 2), it suffices to prove our claim (3.11). We start by proving that

$$\frac{1}{t} (\phi \circ X_{-t} - \phi)$$

viewed as a family of functions of x parametrized by t is bounded in $L^\infty(B, \lambda)$ uniformly in $t \in \mathbf{R}$. This is done as follows: let $\theta \in C_0^\infty(B)$ and $\chi \in C_0^\infty((0, \infty))$. Equation (3.6) shows that

$$\begin{aligned} & \int_0^\infty dt \int_B \phi \circ X_t(x) \theta(x) \chi'(t) d\lambda(x) \\ &= \int_0^\infty dt \int_B (D\phi \circ X_{-t}(x)) \cdot b \circ X_{-t} \theta(x) \chi(t) d\lambda(x). \end{aligned}$$

Therefore

$$\left| \int_0^\infty dt \int_B \phi \circ X_{-t}(x) \theta(x) \chi'(t) d\lambda(x) \right| \leq \|D\phi\|_{L^\infty} \|b\|_{L^\infty} \|\theta\|_{L^1} \|\chi\|_{L^1}$$

whence the map

$$t \mapsto \int_B \phi \circ X_{-t}(x) \theta(x) d\lambda(x)$$

is Lipschitz. Therefore

$$\left| \frac{1}{t} \int_B (\phi \circ X_{-t} - \phi)(x) \theta(x) d\lambda(x) \right| \leq C \|\theta\|_{L^1}$$

and hence

$$\frac{1}{t} (\phi \circ X_{-t} - \phi)$$

is bounded in $L^\infty(B, \lambda)$ (and therefore in $L^2(B, \lambda)$) uniformly in $t \in \mathbf{R}$. As a consequence

$$\frac{1}{t} (\phi \circ X_{-t} - \phi) + b \cdot \nabla \phi \in L^\infty(\mathbf{R}; L^\infty(B, \lambda)),$$

and this expression, viewed as a family of functions of x parametrized by t , is bounded in $L^2(B, \lambda)$ uniformly with respect to $t \in \mathbf{R}$. The Banach-Alaoglu theorem (cf [5]) then shows that the family

$$\frac{1}{t}(\phi \circ X_{-t} - \phi) + b \cdot \nabla \phi$$

is relatively compact in $L^2(B, \lambda)$ equipped with the weak topology. We are going to show that it converges to 0 as $t \rightarrow 0$ in a weaker topology, namely that of the distributions on B . But this is a standard consequence of (3.6) with $\beta = \phi$. Indeed, since $\int (\phi \circ X_t) \chi d\lambda$ is a Lipschitz function of t , one has

$$\begin{aligned} \int_B \frac{1}{t}(\phi \circ X_t - \phi) \chi d\lambda &= \frac{1}{t} \int_0^t \int_B \partial_s(\phi \circ X_s) \chi d\lambda ds \\ &= \frac{1}{t} \int_0^t \int_B ((D\phi) \circ X_s \cdot b \circ X_x) \chi d\lambda ds. \end{aligned} \quad (3.13)$$

Now taking the limit in (3.13) as $t \rightarrow 0$ proves our claim. \square

We now turn to the problem of mollifying a vector field with the regularity assumed in (3.1). This question lies at the heart of DiPerna and Lions' theory. The following lemma is adapted from their paper to the particular case of interest to us. We give the proof first in the flat case as in [6], then extend it to the case of more general spaces below. This is in part to keep the exposition as self-contained as possible, and also to correct a couple of misprints in [6]. We retain the notation introduced above.

LEMMA 2. *Let $v \equiv v(x)$ be a vector field such that*

$$v \in H_{loc}^1(\mathbf{R}^d), \quad \operatorname{div} v = 0 \quad (3.13)$$

and $\rho \equiv \rho(x)$ a function in $C^\infty(\mathbf{R}^d)$ satisfying

$$\rho > 0 \text{ on } B_1(0), \quad \operatorname{supp} \rho = \overline{B}_1(0), \quad \int \rho(x) dx = 1. \quad (3.14)$$

Consider the streaming operator $Af = v \cdot \nabla_x f$ associated with v , and the mollifier $R_\epsilon f = \rho_\epsilon * f$ associated to ρ (where ρ_ϵ is defined by $\rho_\epsilon(x) = \epsilon^{-d} \rho(x/\epsilon)$). The commutator of these two operators $[A, R_\epsilon]$ converges to zero in the following sense:

$$\|[A, R_\epsilon]f\|_{L^1(B_R)} \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0 \quad (3.15)$$

for all $f \in L_{loc}^2(\mathbf{R}^d)$.

Proof. The strategy of the proof is as follows. First, observe that the convergence statement (3.15) is true when both f and v are smooth functions. Convergence under the regularity assumed in (3.13) and for all $f \in L_{loc}^2(\mathbf{R}^d)$ then follows easily from the smooth case and the continuity of the bilinear map

$$(v, f) \mapsto [A, R_\epsilon]f \quad (3.16)$$

from $H_{loc}^1 \times L_{loc}^2$ to L_{loc}^1 (the idea being obviously to approximate both v and f by smooth functions). This continuity will be established in the remaining part of the proof.

A direct calculation shows that

$$[A, R_\epsilon]f = \int (v(x) - v(y)) \cdot \nabla \rho_\epsilon(x - y) f(y) dy.$$

On the other hand

$$\epsilon |\text{grad} \rho_\epsilon|(x) \leq C \rho_{2\epsilon}(x).$$

Therefore

$$\|[A, R_\epsilon]f\|_{L^1(B_R)} \leq C \int_{B_R} dx \int \frac{|v(x) - v(y)|}{\epsilon} \rho_{2\epsilon}(x - y)^{1/2} f(y) \rho_{2\epsilon}(x - y)^{1/2} dy. \quad (3.17)$$

Applying the Cauchy-Schwarz inequality to the right-hand side of (3.17) results in

$$\|[A, R_\epsilon]f\|_{L^1(B_R)} \leq C \|f\|_{L^2(B_{R+1})} \left(\int_{B_R} dx \int \frac{|v(x) - v(y)|^2}{\epsilon^2} \rho_{2\epsilon}(x - y) dy \right)^{1/2}. \quad (3.18)$$

The integral on the right-hand side of (3.18) is transformed into

$$\begin{aligned} & \int_{B_R} dx \int \frac{|v(x) - v(y)|^2}{\epsilon^2} \rho_{2\epsilon}(x - y) dy \\ & \leq C \int_{B_R} dx \frac{1}{\epsilon^d} \int_{|x-y| < 2\epsilon} \frac{|v(x) - v(y)|^2}{\epsilon^2} dy \\ & \leq C \int_{B_R} dx \frac{1}{\epsilon^d} \int_{|x-y| < 2\epsilon} dy \int_0^1 |\text{grad} v(x + t(y - x))|^2 dt \end{aligned}$$

Changing variables from (x, y) to $(x, x - y)$ in this last integral and applying Fubini's theorem leads to

$$\begin{aligned} & \int_{B_R} dx \frac{1}{\epsilon^d} \int_{|x-y| < 2\epsilon} dy \int_0^1 |\text{grad} v(x + t(y - x))|^2 dt \\ & = \frac{1}{\epsilon^d} \int_{|z| < 2\epsilon} dz \int_0^1 dt \int_{B_R} |\text{grad} v(x + tz)|^2 dx \leq C \|\text{grad} v\|_{L^2(B_R)}^2. \end{aligned}$$

Therefore

$$\|[A, R_\epsilon]f\|_{L^1(B_R)} \leq C \|f\|_{L^2(B_{R+1})} \|\text{grad} v\|_{L^2(B_R)}.$$

This last inequality proves the continuity of the bilinear map (3.16). \square

In order to prove the Averaging Theorem in its fully stated generality, we must show that the theory of DiPerna and Lions extends to curvilinear spaces. For the most part, we presume the reader would have no trouble in generalizing all the statements of the present section to the case where \bar{B} of (3.1) is replaced by a smooth compact manifold. For the sake of brevity, we undertake only the most crucial part of this generalization, that of extending Lemma 2.

For this purpose, let X be a compact C^∞ Riemannian manifold of dimension d . Let $(\phi_i)_{1 \leq i \leq N}$ be a finite C^∞ square partition of unity, subordinate to some finite open cover $(U_i)_{1 \leq i \leq N}$ of X . This means that

$$\phi_i \in C^\infty(X), \quad \text{supp} \phi_i \subset\subset U_i, \quad \phi_i \geq 0 \quad \text{and} \quad \sum_{i=1}^N (\phi_i(x))^2 = 1.$$

Without loss of generality, we may assume that each U_i is coordinatized by charts $\chi_i : U_i \rightarrow \mathbf{R}^n$ and that the collection $\{(U_i, \chi_i)\}_{1 \leq i \leq N}$ forms an atlas for X . We denote the inverse of the chart $\chi_i : U_i \rightarrow \chi_i(U_i)$ by θ_i , and the push-forward of the volume element λ under χ_i by $(\chi_i)_*(\lambda)$. Clearly $(\chi_i)_*(\lambda)$ is absolutely continuous with respect to Lebesgue measure, and we denote its density by m_i .

Recalling the definition of ρ_ϵ above (cf (3.14) ff), on each product $U_i \times U_i$ we define the function

$$k_i^\epsilon(x, y) = \rho_\epsilon(\chi_i(x) - \chi_i(y)) m_i(\chi_i(x))^{-1/2} m_i(\chi_i(y))^{-1/2},$$

which in turn allows us to define $K_\epsilon : X \times X \rightarrow \mathbf{R}$ by

$$K_\epsilon(x, y) = \sum_{i=1}^N \phi_i(x) k_i^\epsilon(x, y) \phi_i(y).$$

We then claim that

$$a_\epsilon(x) = \int_X K_\epsilon(x, y) d\lambda(y) \rightarrow 1 \quad \text{as } \epsilon \rightarrow 0$$

uniformly for each $x \in X$, which the reader is invited to check. Since each k_i^ϵ is symmetric, it follows that K_ϵ is also symmetric, and so

$$a_\epsilon(y) = \int_X K_\epsilon(x, y) d\lambda(x) \rightarrow 1 \quad \text{as } \epsilon \rightarrow 0$$

uniformly. It follows that

$$\int_X K_\epsilon(x, y) f(y) d\lambda(y) \rightarrow f(x) \quad \text{as } \epsilon \rightarrow 0$$

in $L^1(X)$ for all $f \in L^1(X)$.

The idea is to write

$$\int_X K_\epsilon(x, y) f(y) d\lambda(y) - f(x) = \int_X K_\epsilon(x, y) (f(y) - f(x)) d\lambda(y) + (a_\epsilon(x) - 1) f(x),$$

the second term of which clearly converges to 0 in $L^1(X)$ as $\epsilon \rightarrow 0$, as does the first term when f is continuous. It remains to estimate

$$\begin{aligned} & \int_X \left| \int_X K_\epsilon(x, y) (f_\eta(y) - f(x)) d\lambda(y) \right| d\lambda(x) \\ & \leq \int_X \int_X K_\epsilon(x, y) |f_\eta(y) - f(x)| d\lambda(y) d\lambda(x) \\ & \leq \int_X a_\epsilon(y) |f_\eta(y) - f(x)| d\lambda(y) \rightarrow 0 \quad \text{as } \eta \rightarrow 0 \end{aligned}$$

uniformly in ϵ , for a sequence of smooth f_η with $f_\eta \rightarrow f$ in L^1 . We therefore conclude that

$$(R_\epsilon f)(x) = \int_X K_\epsilon(x, y) f_\eta(y) d\lambda(y) \tag{3.19}$$

is a bona fide mollifier on X .

We now wish to estimate $\|[A, R_\epsilon]f\|_{L^1(X)}$ in the case where $f \in L^2(X)$ and v is a divergence-free vector field belonging to the Sobolev space $H^1(X)$. We compute

$$\begin{aligned} [A, R_\epsilon]f &= \sum_{i=1}^N \left\{ \int_X A \otimes 1 (\phi_i(x) k_i^\epsilon(x, y)) \phi_i(y) f(y) d\lambda(y) \right. \\ &\quad \left. - \int_X \phi_i(x) k_i^\epsilon(x, y) 1 \otimes A (\phi_i(y) f(y)) d\lambda(y) \right\} \\ &= \sum_{i=1}^N \int_X (A \otimes 1 + 1 \otimes A) (\phi_i(x) k_i^\epsilon(x, y)) \phi_i(y) f(y) d\lambda(y) \end{aligned}$$

which, since v is divergence-free, reduces to

$$\begin{aligned} &\sum_{i=1}^N \int_X A \otimes 1 (\phi_i(x)) k_i^\epsilon(x, y) \phi_i(y) f(y) d\lambda(y) \\ &\quad + \sum_{i=1}^N \int_X (A \otimes 1 + 1 \otimes A) (k_i^\epsilon(x, y)) \phi_i(x) \phi_i(y) f(y) d\lambda(y). \end{aligned}$$

The first term is easily estimated, as there are no derivatives acting on k_i^ϵ . As for the second term, we coordinatize under the integral sign to obtain

$$\begin{aligned} &\int_{U_i} (A \otimes 1 + 1 \otimes A) (k_i^\epsilon(x, y)) \phi_i(x) \phi_i(y) f(y) d\lambda(y) \\ &= \int_{\chi_i(U_i)} (a_\epsilon(x') \cdot \text{grad}_{x'} + a_\epsilon(y') \cdot \text{grad}_{y'}) \\ &\quad \times \left(\rho_\epsilon(x' - y') m_i(x')^{-1/2} m_i(y')^{-1/2} \phi_i(\theta_i(x')) \phi_i(\theta_i(y')) f(\theta_i(y')) \right) d\lambda(y'). \end{aligned}$$

In the above integral, the only difficult term is

$$\begin{aligned} &\int_{\chi_i(U_i)} (a_\epsilon(x') \cdot \nabla_{x'} + a_\epsilon(y') \cdot \text{grad}_{y'}) \left(\rho_\epsilon(x' - y') m_i(x')^{-1/2} m_i(y')^{-1/2} \right) \\ &\quad \times \phi_i(\theta_i(x')) \phi_i(\theta_i(y')) f(\theta_i(y')) d\lambda(y') \end{aligned}$$

which is estimated as in Lemma 2.

It is now a simple matter to retrace the steps in the proof of Lemma 2 to ensure that it indeed extends to the case where \bar{B} in (3.1) is replaced by X . \square

We may now proceed with the proof of the main theorem.

4. Proof of the averaging theorem

PART 1. In order to prove the main result, we follow Neishtadt [11] by introducing the smooth auxiliary function $S \in C_0^\infty(\mathbf{R}^m \times \mathcal{M}; \mathbf{R}^m)$ (to be determined later) and the intermediate variable $P = X + \epsilon S(X, Y)$. (The reader unfamiliar with standard averaging techniques may be interested to see that this change of variables will be chosen so that (2.1a) is transformed into $dP/dt = \epsilon(\bar{f}(P) + o(1))$. We use this only indirectly below; see [10] for a more complete discussion of this most rudimentary of normal form techniques.)

Now integrating the difference $(d/dt)(X - W) = (d/dt)(P - W - \epsilon S)$ from 0 to t leads to

$$\begin{aligned} E_t^\epsilon(x, y) &= X_t^\epsilon(x, y) - W_t^\epsilon(x) \\ &= \epsilon \int_0^t (\bar{f}(X_u^\epsilon(x, y))) - \bar{f}(W_u^\epsilon(x)) du \\ &\quad - \epsilon \left(S(X_t^\epsilon(x, y), Y_t^\epsilon(x, y)) - S(x, y) \right) + \epsilon \int_0^t R^\epsilon(X_u^\epsilon(x, y), Y_u^\epsilon(x, y)) du \end{aligned} \quad (4.1)$$

where

$$R = f - \bar{f} + \epsilon f \cdot \text{grad}_x S + (\omega + \epsilon g) \cdot \text{grad}_y S. \quad (4.2)$$

The justification of (4.1) is straightforward in the case of classical flows; for generalized flows however, some care is needed. We observe that

$$\partial_t(E^\epsilon - \epsilon S(X^\epsilon, Y^\epsilon)) = \epsilon(\bar{f}(X^\epsilon) - \bar{f}(W^\epsilon) + R^\epsilon(X^\epsilon, Y^\epsilon)), \quad (4.3)$$

in the sense of distributions, according to formula (3.6). Observing that the right-hand side of (4.3) lies in $L^\infty(D_R)$, it follows that $E^\epsilon - \epsilon S(X^\epsilon, Y^\epsilon)$ is Lipschitz in time, so, for all $t > 0$, the integrated relation (4.1) holds in the sense of distributions on D_R .

Since all distributions considered in (4.1) are locally integrable functions, equality (4.1) which holds in the sense of distributions on D_R also holds pointwise almost everywhere in (x, y) . This implies that for all $t > 0$,

$$\begin{aligned} |E_t^\epsilon(x, y)| &\leq L\epsilon \int_0^t |X_u^\epsilon - W_u^\epsilon|(x, y) du + \epsilon |S((X_t^\epsilon, Y_t^\epsilon)(x, y))| \\ &\quad + \epsilon |S(x, y)| + \epsilon \int_0^t |R^\epsilon((X_u^\epsilon, Y_u^\epsilon)(x, y))| du \end{aligned} \quad (4.4)$$

where L is the Lipschitz constant for \bar{f} .

Again following Neishtadt, we now set $Z(t) = \int_{D_R} |E_t^\epsilon(x, y)| dx dy$. Then integrating over all initial conditions (x, y) (using Fubini's Theorem for positive measurable functions to reverse the order of integration), we obtain

$$\begin{aligned} Z(t) &\leq L\epsilon \int_0^t Z(u) du + \epsilon \int_{D_R} |S(X_t^\epsilon, Y_t^\epsilon)| dx dy + \epsilon \|S\|_{L^1} \\ &\quad + \epsilon \int_0^t \int_{D_R} |R^\epsilon(X_u^\epsilon, Y_u^\epsilon)| dx dy du. \end{aligned} \quad (4.5)$$

If the flow of (2.1) were C^1 , it would be routine to verify that the Jacobian of the reverse flow $(X_{-t}^\epsilon, Y_{-t}^\epsilon)$ is $O(1)$ over $O(1/\epsilon)$ -time intervals. The analogue of this procedure in the case of generalized flows is estimate (3.5). We use (3.5) together with the fact that the divergence of b (b being in this case the perturbed vector field $(\epsilon f, \omega + \epsilon g)$) is $O(\epsilon)$ to ensure that the dilation or compression of the image measure $(X_t^\epsilon, Y_t^\epsilon)_*(dx dy)$ is $O(1)$, bounded away from 0, on $O(1/\epsilon)$ -time intervals. More precisely, using (3.5) and (H2), we deduce the existence of constants $C_2 \geq C_1 > 0$ such that, given $T > 0$, for all $\epsilon > 0$, the inequality

$$C_1 dx dy \leq (X_t^\epsilon, Y_t^\epsilon)_*(dx dy) \leq C_2 dx dy$$

holds for all $0 \leq t \leq T/\epsilon$. We may therefore write

$$Z(t) \leq L\epsilon \int_0^t Z(u)du + c(\epsilon\|S\|_{L^1} + T\|R\|_{L^1})$$

for some positive constant $c > 0$ and for all $t \in [0, T/\epsilon]$. A standard Gronwall estimate then leads to

$$Z(t) \leq c(\epsilon\|S\|_{L^1} + T\|R\|_{L^1}) \exp(\epsilon Lt). \quad (4.6)$$

But we know (cf (4.2)) that

$$\|R\|_{L^1} \leq \|\tilde{f} - \omega \cdot \text{grad}_y S\|_{L^1} + \epsilon\|f \cdot \text{grad}_x S + g \cdot \text{grad}_y S\|_{L^1} \quad (4.7)$$

(here $\tilde{f} = f - \bar{f}$), so if we can choose the auxiliary function S so as to make the difference

$$\|\tilde{f} - \omega \cdot \text{grad}_y S\|_{L^1} \quad (4.8)$$

sufficiently small with ϵ , we obtain

$$\sup_{t \in [0, T/\epsilon]} Z(t) = \sup_{t \in [0, T/\epsilon]} \int_{D_R} |E_t^\epsilon(x, y)| dx dy \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0 \quad (4.9)$$

by taking the supremum of (4.6) over $[0, T/\epsilon]$. Observe that (4.8) replaces the homological equation for linear flows on tori.

However, the convergence stated in (4.9) is weaker than that announced in the Averaging Theorem, since only the following inequality is true:

$$\sup_{t \in [0, T/\epsilon]} \int_{D_R} |E_t^\epsilon(x, y)| dx dy \leq \int_{D_R} \sup_{t \in [0, T/\epsilon]} |E_t^\epsilon(x, y)| dx dy,$$

and not (at least in general) the equality, as claimed in [11]. The stronger statement in the Averaging Theorem is instead obtained as follows.

Observe first that $B_R(0)$ is invariant under both the perturbed flow X_t^ϵ and W_t^ϵ (cf hypothesis (H3)); hence $|E_t^\epsilon(x, y)| \leq 2R$, and in particular

$$\epsilon \int_0^{T/\epsilon} \int_{D_R} |E_t^\epsilon(x, y)|^2 dx dy dt \leq T \cdot 2R \cdot \sup_{t \in [0, T/\epsilon]} \int_{D_R} |E_t^\epsilon(x, y)| dx dy \rightarrow 0 \quad (4.10)$$

as $\epsilon \rightarrow 0$. Therefore

$$\begin{aligned} E_t^\epsilon(x, y)^2 &= \int_0^t 2E_s^\epsilon(x, y) \partial_s E_s^\epsilon(x, y) ds \\ &\leq 2 \left(\epsilon \int_0^t E_s^\epsilon(x, y)^2 ds \right)^{1/2} \cdot \left(\frac{1}{\epsilon} \int_0^t |\partial_s E_s^\epsilon(x, y)|^2 ds \right)^{1/2} \end{aligned}$$

according to the Cauchy-Schwarz inequality. In particular,

$$\begin{aligned} \sup_{t \in [0, T/\epsilon]} E_t^\epsilon(x, y)^2 &\leq 2 \left(\epsilon \int_0^{T/\epsilon} E_s^\epsilon(x, y)^2 ds \right)^{1/2} \cdot \left(\frac{1}{\epsilon} \int_0^{T/\epsilon} |\partial_s E_s^\epsilon(x, y)|^2 ds \right)^{1/2} \\ &\leq 4\sqrt{T} \|f\|_{L^\infty(D_R)} \left(\epsilon \int_0^{T/\epsilon} E_s^\epsilon(x, y)^2 ds \right)^{1/2} \quad (4.11) \end{aligned}$$

according to assumption (H1) which ensures that f is essentially bounded in D_R . Integrating both sides of the above inequality leads to

$$\begin{aligned}
 & \int_{D_R} \sup_{t \in [0, T/\epsilon]} E_t^\epsilon(x, y)^2 dx |dy| \\
 & \leq 4\sqrt{T} \|f\|_{L^\infty(D_R)} \sqrt{\text{meas}(D_R)} \left(\epsilon \int_{D_R} dx |dy| \int_0^{T/\epsilon} E_s^\epsilon(x, y)^2 ds \right)^{1/2} \\
 & \leq 4\sqrt{2} \|f\|_{L^\infty(D_R)} T \sqrt{R \text{meas}(D_R)} \left(\sup_{t \in [0, T/\epsilon]} \int_{D_R} |E_t^\epsilon(x, y)| dx |dy| \right)^{1/2} \rightarrow 0
 \end{aligned} \tag{4.12}$$

as $\epsilon \rightarrow 0$ according to (4.10). Using the bound $|E_t^\epsilon(x, y)| \leq 2R$ yields the convergence statement of the Averaging Theorem

$$\int_{D_R} \sup_{t \in [0, T/\epsilon]} |E_t^\epsilon(x, y)|^p dx |dy| \rightarrow 0, \quad \text{as } \epsilon \rightarrow 0 \tag{4.13}$$

for all $2 \leq p < \infty$; the fact that this statement also holds for $1 \leq p \leq 2$ follows from the case $p = 2$ and Hölder's inequality, since D_R is of finite volume.

The proof of the Averaging Theorem therefore reduces to approximately solving the homological equation (i.e, to finding a smooth S so that (4.8) is sufficiently small), which we set out to do below in Part 2.

PART 2. In the spirit of Kasuga [9], we use functional analytic methods to approximately 'solve' (i.e., to demonstrate the existence of an approximate weak solution of) the homological equation

$$\tilde{f} = \omega \cdot \text{grad}_y S.$$

To this end, we recall that the unperturbed generalized flow of (2.1) is denoted by (X_t^0, Y_t^0) . To this flow is associated the mapping U_t on $L^2(D_R)$ defined by

$$U_t(F) = F \circ (X_t^0, Y_t^0)$$

for any $F \in L^2(D_R)$. Part (1) of Lemma 1 ensures that U is a strongly continuous group of isometries in $L^2(D_R)$ with infinitesimal generator A defined by

$$(AF) = \omega \cdot \text{grad}_y F$$

with domain

$$\mathcal{D}(A) = \{F \in L^2(D_R) / \omega \cdot \text{grad}_y F \in L^2(D_R)\}.$$

From Stone's theorem, we conclude that A is skew-adjoint, from which it follows that

$$\overline{\mathcal{R}(A)} = \mathcal{N}(A)^\perp.$$

We will now use the ergodicity hypothesis (H5) to show that $\mathcal{N}(A) = \{F(X)\}$, the set of functions depending a.e. only on X .

It is clear that $\mathcal{N}(A) \supset \{F(X)\}$. To see that the reverse inclusion also holds, let $F \in \mathcal{N}(A)$, and $t \neq 0$. Since $\mathcal{N}(A) \subset \mathcal{D}(A)$, we may write $U_t F - F =$

$\int_0^t A(U_s F) ds$, and since A commutes with the group it generates, this last integral becomes $\int_0^t U_s(AF) ds = 0$. F is therefore invariant under U , and by (H5), we must have $F \in \{F(X)\}$.

Since $\tilde{f} = f - \overline{f}$ has zero mean over \mathcal{M} , it follows that $\tilde{f} \in \{F(X)\}^\perp$, which in turn leads to $\tilde{f} \in \mathcal{R}(A)$. Therefore, given any $\eta > 0$, we may choose a transformation $S \in \mathcal{D}(A)$ with

$$\|\tilde{f} - AS\|_{L^2} < \eta. \quad (4.14)$$

The procedure just described approximately ‘solves’ the homological equation, but only in the weak sense—that is, the S appearing in (4.8) need not be a smooth function. To conclude, we must show that we can regularize S while maintaining the estimate (4.8).

We choose a sequence of mollifiers as in Lemma 2, denoted R_ϵ , and consider the regularized auxiliary function $S_\epsilon = R_\epsilon S$. Observe that

$$\begin{aligned} & \|\tilde{f} - AS_\epsilon\|_{L^1(D_R)} \\ & \leq \text{meas}(D_R)^{1/2} \|\tilde{f} - AS\|_{L^2(D_R)} + \|R_\epsilon(AS) - AS\|_{L^1(D_R)} + \|[A, R_\epsilon]S\|_{L^1(D_R)} \end{aligned} \quad (4.15)$$

since

$$AS_\epsilon - AS = AR_\epsilon S - AS = R_\epsilon(AS) - AS + [A, R_\epsilon]S.$$

Now let η be an arbitrary positive number. According to (4.14), we can choose $S \in \mathcal{D}(A)$ so that the first term on the right-hand side of (4.15) is less than η . Since $S \in \mathcal{D}(A)$ and $\text{meas}(D_R) < \infty$, $AS \in L^1(D_R)$ and therefore the second term on the right-hand side of (4.15) can be made smaller than η by choosing $0 < \epsilon < \epsilon_0(\eta)$. As for the third term on the right-hand side of (4.15), Lemma 2 shows that it can also be made less than η by choosing $0 < \epsilon < \epsilon_1(\eta)$. Therefore, for $0 < \epsilon < \min(\epsilon_0(\eta), \epsilon_1(\eta))$, the left-hand side of (4.15) is less than 3η .

This shows that (4.14) can be made as small as desired by choosing ϵ sufficiently small. In view of the remarks following (4.9) and (4.13), the proof of the Averaging Theorem is complete.

Acknowledgments. We thank Patrick Gérard for helpful discussions concerning the theory of DiPerna and Lions. One of us (HSD) was partly supported by the National Science Foundation under Grant DMS-9106812, and would like to thank the mathematics departments of the Universities of Picardy (Amiens), and Paris VII (Jussieu) for their hospitality while much of this work was carried out.

REFERENCES

- [1] R.A. Adams. *Sobolev Spaces*, Academic Press, New York, 1975.
- [2] D.V. Anosov. Averaging in systems of ODEs with rapidly oscillating solutions, *Izv. Akad. Nauk. SSSR* **24** (1960), 721–742 (untranslated).
- [3] V.I. Arnol’d (ed). *Dynamical Systems III Encyclopedia of Mathematical Sciences* Vol. III, Springer, Berlin, 1988.

- [4] V.I. Bakhtin. Averaging in multifrequency systems, *Funktional. Anal. i Priložen.* **20**, No. 2 (1986), 1–7, English translation: *Functional Anal. Appl.* **20**, No. 2 (1986), 83–88.
- [5] H. Brezis. *Analyse fonctionnelle, théorie et applications*, Masson, Paris, 1983.
- [6] R. DiPerna & P.-L. Lions. Ordinary differential equations, transport theory and Sobolev spaces, *Inv. Math.* **98** (1989), 511–548.
- [7] M.M. Dodson, B.P. Rynne, & J.A.G. Vickers. Averaging in multifrequency systems, *Nonlinearity* **2** (1989), 137–148.
- [8] H.S. Dumas. A new proof of Anosov’s averaging theorem, to appear in *Hamiltonian Dynamical Systems: History, Theory, and Applications* (H.S. Dumas, K.R. Meyer, and D.S. Schmidt, eds.), Springer, 1995.
- [9] T. Kasuga. On the adiabatic theorem for the Hamiltonian system of differential equations in classical mechanics (parts I, II, and III), *Proc. Acad. Japan* **37** (1961), 366–382.
- [10] P. Lochak & C. Meunier. *Multiphase Averaging for Classical Systems*, Applied Mathematical Sciences Vol. 72, Springer, New York, 1988.
- [11] A.I. Neishtadt. Averaging in multi-frequency systems, *Dokl. Akad. Nauk. SSSR Mech.* **223 (2)** (1975), 314–317, English translation: *Soviet Phys. Dokl.* **20 (7)** (1975), 492–494; Averaging in multi-frequency systems II, *Dokl. Akad. Nauk. SSSR Mech.* **226 (6)** (1976), 1295–1298, English translation: *Sov. Phys. Dokl.* **21 (2)** (1976), 80–82.