Internet Supplement for Introduction to Mechanics and Symmetry

A Basic Exposition of Classical Mechanical Systems

Second Edition

Jerrold E. Marsden
CDS 107-81
Caltech
Pasadena, CA 91125
USA
marsden@cds.caltech.edu

Tudor S. Ratiu
Departement de Mathematiques
Ecole Polytechnique Federale de Lausanne
CH - 1015 Lausanne
Switzerland
ratiu@dma.epfl.ch

Last modified on 19 December 1998

Contents

Preface

This supplement contains a number of topics that are somewhat peripheral to the main flow of the text itself, so that to keep the book within a reasonable size, we have placed them here. This does not mean that they are any less important, but as usual, one has to make choices, sometimes difficult ones. We have organized the material by Chapter to match that of the text as far as possible.

This supplement is being continually updated and we appreciate comments and suggestions from readers. Please also note that you can get the current errata for the main text from the site

`http://www.cds.caltech.edu/~marsden`

Jerry Marsden
Pasadena, California

Tudor Ratiu
Lausanne, Switzerland

December, 1998

N6
Cotangent Bundles

N6.A Linearization of Hamiltonian Systems

One process of linearizing a system is by doubling its dimension using the tangent operation. In fact, if P is a symplectic (or even Poisson) manifold, then so is TP in a natural way. We will show how this is established below.

A second method is that of linearizing along a given solution. For example, to linearize a Hamiltonian system on a symplectic manifold at a fixed point, one usually wants the linearized Hamiltonian to be the second variation of the original Hamiltonian at the fixed point. The tangent linearization does *n*ot give this; in canonical coordinates q^i, p_i, the tangent linearized symplectic structure is

$$dq^i \wedge d(\delta p_i) + d(\delta q^i) \wedge dp_i \tag{N6.A.1}$$

in the variables $(q^i, p_i, \delta q^i, \delta p_i)$. However, at a fixed point, it is often desirable to use the given symplectic form simply evaluated at the fixed point, which has the expression

$$d(\delta q^i) \wedge d(\delta p_i), \tag{N6.A.2}$$

while (N6.A.1) restricts to zero.

One can use "symplectic connections" to compare tangent spaces at different points along the unperturbed curve and thus make the linearization process meaningful. A useful class of intrinsic symplectic connections on cotangent bundles of Lie groups is constructed in Marsden, Ratiu, and

Raugel [1991]. For systems with a symmetry group G, they use a G-invariant connection and this gives, via reduction, a linearization theory for Lie–Poisson systems. For instance, the rigid body and ideal fluid flow is linearized in this fashion. One also gets a generalization of the linearization procedure at a fixed point noted in Holm, Marsden, Ratiu, and Weinstein [1985] and Abarbanel, Holm, Marsden, and Ratiu [1986].

Hamiltonian Systems in $\mathbb{R}^{2n}$. Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ be a Hamiltonian function, which in canonical coordinates (q^i, p_j) gives rise to Hamilton's equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}. \tag{N6.A.3}$$

Linearizing along a solution curve $(q^i(t), p_i(t))$ and calling the new variables $(\delta q^i, \delta p_i)$ we get the equations

$$\begin{aligned}(\delta q^i)^{\cdot} &= \frac{\partial^2 H}{\partial q^j\, \partial p_i}\delta q^j + \frac{\partial^2 H}{\partial p_j\, \partial p_i}\delta p_j,\\ (\delta p_i)^{\cdot} &= -\frac{\partial^2 H}{\partial q^j\, \partial q^i}\delta q^j - \frac{\partial^2 H}{\partial q^i\, \partial p_j}\delta p_j.\end{aligned} \tag{N6.A.4}$$

The matrix of the canonical symplectic form $d(\delta q^i) \wedge d(\delta p_i)$ is

$$\mathbb{J} = \begin{bmatrix} 0 & \mathbb{I} \\ -\mathbb{I} & 0 \end{bmatrix}.$$

Recall (see §2.7) that a linear operator with matrix

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

is infinitesimally symplectic, that is, $T^t\mathbb{J} + \mathbb{J}T = 0$, or equivalently, T is ω-skew, if and only if B and C are symmetric matrices and $D = -A^T$. The linear system (N6.A.4) has a matrix clearly satisfying these conditions and, therefore, it defines a Hamiltonian system in the $(\delta q^i, \delta p_i)$-variables, whose Hamiltonian function is verified to be the second variation:

$$\tfrac{1}{2}\omega(T(\delta q^i, \delta p_i), (\delta q^i, \delta p_i)) = \tfrac{1}{2}\delta^2 H(q^i(t), p_i(t))(\delta q^i, \delta p_i)^2. \tag{N6.A.5}$$

The same argument and formulas hold for infinite-dimensional weak symplectic vector spaces $E \times E'$, where E' and E are (weakly) paired. One of the goals of Marsden, Ratiu, and Raugel [1991] is to generalize this simple procedure to arbitrary symplectic manifolds. Formula (N6.A.5) cannot be correct, in general, since the second variation of a function does not make intrinsic sense, except at critical points. Additional structure is needed to correct the second variation by the addition of terms making the resulting formula invariant.[1]

Infinite Dimensional Systems. There are a number of several interesting infinite-dimensional systems whose phase spaces are of the form $U \times E'$, where U is open in a Banach space E weakly paired with E'. In all of these cases the linearized equations are infinite-dimensional versions of (N6.A.4) and the Hamiltonian function is given by the second variation of the original Hamiltonian along a given integral curve. As we have mentioned, one of the purposes of Marsden, Ratiu, and Raugel [1991] is to generalize this to the nontrivial case. The latter include systems like the rigid body and fluids, charged fluids, Maxwell–Vlasov equations, etc. However, the case with a trivial connection still includes a surprisingly large number of interesting systems. Here are some examples:

Examples

1. The Sine-Gordon equation $u_{tt} - u_{xx} = \sin u$ has phase space $E \times E'$, where E consists of maps $u : \mathbb{R} \to \mathbb{R}$ (one can also use maps $u : \mathbb{R} \to S^1$, but use of the universal covering space $\mathbb{R}$ of S^1 gives a *linear* space) and E' consists of maps $\dot{u} : \mathbb{R} \to \mathbb{R}$; $E \times E'$ has the canonical symplectic structure. The Hamiltonian has the form kinetic plus potential energy (see Chernoff and Marsden [1974] for details).

2. The Yang-Mills equations have phase space $T^*\mathcal{A}$, where $\mathcal{A}$ is the space of connections on a given principal bundle, which is an affine space, so again we can put the trivial symplectic connection on $T^*\mathcal{A}$. The Yang-Mills equations are Hamiltonian on $T^*\mathcal{A}$ relative to the canonical symplectic structure, so again (N6.A.4) is applicable and the Hamiltonian is the second variation of H. See, for example, Arms, Marsden, and Moncrief [1982] for the explicit formula. One of the interesting complications in this example is the presence of a gauge symmetry; the statements above are valid in any gauge. Interestingly, the symplectic form is always canonical, but the Hamiltonian is linear in the so-called atlas fields, representing the gauge freedom (the coefficients of the atlas fields are the momentum map for the gauge group).

[1] Another motivation for working in this general context is to deal with Hamiltonian systems in Lie–Poisson spaces, which, as we explore in detail in Chapters 13 and 14, is equivalent to G-invariant Hamiltonian systems on T^*G, where G is a Lie group. At critical points of $H + C$, where C is a Casimir on $\mathfrak{g}^*$ ($\mathfrak{g}^*$ is the dual of the Lie algebra $\mathfrak{g}$ of G), such a linearization has been carried out in Holm, Marsden, Ratiu, and Weinstein [1985] and Abarbanel, Holm, Marsden, and Ratiu [1986]; as expected, the Hamiltonian function of the linearized equations is the second variation of $H + C$, but the Poisson structure instead of being Lie–Poisson is a "frozen coefficient" Poisson bracket.

3. General relativity (in dynamical form) has phase space $T^* \operatorname{Riem}(M)$, where $\operatorname{Riem}(M)$ is the space of Riemannian metrics on a *fixed* hypersurface M. Again the dynamical equations are Hamiltonian on $T^* \operatorname{Riem}(M)$ relative to the *canonical* symplectic structure (for any choice of gauge). Thus, again we can put the trivial symplectic connection on $T^* \operatorname{Riem}(M)$ and formulas (N6.A.4) and (N6.A.5) (in their obvious infinite-dimensional generalization) apply. These linearized equations are studied in some detail, for the purpose of getting results on the *space* of nonlinear solutions, in Fischer, Marsden, and Moncrief [1980] and Arms, Marsden, and Moncrief [1982]. ♦

An interesting question here is to *couple* these systems to ones with nontrivial phase space. For instance, charged fluids, general relativistic fluids or elasticity, the Maxwell–Vlasov equations, etc., are such systems. All of these will produce nontrivial linearizations by these methods.

The Tangent Symplectic Structure. If (P, Ω) is a symplectic manifold, the "flat map" $\Omega^\flat : TP \to T^*P$ is a diffeomorphism. Then TP becomes an exact symplectic manifold if the map $\Omega^\flat : TP \to T^*P$ is used to pull back the canonical one-form on T^*P. This one-form on TP, denoted Θ_T, has the expression

$$\langle (\Theta_T)_v, w \rangle = \Omega_z(v, T\tau_P(w)), \tag{N6.A.6}$$

where $v \in T_zP$, $w \in T_v(TP)$, $\tau_P : TP \to P$ is the projection, and $\langle\, , \rangle$ denotes the pairing between $T^*(TP)$ and $T(TP)$. In this way, TP becomes a symplectic manifold with symplectic form $\Omega_T = -\mathbf{d}\Theta_T$.

If $f : P \to P$ is a diffeomorphism one verifies that $Tf : TP \to TP$ is symplectic iff f is symplectic.

We remark in passing that a vector field X is locally Hamiltonian if and only if $X(P)$ is a Lagrangian submanifold of (TP, Ω_T) (see Abraham and Marsden [1978], §5.3, and Sánchez de Alvarez [1986, 1989]).

The First Variation Equation. Let φ_t be the flow of a Hamiltonian vector field X_H on a symplectic manifold P and let $\psi_t = T\varphi_t$ be the tangent flow and Y be its generating vector field. Let $s_P : T(TP) \to T(TP)$ be the ***canonical involution*** given locally by $s_P(u, v, \dot{u}, \dot{v}) = (u, \dot{u}, v, \dot{v})$. One verifies that $Y = s_P \circ TX_H$ is Hamiltonian with respect to the symplectic form Ω_T on TP with the Hamiltonian function $\mathcal{H}(v) = \Omega(X_H(p), v), v \in T_pP$, which is given in coordinates by the formula

$$\mathcal{H}(q^i, p_i, v^i, w_i) = v^i \frac{\partial H}{\partial q^i} + w_j \frac{\partial H}{\partial p_j}. \tag{N6.A.7}$$

The Hamiltonian system $Y = X_{\mathcal{H}}$ on TP is called the ***linearized Hamiltonian system*** or ***first variation equation of*** X_H.

If Q is a pseudo-Riemannian manifold and $P = TQ$ with the symplectic form induced by the metric, the linearized Hamiltonian $\mathcal{H}$ of the Hamiltonian given by the kinetic energy of the metric on Q gives rise to the Hamiltonian vector field $X_{\mathcal{H}}$, which coincides with the first variation equation for geodesics, which is an important construction in geometry (see, for instance, Milnor [1965]).

Linearization with Respect to a Parameter. Let H_ϵ be a family of Hamiltonian functions on P depending smoothly on a parameter $\epsilon \in \mathbb{R}$. Let H_0 denote the value of H_ϵ at $\epsilon = 0$ and

$$H^1 = \left.\frac{dH_\epsilon}{d\epsilon}\right|_{\epsilon=0}.$$

Let φ_t^ϵ be the flow of the Hamiltonian vector field with Hamiltonian H_ϵ and let

$$\left.\frac{d}{d\epsilon}\varphi_t^\epsilon(p)\right|_{\epsilon=0} = \varphi_t'(p) \in T_{\varphi_t(p)}P.$$

Since $\varphi_0^\epsilon(p) = p$, we have $\varphi_0'(p) = 0$. Thus φ_t' *is an integral curve of the Hamiltonian vector field* $X_{\mathcal{H}^1}$ *on* $(TP, \mathbf{d}\Theta_T)$*, where*

$$\mathcal{H}^1 = \langle \mathbf{d}H_0, \cdot\rangle + \tau_P^* H^1, \tag{N6.A.8}$$

with$\tau_P : TP \to P$ *the canonical tangent bundle projection,* $\langle\, ,\rangle$ *the pairing between* T^*P *and* TP*, and* $\langle \mathbf{d}H_0, \cdot\rangle : TP \to \mathbb{R}$ *is given by*

$$\langle \mathbf{d}H_0, \cdot\rangle (v_p) := \langle \mathbf{d}H_0(p), v_p\rangle \tag{N6.A.9}$$

for $v_p \in T_pP$. In local coordinates (q^i, p_i, v^i, w_i) on TP,

$$\begin{aligned}\mathcal{H}^1(q^i, p_i, v^i, w_i) &= v^i \frac{\partial H_0}{\partial q^i} + w_i \frac{\partial H_0}{\partial p_i} + H^1(q^i, p_i)\\ &= \mathcal{H}_0(q^i, p_i, v^i, w_i) + H^1(q^i, p_i),\end{aligned} \tag{N6.A.10}$$

where $\mathcal{H}_0$ is given in terms of H_0 by (N6.A.7). Hamilton's equations for $\mathcal{H}^1$ on TP relative to the symplectic form Ω_T are

$$\left.\begin{aligned}\frac{dq^i}{dt} &= \frac{\partial H_0}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H_0}{\partial q^i},\\ \frac{dv^i}{dt} &= \left(v^j \frac{\partial}{\partial q^j} + w_j \frac{\partial}{\partial p_j}\right)\frac{\partial H_0}{\partial p_i} + \frac{\partial H^1}{\partial p_i},\\ \frac{dw_i}{dt} &= -\left(v^j \frac{\partial}{\partial q^j} + w_j \frac{\partial}{\partial p_j}\right)\frac{\partial H_0}{\partial q^i} - \frac{\partial H^1}{\partial q^i}.\end{aligned}\right\} \tag{N6.A.11}$$

One calls this the ***first variation equation relative to a parameter***. If we set $H^1 = 0$ we recover the first variation equation (N6.A.4) for $X_{\mathcal{H}_0}$ discussed earlier, with $\mathcal{H}_0 = H, v^i = \delta q^i$, and $w_i = \delta p_i$.

Further details on the linearization of Hamiltonian systems and the use of symplectic connections to accomplish this may be found in Marsden, Ratiu, and Raugel [1991].

N7
Lagrangian Mechanics

N7.A The Classical Limit and the Maslov Index

The purpose of this section is to give a brief introduction through the simplest examples, of the quantum–classical relationship and the Maslov index, following the exposition in Marsden and Weinstein [1979]. For further information and generalizations, the reader may consult Guillemin and Sternberg [1977, 1984]; Woodhouse [1992] and Bates and Weinstein [1997]. We also refer to Littlejohn [1988] for an interpretation of the Maslov index in terms of Berry's phase. We also will not attempt to make every step absolutely rigorous. See Eckmann and Seneor [1976] for details.

We begin with the ***one–dimensional Schrödinger equation***. Let $V : \mathbb{R} \to \mathbb{R}$ be a given potential, let $\psi : \mathbb{R} \to \mathbb{C}$ be a wave function, and let $E, \hbar, m$ be constants (energy, Planck's constant, and mass, respectively). Consider the ***stationary Schrödinger equation***:

$$L\psi = E\psi, \tag{N7.A.1}$$

where

$$L\psi = -\frac{\hbar^2}{2m}\psi'' + V\psi \tag{N7.A.2}$$

and the time-independent Hamilton–Jacobi equation for the function $S : \mathbb{R} \to \mathbb{R}$:

$$\frac{1}{2m}(S')^2 + V = E. \qquad \text{(N7.A.3)}$$

In this one dimensional case, the Hamilton–Jacobi equation is related to Hamilton's equations

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}, \qquad \text{(N7.A.4)}$$

where $H(q,p) = p^2/2m + V(q)$, in a very simple way: if $S(q)$ satisfies the Hamilton–Jacobi equation, and if $\dot{q}(t) = p(t)/m$ and if $p = S'(q) \neq 0$, then $(q(t), p(t))$ satisfies Hamilton's equations and has energy E.

Two related central questions are:

1. How does one pass from classical objects to quantum objects? Here, "objects" can refer to the equations themselves, to solutions, or to properties of the equations or solutions.

2. In what sense are solutions of the Hamilton–Jacobi equation a limit of solutions of the Schrödinger equation as $\hbar \to 0$?[1]

Progress with these questions was made with the basic work of Weyl, Birkhoff, van Hove, and, among many others, Keller, Maslov, Souriau, and Kostant (see the preceding references for the literature citations). van Hove showed that there is no general quantization having all the properties one would want.[2] In studying question 2 using the WKB method, Keller and Maslov discovered the topological meaning of the corrected Bohr–Sommerfeld quantization rules. The invariant they discovered is commonly called the ***Maslov index***. (See Arnold [1967]). Our one–dimensional example will contain many of the features of the general case.

If S is a solution of (N7.A.3), we try to solve (N7.A.1) with

$$\psi = \exp(iS/\hbar). \qquad \text{(N7.A.5)}$$

Substitution of (N7.A.5) in (N7.A.2) gives

$$E\psi = L\psi + \frac{i\hbar}{2m}\psi S'' \qquad \text{(N7.A.6)}$$

[1]Of course Planck's constant is a constant and cannot *literally* tend to zero, any more than the velocity of light can tend to infinity. However, when $\hbar$ is small, compared to quantities of interest in classical mechanics, this is expressed by mathematically by taking the limit $\hbar \to 0$ or by letting related parameters tend to zero (see Littlejohn [1988] and de de Gosson [1997]).

[2]van Hove's theorem in $\mathbb{R}^n$ is proved in Abraham and Marsden [1978], §5.4. van Hove also found some positive results that were extended by Segal, Souriau and Kostant in a procedure now called ***prequantization.*** Recent references in this direction may be found in Gotay, Grundling, and Tuynman [1996].

by using (N7.A.3). Equation (N7.A.6) differs from (N7.A.1) by a term of order $\hbar$. Next, try

$$\psi = a \exp(iS/\hbar) \tag{N7.A.7}$$

for $a : \mathbb{R} \to \mathbb{R}$. Substituting this into (N7.A.2) and using the Hamilton–Jacobi equation, we get

$$L\psi = E\psi - \hbar \frac{i}{2m} \left(S''a + 2S'a' \right) \frac{\psi}{a} - \frac{\hbar^2}{2m} \frac{a''}{a} \psi$$

or

$$E\psi = L\psi + \hbar \frac{i}{2m} \left(S''a + 2S'a' \right) \frac{\psi}{a} + \frac{\hbar^2}{2m} \frac{a''}{a} \psi. \tag{N7.A.8}$$

This equation differs from (N7.A.1) by a term of order $\hbar^2$ if a satisfies the ***transport equation***

$$2a'S' + aS'' = 0, \tag{N7.A.9}$$

whose solution is $a = (\text{constant})/|S'|^{1/2}$. Thus, (N7.A.8) becomes

$$E\psi = L\psi + \frac{\hbar^2}{2m} \frac{a''}{a} \psi \tag{N7.A.10}$$

which differs from (N7.A.1) by a term of order $\hbar^2$. The idea is now to continue this process by writing

$$\psi = \left(\sum_{k=0}^{N} a_k (i\hbar)^k \right) \exp(iS/\hbar) \tag{N7.A.11}$$

for some functions $a_k : \mathbb{R} \to \mathbb{R}$ and requiring ψ to satisfy (N7.A.1) up to an error term of order $\hbar^{N+2}$. This procedure is usually called the ***WKB method*** (after G. Wentzel, H. A. Kramers, and L. Brillouin, although it goes back to Liouville, Green, and Lord Rayleigh).

Substituting (N7.A.11) into (N7.A.2) and using, as before, the Hamilton–Jacobi equation (N7.A.3) yields

$$\begin{aligned} L\psi = E\psi - \frac{\exp(iS/\hbar)}{2m} \Bigg[& (S''a_0 + 2S'a_0')i\hbar \\ & + i\hbar \sum_{k=1}^{N} (S''a_k + 2S'a_k' - a_{k-1}'')(i\hbar)^k + i^N a_N'' \hbar^{N+2} \Bigg]. \end{aligned} \tag{N7.A.12}$$

Imposing the ***transport equations***

$$S''a_k + 2S'a_k' - a_{k-1}'' = 0, \quad k = 0, 1, \ldots, N, \quad a_{-1} \equiv 0, \tag{N7.A.13}$$

which can be solved recursively, we see that (N7.A.12) reduces to

$$E\psi = L\psi + \frac{i^N \exp(iS/\hbar)}{2m} a_N'' \hbar^{N+2}. \tag{N7.A.14}$$

Thus, we have "solved" (N7.A.1) up to an error of order $\hbar^{N+2}$. Therefore, if we let $N \to \infty$ we have found an asymptotic solution

$$\psi \sim \left(\sum_{k=0}^{\infty} a_k \hbar^k \right) \exp(iS/\hbar) \tag{N7.A.15}$$

of (N7.A.1). *The key observation in this procedure is that once S is determined, the coefficients a_k are obtained recursively as solutions of linear ordinary differential equations*. The solutions are *a fortiori* only local since S given by (N7.A.3) is only local, as we shall see below.

Suppose the energy surface for the classical system has the form shown in Figure N7.A.1.

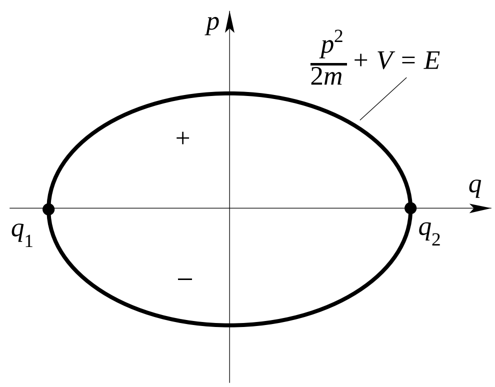

FIGURE N7.A.1. A sample classical energy surface.

There correspond two solutions of (N7.A.3):

$$S = \pm \int p(q)\, dq + C_{\pm}, \tag{N7.A.16}$$

where $p(q) = \sqrt{2m(E - V(q))}$, and $C_{\pm}$ are constants. Thus if ψ is given by (N7.A.11), or asymptotically by (N7.A.15), then the first transport equation (N7.A.9) for $k = 0$ yields

$$a_{0\pm} = \frac{d_{\pm}}{[2m(E - V(q))]^{1/4}} \tag{N7.A.17}$$

for some constants $d_{\pm}$. This expression diverges at q_1 and q_2 and becomes imaginary outside the interval $[q_1, q_2]$.

The subtlety of questions 1 and 2 centers on the multiple valuedness of S and the presence of the turning points at q_1 and q_2. To get around these difficulties there have been several approaches.

1. Use analytic continuation methods to avoid the turning points. This approach was developed by Zwaan.

2. Approximate the potential by a linear one near each turning point. Schrödinger's equation then yields an Airy function which is asymptotically matched by Bessel functions (Langer and Jeffreys).

3. Use a modified WKB method near the turning point and an asymptotic expansion (Maslov). We shall describe this method shortly.

There are other approaches too. For instance, Miller and Good [1953] effectively used area–preserving maps to deform Figure N7.A.1 into that for a harmonic oscillator. The same idea was used by Maslov [1965] for higher superpositions of such expressions.

To study the behavior near q_1 and q_2, we replace $\psi = a\exp(iS/\hbar)$ by a superposition of such expressions, that is, by

$$\psi(q) = \int_{-\infty}^{\infty} a(q,p)\exp(i\varphi(q,p)/\hbar)\,dp \qquad \text{(N7.A.18)}$$

where $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is positive. This integral is called an ***oscillatory function***; the theory of such integrals parallels that of Fourier integrals. Let us take $\varphi(q,p) = qp - T(p)$ for some real–valued function T defined in a neighborhood of the origin whose second derivative never vanishes, that is,

$$\psi(q) = \int_{-\infty}^{\infty} a(q,p)\exp\left[i\frac{pq - T(p)}{\hbar}\right]dp, \qquad \text{(N7.A.19)}$$

and try to solve (N7.A.1). A direct computation shows that

$$\begin{aligned} L\psi - E\psi = \int_{-\infty}^{\infty} &\left[a\left(\frac{p^2}{2m} + V(q) - E\right)\right. \\ &\left. - \left(\frac{i\hbar p}{m}\frac{\partial a}{\partial q} + \frac{\hbar^2}{2m}\frac{\partial^2 a}{\partial q^2}\right)\right]\exp\left[i\frac{pq - T(p)}{\hbar}\right]dp. \end{aligned} \qquad \text{(N7.A.20)}$$

To evaluate the right–hand side of (N7.A.20) asymptotically in $\hbar$ we need the following:

Theorem N7.A.1 (Stationary Phase Formula). *Let $a, \varphi : \mathbb{R} \to \mathbb{R}$ be C^∞ functions, φ having finitely many nondegenerate critical points. Then*

$$\int_{-\infty}^{\infty} a(x) \exp(i\varphi(x)/\hbar)\, dx$$
$$= \sqrt[2\pi\hbar]{\sum\nolimits_{\varphi'(y)=0}} \exp\left(\frac{i\pi}{4} \operatorname{sgn} \varphi''(y)\right) \frac{a(y)\exp(i\varphi(y)/\hbar)}{|\varphi''(y)|^{\frac{1}{2}}} + O(\hbar^{\frac{3}{2}}) \tag{N7.A.21}$$

where the sum is taken over all critical points y of φ. (Recall that a critical point y of φ is nondegenerate iff $\varphi''(y) \neq 0$.)

Proof (After Guillemin and Sternberg [1977]). Let $\{\chi_n\}$ be a C^∞ partition of unity on the real line, that is, each χ_n is $C^\infty, 0 \le \chi_n \le 1$, $\operatorname{supp} \chi_n =$ closure of $\{x \in \mathbb{R} \mid \chi_n(x) \neq 0\}$ is compact, each $x \in \mathbb{R}$ has a neighborhood intersecting only finitely many of $\operatorname{supp} \chi_n$, and $\sum_n \chi_n(x) = 1$ for each $x \in \mathbb{R}$. Since there are only finitely many critical points of φ, we can arrange the supports of χ_n such that each $\operatorname{supp} \chi_n$ contains at most one critical point of φ. Writing

$$\int_{-\infty}^{\infty} a(x) \exp\left[\frac{i\varphi(x)}{\hbar}\right] dx = \sum_n \int_{-\infty}^{\infty} \chi_n(x) a(x) \exp\left[\frac{i\varphi(x)}{\hbar}\right] dx,$$

we see that each integral on the right–hand side is a definite integral on $\operatorname{supp} \chi_n$ and that there are only a finite number of integrals that have overlapping domains of integration. Some of these integrals have domains which contain critical points of f, others do not.

We begin by studying those integrals that do not have a critical point of φ in their domain. Thus, we can assume that $\operatorname{supp} a$ is compact and that $\varphi' \neq 0$ on $\operatorname{supp} a$. Integrating by parts,

$$\begin{aligned}\int_{-\infty}^{\infty} a(x) \exp\left[\frac{i\varphi(x)}{\hbar}\right] dx &= \int_{-\infty}^{\infty} a(x) \frac{\hbar}{i\varphi'(x)} \frac{d}{dx}\left(\exp\left[\frac{i\varphi(x)}{\hbar}\right]\right) dx \\ &= i\hbar \int_{-\infty}^{\infty} \frac{d}{dx}\left(\frac{a(x)}{\varphi'(x)}\right) \exp\left[\frac{i\varphi(x)}{\hbar}\right] dx,\end{aligned}$$

which is an integral of the same type since $\frac{d}{dx}[a(x)/\varphi'(x)]$ is again C^∞ with compact support inside $\operatorname{supp} a$. Thus the procedure can be repeated any number of times yielding

$$\int_{-\infty}^{\infty} a(x) \exp\left[\frac{i\varphi(x)}{\hbar}\right] dx = O(\hbar^N)$$

for any $N \in \mathbb{N}$. Thus, to prove (N7.A.21), it suffices to establish it if $\operatorname{supp} a$ is compact and contains exactly one critical point x_0 of φ. This will be carried out in several steps.

Step 1 (Morse Lemma)**.** There is a change of variables $x \mapsto z$ such that

$$\varphi(x(z)) = \varphi(x_0) + \tfrac{1}{2}(\operatorname{sgn} \varphi''(x_0))(z - z_0)^2,$$

where $x(z_0) = x_0$.

To show this, we can clearly assume that

$$x_0 = 0, \varphi(x_0) = 0, \quad \varphi'(x_0) = 0, \quad \varphi''(x_0) \neq 0.$$

Write first

$$\varphi(x) = \int_0^1 \frac{d}{dt}\varphi(tx)\,dt = x\int_0^1 \varphi'(tx)\,dt = x\alpha(x),$$

where

$$\alpha(x) = \int_0^1 \varphi'(tx)\,dt$$

is again a C^∞ function. Since

$$\varphi'(x) = \alpha(x) + x\alpha'(x),$$

and $\varphi'(0) = \alpha(0) = 0$, the same argument shows that $\alpha(x) = x\beta(x)$ for some C^∞ function $\beta(x)$. Therefore,

$$\varphi(x) = x^2\beta(x) \quad \text{and} \quad \beta(x) = \int_0^1 \alpha'(tx)\,dt,$$

whence

$$\beta(0) = \alpha'(0) = \tfrac{1}{2}\varphi''(0).$$

Define $z(x) = \sqrt{2}|\beta(x)|^{\frac{1}{2}}x$ which is C^∞ in a neighborhood of 0, since $\beta(0) \neq 0$, and satisfies

$$z'(0) = \sqrt{2}|\beta(0)|^{\frac{1}{2}} \neq 0.$$

Therefore $x \mapsto z$ is a diffeomorphism in a neighborhood of 0 and in this neighborhood, suitably shrunk if necessary, $\beta(x)$ does not change sign. Thus

$$\varphi(x) = x^2\beta(x) = (\operatorname{sgn}\beta(x))x^2|\beta(x)| = \tfrac{1}{2}(\operatorname{sgn}\beta(0))z^2 = \tfrac{1}{2}(\operatorname{sgn}\varphi''(0))z^2.$$

Step 2. Performing the change of variables $x \mapsto z$ given in **Step 1** we get

$$\begin{aligned}
&\int_{-\infty}^{\infty} a(x)\exp(i\varphi(x)/\hbar)\,dx \\
&\qquad = \frac{a(x_0)\exp(i\varphi(x_0)/\hbar)}{|\varphi''(x_0)|^{\frac{1}{2}}}\int_{-\infty}^{\infty}\exp\left[\frac{\pm i(z-z_0)^2}{2\hbar}\right]dz \\
&\qquad\quad + \exp\left[\frac{i\varphi(x_0)}{\hbar}\right]\int_{-\infty}^{\infty}(z-z_0)\gamma(z)\exp\left[\frac{\pm i(z-z_0)^2}{2\hbar}\right]dz,
\end{aligned}$$

where $+$ or $-$ is taken in accordance with sgn $\varphi''(x_0)$ and $\gamma(z)$ is C^∞ with $(z-z_0)\gamma(z)$ bounded together with all its derivatives. (The bound for each derivative may be different.)

Indeed,

$$\int_{-\infty}^{\infty} a(x)\exp\left[\frac{i\varphi(x_0)}{\hbar}\right] dx = \int_{-\infty}^{\infty} a(x(z))\exp\left[\frac{i\varphi(x_0)}{\hbar} \pm \frac{i(z-z_0)^2}{2\hbar}\right]\left|\frac{dx}{dz}\right| dz$$

and note that

$$\frac{dx}{dz}(z_0) = \frac{1}{\sqrt{2}}|\beta(x_0)|^{\frac{1}{2}} = |\varphi''(x_0)|^{-\frac{1}{2}}$$

so that proceeding as in **Step 1** we can write

$$a(x(z))\left|\frac{dx}{dz}\right| - a(x_0)\left|\frac{dx}{dz}(z_0)\right| = (z-z_0)\gamma(z)$$

for some C^∞ function $\gamma(z)$ (z_0 denotes the point given by $x(z_0)=x_0$), that is,

$$a(x(z))\left|\frac{dx}{dz}\right| = a(x_0)\frac{1}{|\varphi''(x_0)|^{\frac{1}{2}}} + (z-z_0)\gamma(z).$$

Since

$$(z-z_0)\gamma(z) = (z-z_0)\int_0^1 \frac{d}{dt}\left(a(x(z_t)\left|\frac{dx}{dz}(z_t)\right|\right) dt,$$

where $z_t = tz + (1-t)z_0$, we see that on its domain of definition $\gamma(z)$ is smooth and has itself and all its derivatives bounded because $a(x)$ has compact support.

To show that each integral in **Step 2** is well defined, we prove:

Step 3. Let $h(z)$ be a C^2 function of a real variable such that the three functions $|h(z)|, |h'(z)|, |h''(z)|$ are all bounded by $M > 0$. If $\lambda \in \mathbb{C}$, the integral

$$\int_{-\infty}^{\infty} e^{-\lambda z^2/2} h(z)\, dz$$

is uniformly convergent for $\operatorname{Re}\lambda \geq 0, |\lambda| \geq 1$, bounded by a constant depending on M only, holomorphic for $\operatorname{Re}\lambda > 0$, and continuous for $\operatorname{Re}\lambda \geq 0$.

It suffices to prove this for $\int_0^\infty e^{-\lambda z^2/2} h(z)\, dz$ for then, by changing variables $z \mapsto -z$, the same result holds for the integral from $-\infty$ to 0 and hence for the sum. Let $0 < A < B$. Then

$$\begin{aligned}
&\int_A^B e^{-\lambda z^2/2} h(z)\,dz \\
&\quad = -\lambda^{-1} \int_A^B \frac{1}{z}(e^{-\lambda z^2/2})' h(z)\,dz \\
&\quad = -(\lambda z)^{-1} e^{-\lambda z^2/2} h(z) \ \Big|_A^B + \lambda^{-1} \int_A^B e^{-\lambda z^2/2} \left(\frac{h(z)}{z}\right)' dz \\
&\quad = -(\lambda z)^{-1} e^{-\lambda z^2/2} h(z) \ \Big|_A^B - \lambda^{-2} \int_A^B (e^{-\lambda z^2/2})' \frac{1}{z} \left(\frac{h(z)}{z}\right)' dz \\
&\quad = -(\lambda z)^{-1} e^{-\lambda z^2/2} \left(h(z) + \lambda^{-1} \left(\frac{h(z)}{z}\right)' \right) \Bigg|_A^B \\
&\qquad + \lambda^{-2} \int_A^B e^{-\lambda z^2/2} \left[\frac{1}{z} \left(\frac{h(z)}{z}\right)' \right]' dz.
\end{aligned}$$

The first term tends to zero as $A \to \infty$ by boundedness of $|h|, |h'|$ if Re $\lambda \geq 0, |\lambda| \geq 1$. The integral in the second term can also be bounded in absolute value for the same range of λ by a constant depending only on M since $|h''|$ is bounded. In particular, the integral

$$\int_0^\infty e^{-\lambda z^2/2} h(z)\,dz$$

is uniformly convergent.

Arguing in the same manner for the λ–derivative, we conclude that the integral is holomorphic for Re $\lambda > 0$. Similarly one shows continuity for Re $\lambda \geq 0$.

Step 4.

$$\int_{-\infty}^\infty \exp\left[\frac{\pm i(z - z_0)^2}{2\hbar}\right] dz = \sqrt{2\pi\hbar}\, e^{\pm \pi i/4}.$$

From the previous step it follows that this integral exists, by taking $\lambda = \mp i$ and $h(z) = 1$. Moreover, the classical formula

$$\int_{-\infty}^\infty e^{-u^2/2}\,du = \sqrt{2\pi}$$

implies that for real positive λ we have

$$\int_{-\infty}^\infty e^{-\lambda u^2/2}\,du = \sqrt{2\pi/\lambda}.$$

By analytically continuing both sides for Re $\lambda > 0$, the same formula holds for complex λ in the right half–plane. Now let $\lambda \to \mp i/\hbar$ to obtain

$$\int_{-\infty}^{\infty} \exp\left[\frac{\pm i(z-z_0)^2}{2\hbar}\right] dz = \sqrt{2\pi}\sqrt{\hbar}\sqrt{\exp\left(\pm\frac{\pi}{2}i\right)} = \sqrt{2\pi\hbar}\exp\left(\pm\frac{\pi i}{4}\right).$$

Step 5. The second integral in **Step 2** is $O(\hbar^{3/2})$.

Indeed, the integral exists by **Step 3** and

$$\begin{aligned}
&\int_{-\infty}^{\infty} (z-z_0)\gamma(z)\exp\left[\frac{\pm i(z-z_0)^2}{2\hbar}\right] dz \\
&= \int_{-\infty}^{\infty} z\gamma(z+z_0)\exp\left(\frac{\pm iz^2}{2\hbar}\right) dz \\
&= \pm\frac{\hbar}{i}\int_{-\infty}^{\infty} \left[\exp\left(\frac{\pm iz^2}{2\hbar}\right)\right]' \gamma(z+z_0)\, dz \\
&= \pm i\hbar\int_{-\infty}^{\infty} \gamma'(z+z_0)\exp\left(\frac{\pm iz^2}{2\hbar}\right) dz.
\end{aligned}$$

The boundary terms vanish if γ vanishes sufficiently fast at ∞. This integral has exactly the form of the original integral and therefore can be written as a sum of two integrals, the first of order $O(\hbar^{1/2})$ by **Step 4** and the second $\hbar$ times again an integral of the same type. Thus this integral is of order $\hbar\hbar^{1/2} = \hbar^{3/2}$.

From **Steps 2, 4,** and **5**, we conclude that if there is a single critical point x_0 of φ in $\operatorname{supp} a$ we get

$$\begin{aligned}
&\int_{-\infty}^{\infty} a(x)\exp\left[\frac{i\varphi(x)}{\hbar}\right] dx \\
&= \sqrt{2\pi\hbar}\exp\left(\pm i\frac{\pi}{4}\right)\frac{a(x_0)\exp(i\varphi(x_0)/\hbar)}{|\varphi''(x_0)|^{1/2}} + O(\hbar^{3/2}). \qquad \blacksquare
\end{aligned}$$

The previous proof shows that the same formula holds if all functions depend smoothly on additional parameters. In particular, we shall use the following expression in analyzing the right–hand side of (N7.A.19):

$$\begin{aligned}
&\int_{-\infty}^{\infty} c(q,p)\exp\left[i\frac{f(q,p)}{\hbar}\right] dp \\
&= \sqrt{2\pi\hbar}\sum_{f_p=0}\exp\left(\frac{i\pi}{4}\operatorname{sgn} f_{pp}\right)\frac{c(q,p)\exp(if/\hbar)}{|f_{pp}|^{1/2}} + O(\hbar^{3/2}), \qquad \text{(N7.A.22)}
\end{aligned}$$

where the sum is over all p such that $f_p = \partial f/\partial p$ vanishes; these critical points are assumed to be finite in number and nondegenerate, that is, $f_{pp} = \partial^2 f/\partial p^2 \neq 0$.

Applying (N7.A.22) to (N7.A.20) gives

$$\begin{aligned} L\psi - E\psi \\ &= \sqrt{2\pi\hbar} \sum_{q=T'(p)} \frac{\exp[-i\pi \operatorname{sgn} T''(p)/4]}{|T''(p)|^{1/2}} a(q,p) \left(\frac{p^2}{2m} + V(q) - E \right) \\ &\qquad \times \exp\left[i \frac{pq - T(p)}{\hbar} \right] + O(\hbar^{3/2}), \end{aligned} \tag{N7.A.23}$$

provided the number of critical points in p of the q-dependent function $f(q,p) = qp - T(p)$ is finite and all these p-critical points are nondegenerate. By assumption, T'' never vanishes for p near zero and thus $T'(p)$ is either strictly increasing or strictly decreasing. Thus, for a fixed q, there is exactly one p such that $q = T'(p)$, that is, (N7.A.23) reads

$$\begin{aligned} L\psi - E\psi &= \sqrt{2\pi\hbar} \frac{\exp[-i\pi \operatorname{sgn} T''(p)/4]}{|T''(p)|^{1/2}} a(T'(p), p) \\ &\times \left(\frac{p^2}{2m} + V(T'(p)) - E \right) \exp\left[i \frac{pT'(p) - T(p)}{\hbar} \right] + O\left(\hbar^{3/2} \right). \end{aligned}$$

Now we require that $L\psi - E\psi = O(\hbar^{3/2})$, which forces the first term to vanish (since $a(q,p)$ is not the zero function), that is,

$$\frac{p^2}{2m} + V(T'(p)) = E. \tag{N7.A.24}$$

Thus the graph of $q = T'(p)$ (as a function of p) is contained in the energy surface. Equation (N7.A.24) is the Hamilton–Jacobi equation in the variable p, which is approximated near the turning points q_1 and q_2.

Applying (N7.A.22) to formula (N7.A.19) gives

$$\begin{aligned} \psi(q) &= \sqrt{2\pi\hbar} \sum_{q=T'(p)} \frac{1}{|T''(p)|^{1/2}} \exp[-i\pi \operatorname{sgn} T''(p)/4] \\ &\qquad \times \exp\left[i \frac{pq - T(p)}{\hbar} \right] a(q,p) + O(\hbar^{3/2}) \\ &= \sqrt{2\pi\hbar} \frac{1}{|T''(p)|^{1/2}} \exp[-i\pi \operatorname{sgn} T''(p)/4] \\ &\qquad \times \exp\left[i \frac{pT'(p) - T(p)}{\hbar} \right] a(T'(p), p) + O(\hbar^{3/2}) \\ &= O(\hbar^{1/2}). \end{aligned} \tag{N7.A.25}$$

We now seek to represent ψ near q_1 and q_2 using functions T_1 and T_2 given by (N7.A.24) and seek to represent ψ on the $\pm$ portions in the form (N7.A.19). We are, in effect, using a superposition of two WKB approximations.

Notice that if $q - T'(p) = 0$, as above, then

$$\frac{d}{dq}(pq - T(p)) = p + \frac{dp}{dq}q - T'(p)\frac{dp}{dq} = p, \tag{N7.A.26}$$

so both S and $pq - T(p)$ are given by integrating p with respect to q; that is, they are both *actions*.

Since $q = T'(p)$ along the energy curve in Figure N7.A.1, we see that $T''(p) > 0$ on the $+$ side and $T''(p) < 0$ on the $-$ side of the p–axis. Thus the term

$$e^{-i\pi \operatorname{sgn} T''(p)/4} \tag{N7.A.27}$$

in (N7.A.25) jumps, or suffers a phase shift, as p crosses the q–axis. In Figure N7.A.2 we show the different regions and functions being considered.

So now we have obtained ψ on four different regions: The upper and lower part of the energy surface and the parts around the two turning points $(q_1, 0), (q_2, 0)$; see Figure N7.A.2. The structure of this function is that of a product of an amplitude times an exponential plus higher-order terms. We shall require that they all match on the overlaps at first order. Since there are constants of integration in these formulae (as in (N7.A.17), for example), matching at points A, B, and C determines all the constants. Thus, the consistency condition is the match of these solutions at the point D. This will happen only if the phases in (N7.A.25) match.

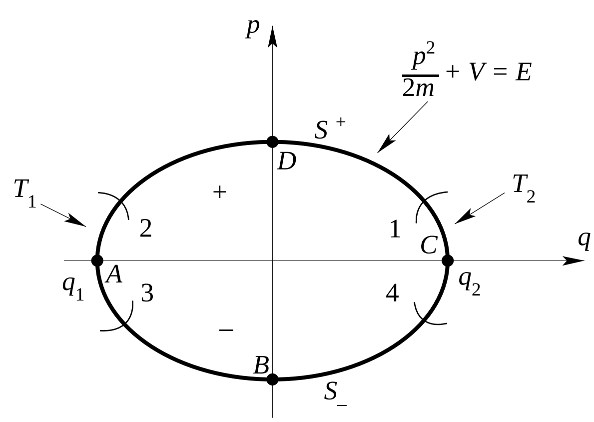

FIGURE N7.A.2. Matching phases.

The phase changes in the exponentials $\exp(iS/\hbar)$ and $\exp[i(pq-T(p))/\hbar]$ are given by

$$\frac{1}{\hbar}\oint p\,dq, \tag{N7.A.28}$$

since both S and $pq-T(p)$ are given by integrating p and the line integral is over the energy curve. On the other hand, the phase change due to the term (N7.A.27) is

$$-2\times\left[\frac{\pi}{4}-\left(-\frac{\pi}{4}\right)\right]=-\pi, \tag{N7.A.29}$$

so the consistency condition is

$$\frac{1}{\hbar}\oint p\,dq-\pi=2\pi n, \text{i. e.}, \oint\frac{p\,dq}{2\pi\hbar}=n+\tfrac{1}{2}. \tag{N7.A.30}$$

The $\frac{1}{2}$ is the correction to the Bohr–Sommerfeld rules which one sees, for example, in the harmonic oscillator solution. Equation (N7.A.30) is the ***quantization condition***. Its generalization to arbitrary manifolds reads

$$\frac{1}{2\pi\hbar}\oint_\gamma p_i\,dq^i-\tfrac{1}{4}I_\gamma=\text{integer}, \tag{N7.A.31}$$

where I_γ is the ***Maslov index*** of a closed curve γ. This topological invariant is thus arrived at via the WKB method. To understand it in higher dimensions requires a lengthy excursion into the theory of Lagrangian submanifolds. However, our simplified example shows that starting with a study of the asymptotic limit $\hbar\to 0$, one is led to quantization conditions; that is, questions 1 and 2, formulated at the beginning of this section are intimately related.

The overall aims of quantization and geometric asymptotics become clearer if one has in mind some of the classical–quantum correspondences. To this end, we present the table below (see Slawianowski [1971]). The basic classical object is a symplectic manifold (T^*Q,Ω) and the quantum object is the intrinsic Hilbert space $\mathcal{H}=L^2(Q)$ of half densities on Q. The dictionary sets up a correspondence between operations on each.

Classical Mechanics	Quantum Mechanics
immersed Lagrangian manifold $\Lambda \to (T^*Q, \Omega)$	element of $L^2(Q)$ or $\mathcal{D}'(Q)$
$\Lambda =$ graph of $\mathbf{d}S$	$\psi = \exp(iS/\hbar)$
multiplication by (-1) on fibers	complex conjugation
T^*Q	Hilbert space
$(T^*Q, -\Omega)$	dual space
Cartesian product	tensor product
disjoint union	direct product
Lagrangian manifold $\Omega \subset (T^*Q, \Omega_Q) \times (T^*R, -\Omega_R)$	(unbounded) operator from $L^2(R)$ to $L^2(Q)$
composition of canonical relations	composition of operators
graphs of canonical relations	unitary operators
Hamilton–Jacobi equation	Schrödinger equation
coisotropic submanifold $C \subset T^*Q$	involutive system of linear differential equations
reduced space C/C^{Ω}	solution space
reduction of Lagrangian submanifolds	projection onto solution space
symplectic action (Hamiltonian G–space)	unitary representation
coadjoint orbits (homogeneous Hamiltonian G–spaces)	irreducible representations
reduction of phase space by a symmetry group	multiplicities of irreducibles
momentum mapping	associated representation of the group algebra
polarization	complete set of observables
special symplectic structure	representation of a complete set of observables
change of special symplectic structure (Tulczyjew [1977])	Fourier integral operator

N9
Lie Groups

Lie groups is a large subject and Chapter 9 of the text as well as this supplement cover only a part of the subject.

N9.A Automatic Smoothness

We begin with a proof of Proposition 9.1.4 in the text. We recall the statement.

Proposition 9.1.4. *Let $\gamma : \mathbb{R} \to G$ be a continuous one–parameter subgroup of a Lie group G. Then γ is smooth and hence $\gamma(t) = \exp(\xi t)$ for some $\xi \in g$.*

Proof. It suffices to prove smoothness of γ for $|t| < \epsilon$ for some small $\epsilon > 0$. Indeed, $\gamma(t+s) = \gamma(t)\gamma(s)$ shows that if $|s| < \epsilon$ then γ is smooth in an ϵ–neighborhood of each t; thus $\gamma(t)$ is smooth in a 2ϵ–neighborhood of zero. Repeating, we see γ is smooth everywhere.

To show that γ is smooth for $|t|$ small, the strategy is to show that it coincides with $\exp(t\zeta)$ for some $\zeta \in \mathfrak{g}$ and for small t. The strategy of the proof is to show this equality for small rational numbers t using algebraic properties of γ and exp and then to invoke continuity for a limiting argument.

To carry this strategy out, fix some $n \in \mathbb{N}$ and let B_R be the open ball of radius R about the origin in $\mathfrak{g}$ on which exp is a diffeomorphism. By continuity of γ, there is some $\epsilon > 0$ such that $\gamma(t) \in \exp\left(B_{R/2}\right)$ for all

$|t| < \epsilon$. Fix $s > 0, s < \epsilon$ and define $\eta \in B_{R/2}$ by $\exp \eta = \gamma(s)$. Similarly, since $s/n < \epsilon$, define $\xi_n \in B_{R/2}$ by $\gamma(s/n) = \exp \xi_n$ and note that

$$\exp (n\xi_n) = (\exp \xi_n)^n = \gamma (s/n)^n = \gamma(s) = \exp \eta$$

which would imply, by bijectivity of exp on B_R, that $n\xi_n = \eta$, if we knew in advance that $n\xi_n \in B_R$. To see this, we begin by observing that $2\xi_n \in B_R$ and that

$$\exp (2\xi_n) = (\exp \xi_n)^2 = \gamma(s/n)^2 = \gamma(2s/n) \in \exp(B_{R/2})$$

since $2s/n < \epsilon$ if $2 \leq n$. Thus, $2\xi \in B_{R/2}$. Repeating this argument for $3\xi, 4\xi, \ldots$, we conclude that $n\xi \in B_{R/2}$ and so $n\xi_n = \eta$.

Let now $k \in \mathbb{N}, 1 \leq k \leq n$. Then

$$\gamma(ks/n) = \gamma(s/n)^k = \exp(\xi_n)^k = \exp(k\xi_n) = \exp(k\eta/n)$$

since $\xi_n = \eta/n$. Also,

$$\gamma(-ks/n) = \gamma(ks/n)^{-1} = \exp(k\eta/n)^{-1} = \exp(-k\eta/n)$$

which shows that for any rational number $q, |q| \leq 1$, we have

$$\gamma(qs) = \exp(q\eta).$$

Now let q_n be a sequence of rational numbers convergent to t/s for $|t| \leq s < \epsilon$. Continuity for γ and exp imply then that $\gamma(q_n s) \to \gamma(t)$ and $\exp(q_n \eta) \to \exp(t\eta/s)$ as $n \to \infty$. We conclude that $\gamma(t) = \exp(t\zeta)$ where $\zeta = \eta/s$, for all $|t| \leq s$. ■

Next we generalize this result to Theorem 9.1.9 of the text. Again, we recall the statement.

Theorem 9.1.9. *Let $f : G \to H$ be a continuous homomorphism of finite dimensional Lie groups. Then f is smooth.*

Notice that if $G = \mathbb{R}$, this statement gives the preceding proposition. In fact the strategy is to use the special case 9.1.4 to prove the more general one 9.1.9.

Proof. Note that if $\xi_1, \ldots, \xi_n \in \mathfrak{g}$ is a basis, then $\psi : \mathbb{R}^n \to G, n = \dim G$, given by

$$\psi(x^1, \ldots, x^n) = \exp(x^1\xi_1) \cdots \exp(x^n\xi_n)$$

has derivative at the origin equal to the identity map (if we identify $\mathfrak{g}$ with $\mathbb{R}^n$ via the chosen basis). Therefore, one can find open neighborhoods V of e in G and U of the origin in $\mathbb{R}^n$ such that $\psi|U : U \to V$ is a diffeomorphism. Let $\varphi : V \to U$ be given by $\varphi = (\psi|V)^{-1}$. Then (V, φ)

is a chart at the identity, called ***exponential chart of the second kind*** (as opposed to the ***exponential chart of the first kind*** given by the inverse of the exponential map on a neighborhood of the identity). However, $t \mapsto f(\exp t\xi_i)$ is a continuous one–parameter subgroup of H and is hence smooth by Proposition 9.1.4. Therefore $f \circ \psi$ is smooth which implies that $f|V = (f \circ \psi) \circ \varphi : V \to H$ is smooth. Thus f is smooth in a chart around $e \in G$.

Since an atlas of G can be obtained by left translating this chart and since

$$f|L_g(V) = L_{f(g)} \circ f \circ L_{g^{-1}}|L_g(V),$$

because f is a homomorphism, we see that f is smooth on $Lg(V)$ and hence on G. Here, $L_{f(g)}$, as usual, denotes left translation by $f(g)$ in H and L_g denotes left translation by g in G. ■

N9.B Abelian Lie Groups

In this section we prove Theorem 9.1.11, the main structure theorem for Abelian Lie groups.

Theorem 9.1.11. *Every connected Abelian n–dimensional Lie group G is isomorphic to a cylinder, that is, to $\mathbb{T}^k \times \mathbb{R}^{n-k}$ for some $k = 0, 1, \cdots, n$.*

Proof. Since G is Abelian, the map $t \mapsto (\exp t\xi)(\exp t\eta)$ is a one–parameter subgroup of G for any $\xi, \eta \in \mathfrak{g}$. The derivative at $t = 0$ of this one–parameter subgroup is $\xi + \eta$ and so by uniqueness, we conclude that

$$(\exp t\xi)(\exp t\eta) = \exp t(\xi + \eta).$$

In particular, setting $t = 1$, we see that $\exp : \mathfrak{g} \to G$ is a Lie group homomorphism. In addition, since G is connected, it is generated by an open neighborhood of the identity. Since exp is a local diffeomorphism around the origin, G is generated by $\exp(\mathfrak{g})$ and hence $\exp(\mathfrak{g}) = G$ because exp is a homomorphism. Therefore, $\exp : \mathfrak{g} \to G$ is a surjective Lie group homomorphism that is also a local diffeomorphism. Consequently, its kernel is a zero dimensional submanifold of $\mathfrak{g}$ and thus is a discrete subgroup of $\mathfrak{g}$. Consequently, $\mathfrak{g}/\ker\ \exp$ is isomorphic to G as groups and diffeomorphic to G as manifolds, by working in a small neighborhood of the origin in $\mathfrak{g}$ where exp is a diffeomorphism with an open neighborhood of the identity element in G. Thus $\mathfrak{g}/\ker\ \exp$ and G are isomorphic Lie groups. The theorem is then a consequence of the following lemma. ■

Lemma N9.B.1. *Any closed discrete subgroup C of $\mathbb{R}^n$ is of the form*

$$C = \left\{ \sum_{i=1}^{k} k_i \mathbf{e}_i \mid k_i \in \mathbb{Z} \right\},$$

where $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ is a set of linearly independent vectors of $\mathfrak{g}$.

Proof. If $C = \{0\}$, there is nothing to prove. If not, there is some $\mathbf{e}_1 \neq 0, \mathbf{e}_1 \in C$. Since C is discrete in $\mathbb{R}^n$, there is an open ball B centered at the origin in $\mathbb{R}^n$ such that $C \cap B = \{0\}$. Thus $\mathbf{e}_1$ can be chosen such that $\|\mathbf{e}_1\| \leq \|\mathbf{c}\|$ for all $\mathbf{c} \in C$. Moreover, $\operatorname{span}\{\mathbf{e}_1\} \cap C = \mathbb{Z}\mathbf{e}_1$. Indeed, if $t\mathbf{e}_1 \in C$ and $[t]$ denotes the integer part of t, that is, $[t] \leq t < [t] + 1, [t] \in \mathbb{Z}$, then

$$t\mathbf{e}_1 - [t]\,\mathbf{e}_1 \in C \quad \text{and} \quad \|(t - [t])\mathbf{e}_1\| < \|\mathbf{e}_1\|,$$

which implies that $t = [t] \in \mathbb{Z}$.

Now consider the projection $\rho : \mathbb{R}^n \to \mathbb{R}^n/\operatorname{span}\{\mathbf{e}_1\} \cong \mathbb{R}^{n-1}$. Since ρ is an open map, it follows that $\pi(c)$ is a discrete subgroup of $\mathbb{R}^n/\operatorname{span}\{\mathbf{e}_1\}$. Inductively, we can find linearly independent vectors $\rho(\mathbf{e}_2), \ldots, \rho(\mathbf{e}_k)$ in $\mathbb{R}^n/\operatorname{span}\{\mathbf{e}_1\}$ such that every element of $\rho(c)$ is a linear combination of $\rho(\mathbf{e}_2), \ldots, \rho(\mathbf{e}_k)$ with coefficients in $\mathbb{Z}$.

It follows that $\mathbf{e}_1, \ldots, \mathbf{e}_k$ satisfy the conditions of the lemma. Indeed, since $\rho(\mathbf{e}_2) \ldots \rho(\mathbf{e}_k)$ are linearly independent in $\mathbb{R}^n/\operatorname{span}\{\mathbf{e}_1\}$, it follows that $\mathbf{e}_1, \ldots, \mathbf{e}_k$ are linearly independent in $\mathbb{R}^n$. Moreover, if

$$c = t, \mathbf{e}_1 + t_2\mathbf{e}_2 + \cdots + t_k\mathbf{e}_k \in C,$$

then

$$\rho(c) = t_2\rho(\mathbf{e}_2) + \cdots + t_k\rho(\mathbf{e}_k) \in \rho(c)$$

so by the inductive hypothesis, $t_2, \ldots, t_k \in \mathbb{Z}$. But then

$$t_1\mathbf{e}_1 = c - t_2\mathbf{e}_2 - \cdots - t_k\mathbf{e}_k \in C$$

and hence $t_1 \in \mathbb{Z}$. ■

N9.C Lie Subgroups

This section is devoted to the proof of the following theorem stated in the text.

Theorem 9.1.14. *If H is a closed subgroup of a finite dimensional Lie group G, then H is a regular Lie subgroup. Conversely, if H is a regular Lie subgroup, then H is closed.*

Proof. Assume that H is a closed subgroup of G. The proof given below that H is a regular Lie subgroup of G is due to Adams [1969] and consists of four steps. We shall fix once and for all an inner product on $\mathfrak{g}$ and denote the associated norm by $\| \cdot \|$.

Step 1. Assume that $\{\zeta_n\}$ is a sequence in $\mathfrak{g}$ such that $\zeta_n \neq 0$ for all $n \in \mathbb{N}, \zeta_n \to 0$, and $\zeta_n/\|\zeta_n\| \to \zeta \in \mathfrak{g}$ as $n \to \infty$. If $\exp \zeta_n \in H$, then we will show that $\exp t\zeta \in H$ for all $t \in \mathbb{R}$.

To see this assume first that $t > 0$ and let $m_n = [t/\|\zeta_n\|] \in \mathbb{N}$ be the integer part of $t/\|\zeta_n\|$, that is,

$$0 \leq m_n \leq \frac{t}{\|\zeta_n\|} < m_n + 1.$$

Then we have

$$m_n\|\zeta_n\| - t \leq 0 < m_n\|\zeta_n\| - t + \|\zeta_n\|,$$

which implies that

$$0 \leq t - m_n\|\zeta_n\| < \|\zeta_n\|,$$

whence $m_n\|\zeta_n\| \to t$ as $n \to \infty$. Therefore, $m_n\zeta_n \to t\zeta$ as $n \to \infty$, and hence

$$(\exp \zeta_n)^{m_n} = \exp (m_n\zeta_n) \to \exp t\zeta, \quad \text{as} \quad n \to \infty.$$

Since $\exp \zeta_n \in H$ by hypothesis and H is closed, this implies that $\exp t\zeta \in H$ for all $t > 0$. If $t < 0$, we have $\exp t\zeta = [\exp(-t\zeta)]^{-1} \in H$, since $\exp(-t\zeta) \in H$ by what we just proved.

Step 2. Define $\mathfrak{h} = \{\xi \in \mathfrak{g} \mid \exp t\xi \in \mathbb{R} \text{ for all } t \in \mathbb{R}\}$. We will show that $\mathfrak{h}$ is a linear subspace of $\mathfrak{g}$.

It is clear that if $\lambda \in \mathbb{R}$ and $\xi \in \mathfrak{h}$ then $\lambda\xi \in \mathfrak{h}$. Next, let $\xi, \eta \in \mathfrak{h}$ and assume that $\xi + \eta \neq 0$. If $t \in \mathbb{R}$ is sufficiently small, since exp is a diffeomorphism of a neighborhood of zero in $\mathfrak{g}$ with a neighborhood of e in G, it follows that

$$(\exp t\xi)(\exp t\eta) = \exp(f(t))$$

for some $f(t) \in \mathfrak{g}$ satisfying $f(0) = 0$ and f is smooth around 0. Since

$$\xi + \eta = \left.\frac{d}{dt}\right|_{t=0} (\exp t\xi)(\exp t\eta) = \left.\frac{d}{dt}\right|_{t=0} \exp(f(t)),$$

it follows that $f(t)/t \to \xi + \eta$ as $t \to 0$. Since $f(t) \to 0$ as $t \to 0$, letting $\zeta_n = f(1/n)$ and $\zeta = (\xi + \eta)/\|\xi + \eta\|$ we see that the hypotheses of Step 1 hold and hence we conclude that $\exp t\xi \in H$ for all $t \in \mathbb{R}$. Therefore $\zeta \in \mathfrak{h}$ which implies that $\xi + \eta \in \mathfrak{h}$.

Step 3. Let $\mathfrak{h}^\perp$ be the orthogonal complement to $\mathfrak{h}$ in $\mathfrak{g}$ and define the map

$$\varphi : \mathfrak{h}^\perp \oplus \mathfrak{h} \to G \quad \text{by} \quad \varphi(\xi, \eta) = (\exp \xi)(\exp \eta),$$

for $\xi \in \mathfrak{h}^\perp, \eta \in \mathfrak{h}$. Then we will show that there are neighborhoods of the origin $U' \subset \mathfrak{h}^\perp, U \subset \mathfrak{h}$, and V of e in G, such that

(i) $\varphi : U' \times U \to V$ is a diffeomorphism,

(ii) $V \cap H = \exp(U)$.

It follows from this that $\exp : \{0\} \times U \to \exp(U) \subset H$ is bijective.

To see (i) and (ii), note that $T_{(0,0)} \exp$ equals the identity map of $\mathfrak{g}$ and hence φ is a local diffeomorphism around the origin in $\mathfrak{g}$. In particular, there are balls $B' \subset \mathfrak{h}^{\perp}$ and $B \subset \mathfrak{h}$, both of radius r, centered at the origin such that both the map

$$\varphi : B' \times B \to \varphi\left(B' \times B\right) = \exp\left(B'\right)\exp\left(B\right)$$

and the map

$$\exp : B' \times B \to \exp\left(B' \times B\right)$$

are diffeomorphisms. Let B'_n, B_n denote the balls of radius r/n centered at the origin in $\mathfrak{h}^{\perp}$ and $\mathfrak{h}$ respectively. We claim that for some n large enough,

$$\exp\left(B_n\right) = \varphi\left(B'_n \times B_n\right) \cap H = \left[\exp\left(B'_n\right)\exp\left(B_n\right)\right] \cap H.$$

The definition of $\mathfrak{h}$ immediately implies that $\exp\left(B_n\right) \subset H$ and hence that

$$\exp\left(B_n\right) \subset \left[\exp\left(B'_n\right)\exp\left(B_n\right)\right] \cap H.$$

To show the converse, assume the contrary, namely that for any $n \in \mathbb{N}$ there exists a $\xi_n \in B'_n$ such that $\exp \xi_n \in H$ but $\xi_n \neq 0$. Clearly, $\xi_n \to 0$ as $n \to \infty$ and by compactness of the unit sphere, $\xi_n / \|\xi_n\|$ has a convergent subsequence $\xi_{n_k}/\|\xi_{n_k}\| \to \xi \in \mathfrak{h}^{\perp}, \|\xi\| = 1$. Step 1 then implies that $\exp t\xi \in H$ for all $t \in \mathbb{R}$, that is, $\xi \in \mathfrak{h}$, by definition of $\mathfrak{h}$. Thus $\xi \in \mathfrak{h}^{\perp} \cap \mathfrak{h} = \{0\}$ which contradicts $\|\xi\| = 1$.

Therefore, if n is large enough,

$$\exp(B_n) = [\exp(B'_n)\exp(B_n)] \cap H$$

and so (i) and (ii) are proved, by taking $U' = B'_n, U = B_n$, and $V = \varphi\left(U' \times U\right) = \exp\left(U'\right)\exp U$.

Step 4. Define $\psi : \exp(U) = V \cap H \to \{0\} \times U$ to be the inverse of the bijective map in Step 3. Taking as a chart around e in G the inverse of φ on $\exp(U' \times U)$, that is, we consider the chart (V, φ^{-1}) at e in G, Step 3 guarantees that $\varphi^{-1}(V \cap H) = \{0\} \times U$, that is, (V, φ^{-1}) has the submanifold property relative to H. Moreover, the induced chart at e on H is (W, ψ). Now we left translate (V, φ^{-1}) to any point in G. In particular, the left translated chart

$$V_h := L_h(V),\, \varphi_h^{-1} : V_h \to U' \times U,\, \varphi_h^{-1}\left(k\right) = \varphi^{-1}\left(h^{-1}k\right),$$

has the submanifold relative to H inducing the chart

$$(W_h, \psi_h), W_h := L_h(W) = k\exp(U), \psi_h(k) = \psi(h^{-1}k),$$

on H. Thus, H is a smooth submanifold of G.

Finally, since the group operations in H are the restrictions of those in G which are smooth in the manifold structure of G, it follows that they are smooth in the manifold structure of H, since H is a smooth submanifold on G.

Conversely, assume that H is a regular Lie subgroup of G. We shall prove that H is closed. Let $\{h_n\}$ be a sequence in H convergent in G to some element $h \in G$. Since H is a submanifold of G, there is a chart (V, χ) at e in G with the properties

$$\chi : V \to U' \times U,\ U' \subset \mathfrak{h}^{\perp},\ U \subset \mathfrak{h}$$

open balls at the origin and $\chi(V \cap H) = \{0\} \times U$. For all $n \geq N$, $h^{-1}h_n \in V$. On the other hand $h_N^{-1}h_n \in H$, so

$$\chi\left(h_N^{-1}h_n\right) \to \chi\left(h_N^{-1}h\right) \in \{0\} \times U,$$

since we can always chose V such that $V^{-1} = V$ and $VV \subset V$. Therefore $h_N^{-1}h \in H$, since $\chi : V \cap H \to \{0\} \times U$ is a diffeomorphism. Since $h_N \in H$, this implies $h \in H$ and so H is closed in G. ■

N9.D Lie's Third Fundamental Theorem

Recall the statement of this result.

Theorem 9.1.15. *Let G be a Lie group with Lie algebra $\mathfrak{g}$ and let $\mathfrak{h}$ be a Lie subalgebra of $\mathfrak{g}$. Then there exists a unique connected (immersed) Lie subgroup H of G whose Lie algebra is $\mathfrak{h}$.*

Proof. Define the smooth vector subbundle $\tilde{\mathfrak{h}} \subset TG$ by left translating $\mathfrak{h}$ to any point of G, that is, the fiber $\tilde{\mathfrak{h}}_g$ at g equals $T_eL_g(\mathfrak{h})$. We prove now that $\tilde{\mathfrak{h}}$ is an involutive subbundle.

Let X, Y be vector fields on G with values in $\tilde{\mathfrak{h}}$, that is, they are sections of $\tilde{\mathfrak{h}}$. We will show that $[X, Y]$ is also a section of $\tilde{\mathfrak{h}}$. Fix $g \in G$ and let

$$\xi = T_gL_{g^{-1}}(X(g)) \quad \text{and} \quad \eta = T_gL_{g^{-1}}(Y(g)).$$

Let ξ_L and η_L denote the left invariant vector fields on G generated by ξ and η respectively. The definition of the Lie bracket on $\mathfrak{g}$ implies that $[\xi, \eta]_L = [\xi_L, \eta_L]$. We have

$$\begin{aligned}[X, Y](g) = {} & [X - \xi_L, Y - \eta_L](g) \\ & + [\xi_L, Y - \eta_L](g) + [X - \xi_L, \eta_L](g) + [\xi_L, \eta_L](g).\end{aligned}$$

The last term equals $[\xi, \eta]_L(g) \in \tilde{\mathfrak{h}}_g$. The first three terms all have the following structure: U and V are sections of $\tilde{\mathfrak{h}}$ and $V(g) = 0$. If we can prove that $[U, V](g) \in \tilde{\mathfrak{h}}_g$, this will show that each of the first three terms lies in $\tilde{\mathfrak{h}}_g$ and we can then conclude that $\tilde{\mathfrak{h}}$ is an involutive distribution.

The following Lemma solves this problem.

Lemma N9.D.1. *Let M be a manifold and let E be a subbundle of TM. If Y is a section of E such that $Y(m_0) = 0$ for a given point $m_0 \in M$, then $[X, Y](m_0) \in E_{m_0}$ for any $X \in \mathfrak{X}(M)$.*

Proof. Let $\mathbf{E}$ be the Banach space modeling M. Since the problem is local, we can replace M by an open neighborhood U of $\mathbf{0} \in \mathbf{E}$, TM by $U \times \mathbf{E}$, and m_0 by $\mathbf{0}$. Because E is a subbundle of TM, there is a splitting $\mathbf{E} = \mathbf{E}_1 \times \mathbf{E}_2$ such that, locally, E can be replaced by $U \times \mathbf{E}_1$. A section Y of E is of the form $x \in U \to (x, (f(x), 0))$, where $f : U \to \mathbf{E}_1$ is a smooth function. The condition $X(m_0) = 0$ is equivalent to $f(\mathbf{0}) = \mathbf{0}$.

Let $X \in \mathfrak{X}(M)$ be arbitrary. Represent it locally in the chart with domain U by $X(x) = (x, g(x))$, where $g : U \to \mathbf{E}$ is a smooth function. Then, locally, in U,

$$\begin{aligned} [X, Y](\mathbf{0}) &= \mathbf{D}(f, 0)(\mathbf{0}) \cdot g(0) - \mathbf{D}g(\mathbf{0}) \cdot (f(0), 0) \\ &= \mathbf{D}f(\mathbf{0}) \cdot g(0) \in F_1 \end{aligned}$$

since $f(0) = 0$ and since $\mathbf{D}f(0) \in L(\mathbf{E}, \mathbf{E}_1)$. Therefore $[X, Y](m_0) \in E_{m_0}$. ■

Returning to the proof of the theorem and applying the theorem of Frobenius to the involutive subbundle $\tilde{\mathfrak{h}} \in TG$, it follows that $\tilde{\mathfrak{h}}$ is integrable. Let H be the maximal integral submanifold of $\tilde{\mathfrak{h}}$ through the identity, that is, $e \in H$, $T_h H = \tilde{\mathfrak{h}}_h$ for any $h \in H$, and H is the maximal (relative to the inclusion) immersed submanifold of G having these properties.

We shall prove that H is a subgroup of G. If $g \in G$, then gH is the maximal integral manifold containing g. Indeed, $g \in gH$ since $e \in H$ and if $h \in H$, then

$$\begin{aligned} T_{gh}(gH) &= T_{gh}(L_g H) = T_h L_g(T_h H) \\ &= T_h L_g(\tilde{\mathfrak{h}}_h) = T_h L_g(T_e L_h \mathfrak{h}) \\ &= T_e L_{gh}(\mathfrak{h}) = \tilde{\mathfrak{h}}_{gh} \end{aligned}$$

since $L_g : G \to G$ is a diffeomorphism. Therefore, by uniqueness of the maximal integral manifolds, if $h \in H$, it follows that $hH = H$. Thus, if $k \in H$, then $hk \in hH = H$. Moreover, if $h \in H$, then $h^{-1}H$ is the maximal integral manifold through h^{-1} and this integral manifold contains $h^{-1}h = e$ since $h \in H$. Thus $h^{-1}H$ contains e and, again by uniqueness of the maximal integral manifolds, it follows that $h^{-1}H = H$, that is, $h^{-1} \in H$.

Next, we show that H is a Lie subgroup of G. Indeed,

$$(h, k) \in H \times H \hookrightarrow G \times G \mapsto hk \in H \hookrightarrow G$$

is a smooth map from $H \times H$ to G. However, since $hk \in H$ the map is smooth from $H \times H$ to H.

By construction, $T_eH = \mathfrak{h}$, so $\mathfrak{h}$ is the Lie algebra of H.

Finally we prove that H is the unique connected Lie subgroup of G with Lie algebra $\mathfrak{h}$. Suppose that H_1 was another such Lie subgroup. Then, if

$$h \in H_1,\ T_hH_1 = T_eL_h(T_eH_1) = T_eL_h(\mathfrak{h}) = \mathfrak{h}_h,$$

that is, H_1 is an integral submanifold of $\tilde{\mathfrak{h}}$. Therefore $H_1 \subset H$ is an open subgroup, hence it is also closed, and by connectedness of H it follows that $H_1 = H$. ∎

N9.E Relations between the Symplectic, Orthogonal, and Unitary Groups

We now want to relate $\mathrm{Sp}(2n, \mathbb{R})$, $\mathrm{O}(2n)$, and $\mathrm{U}(n)$. Following this, we shall discuss their quaternionic counterparts.

Our first goal is to show that

$$\mathrm{Sp}(2n, \mathbb{R}) \cap \mathrm{O}(2n, \mathbb{R}) = \mathrm{U}(n).$$

To make this meaningful, we identify $\mathbb{C}^n = \mathbb{R}^n \oplus i\mathbb{R}^n$ and we express the Hermitian inner product on $\mathbb{C}^n$ as a pair of real bilinear forms, namely, if we use the notation

$$\mathbf{x}_1 + i\mathbf{y}_1,\ \mathbf{x}_2 + i\mathbf{y}_2 \in \mathbb{C}^n \quad \text{for} \quad \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n,$$

then

$$\langle \mathbf{x}_1 + i\mathbf{y}_1, \mathbf{x}_2 + i\mathbf{y}_2 \rangle = \langle \mathbf{x}_1, \mathbf{y}_1 \rangle + \langle \mathbf{x}_2, \mathbf{y}_2 \rangle + i\left(\langle \mathbf{x}_2, \mathbf{y}_1 \rangle - \langle \mathbf{x}_1, \mathbf{y}_2 \rangle \right).$$

Thus, identifying $\mathbb{C}^n$ with $\mathbb{R}^n \times \mathbb{R}^n$ and $\mathbb{C}$ with $\mathbb{R} \times \mathbb{R}$, we can write

$$\langle (\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \rangle = \left((\mathbf{x}_1, \mathbf{x}_2) \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, -(\mathbf{x}_1, \mathbf{x}_2) \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \right). \tag{N9.E.1}$$

The next task is to represent elements of $\mathrm{U}(n)$ as $2n \times 2n$ matrices with real entries. Since $\mathrm{U}(n)$ is a closed subgroup of $\mathrm{GL}(n, \mathbb{C})$ we begin by representing the elements of $\mathfrak{gl}(n, \mathbb{C})$ in this way. Let $A + iB \in \mathfrak{gl}(n, \mathbb{C})$ with $A, B \in \mathfrak{gl}(n, \mathbb{R})$ and let $\mathbf{x} + i\mathbf{y} \in \mathbb{C}^n$. Then

$$(A + iB)(\mathbf{x} + i\mathbf{y}) = (A\mathbf{x} - B\mathbf{y}) + i(A\mathbf{y} + B\mathbf{x})$$

suggest that the map

$$\mathcal{C} : A + iB \in \mathrm{GL}(n, \mathbb{C}) \mapsto \begin{bmatrix} A & -B \\ B & A \end{bmatrix} \in \mathrm{GL}(2n, \mathbb{R}) \tag{N9.E.2}$$

is the desired embedding of $\mathrm{GL}(n, \mathbb{C})$ into $\mathrm{GL}(2n, \mathbb{R})$. It is straightforward to verify that the map $\mathcal{C}$ is an injective Lie group homomorphism, so we can identify $\mathrm{GL}(n, \mathbb{C})$ with all invertible $2n \times 2n$ matrices of the form

$$\begin{bmatrix} A & -B \\ B & A \end{bmatrix} \tag{N9.E.3}$$

with $A, B \in \mathfrak{gl}(n, \mathbb{R})$. It is obvious that

$$\mathcal{C}\left((A + iB)^{\dagger}\right) = [\mathcal{C}(A + iB)]^T$$

and

$$\operatorname{trace} \mathcal{C}(A + iB) = 2 \operatorname{Re} \operatorname{trace} (A + iB).$$

The relation

$$\det \mathcal{C}(A + iB) = |\det(A + iB)|^2$$

shows that $A + iB \in GL(n, \mathbb{C})$ if and only if $\mathcal{C}(A + iB) \in GL(2n, \mathbb{R})$. To prove this identity, bring $A + iB$ into Jordan canonical form, so that its determinant equals the product of the diagonal entries: $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. Since $\mathcal{C}$ is a group homomorphism, the identity holds if we can prove it for complex matrices in Jordan canonical form, that is,

$$A = \operatorname{diag}(\operatorname{Re}\lambda_1, \ldots, \operatorname{Re}\lambda_n), \quad \text{and} \quad B = \operatorname{diag}(\operatorname{Im}\lambda_1, \ldots, \operatorname{Im}\lambda_n) + N,$$

where N is the nilpotent matrix with 1's occupying some places on the first upper diagonal given by the complex Jordan canonical form

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) + N.$$

Interchanging columns and rows (for each column interchange do the same for the rows) one can transform this matrix to a block upper triangular matrix, each block on the diagonal being of the form

$$\begin{bmatrix} \operatorname{Re} \lambda_1 & -\operatorname{Im} \lambda_1 \\ \operatorname{Im} \lambda_1 & \operatorname{Re} \lambda_1 \end{bmatrix},$$

the upper 2×2 blocks being either the zero matrix or the matrix

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

and all other entries being zero. This matrix has the same determinant as the original one (because an even number of columns and row changes have been performed) and, since it is block upper triangular, this determinant equals the product of the determinants of the diagonal blocks, that is, $(\text{Re } \lambda_k)^2 + (\text{Im } \lambda_k)^2 = |\lambda_k|^2$, which proves the statement.

Using the embedding $\mathcal{C}$ defined above, it follows that, $\mathrm{U}(n)$ is embedded in $\mathrm{GL}(2n, \mathbb{R})$ as the set of matrices of the form (N9.E.3) with a certain additional property to be determined below. If $A + iB \in \mathrm{U}(n)$ then

$$(A + iB)^\dagger (A + iB) = I.$$

However, under the homomorphism (N9.E.2)

$$(A + iB)^\dagger = A^T - iB^T$$

is sent to the matrix

$$\begin{bmatrix} A^T & B^T \\ -B^T & A^T \end{bmatrix}.$$

Therefore,

$$(A + iB)^\dagger (A + iB) = I$$

becomes

$$\begin{aligned} \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} &= \begin{bmatrix} A^T & B^T \\ -B^T & A^T \end{bmatrix} \begin{bmatrix} A & -B \\ B & A \end{bmatrix} \\ &= \begin{bmatrix} A^T A + B^T B & -A^T B + B^T A \\ -B^T A + A^T B & B^T B + A^T A \end{bmatrix} \end{aligned}$$

which is equivalent to

$$A^T A + B^T B = I \quad \text{and} \quad A^T B \text{ is symmetric.} \tag{N9.E.4}$$

Proposition N9.E.1. *The following holds:*

$$\mathrm{Sp}(2n, \mathbb{R}) \cap \mathrm{O}(2n, \mathbb{R}) = \mathrm{U}(n).$$

As we shall see, this is the first in a series of three parallel results of this sort.

Proof. We have already seen that $A + iB \in \mathrm{U}(n)$ iff (N9.E.4) holds.

Now let us characterize all matrices of the form

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{Sp}(2n, \mathbb{R}) \cap \mathrm{O}(2n, \mathbb{R}).$$

Recall from the main text that a block matrix like this is symplectic iff

$$A^T D - C^T B = I \quad \text{and} \quad A^T C, B^T D \text{ are symmetric.} \tag{N9.E.5}$$

Since this matrix is also in $\mathrm{O}(2n)$, we have

$$\begin{aligned}\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A^T & C^T \\ B^T & D^T \end{bmatrix} \\ &= \begin{bmatrix} AA^T + BB^T & AC^T + BD^T \\ CA^T + DB^T & CC^T + DD^T \end{bmatrix}\end{aligned}$$

which is equivalent to

$$AA^T + BB^T = I, \quad AC^T + BD^T = 0, \quad CC^T + DD^T = I. \qquad \text{(N9.E.6)}$$

Now, multiply on the right by D the first identity in (N9.E.6), to get from (N9.E.5)

$$\begin{aligned} D &= AA^T D + BB^T D \\ &= A(I + C^T B) + BB^T D \\ &= A + AC^T B + BD^T B \\ &= A + (AC^T + BD^T)B = A \end{aligned}$$

by the second identity in (N9.E.6). Next, multiply on the right by B the last identity in (N9.E.6) and use, as before, (N9.E.5) to get

$$\begin{aligned} B &= CC^T B + DD^T B \\ &= C(A^T D - I) + DD^T B \\ &= CA^T D - C + DB^T D \\ &= -C + (CA^T + DB^T)D = -C \end{aligned}$$

by the second identity in (N9.E.6). We have thus shown that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{Sp}(2n, \mathbb{R}) \cap \mathrm{O}(2n)$$

iff $A = D$, $B = -C$, $A^T A + C^T C = I$, and $A^T C$ is symmetric, which coincide with the conditions (N9.E.4) characterizing $\mathrm{U}(n)$. ■

Notice that it follows from this and the fact that elements of $\mathrm{Sp}(2n, \mathbb{R})$ have determinant 1, that

$$\mathrm{Sp}(2n, \mathbb{R}) \cap \mathrm{SO}(2n, \mathbb{R}) = \mathrm{SU}(n).$$

The Group $\mathrm{GL}(n, \mathbb{H})$. By analogy to $\mathbb{R}^n$ and $\mathbb{C}^n$ we define quaternionic n space by

$$\mathbb{H}^n = \{\mathbf{a} = (a_1, \ldots, a_n) \mid a_i \in \mathbb{H}\}.$$

This satisfies all axioms of an n-dimensional vector space over $\mathbb{H}$ with the sole exception that $\mathbb{H}$ is not a field, being non-commutative.The group $\mathrm{GL}(n, \mathbb{H})$ is defined to be the set of all invertible $\mathbb{H}$–linear maps $T : \mathbb{H}^n \to \mathbb{H}^n$ defined by left multiplication by a $n \times n$ matrix $[t_r^p]$, with $t_r^p \in \mathbb{H}$, that is,

$$(T\mathbf{a})_r = \sum_{p=1}^{n} t_r^p a_p,$$

for $\mathbf{a} \in \mathbb{H}^n$. Because of non–commutativity of $\mathbb{H}$, care has to be taken with the concept of $\mathbb{H}$–linearity. It is straightforward to note that

$$T(\mathbf{a}\alpha) = (T\mathbf{a})\alpha,$$

for any $\alpha \in \mathbb{H}$, but that $T(\alpha\mathbf{a}) \neq \alpha(T\mathbf{a})$, in general. Therefore, *usual matrix multiplication is a right–linear map and, in general, it is not left–linear over* $\mathbb{H}$. In complete analogy with the real case, $\mathbb{C}^{2n}$ and $\mathbb{H}^n$ are isomorphic. However, there is a lot of structure that we shall exploit below by realizing left quaternionic matrix multiplication as a complex linear map. To achieve this, we shall identify, as before, $i \in \mathbb{C}$ with the quaternion $\mathbf{i} \in \mathbb{H}$ and will define the fundamental right complex isomorphism

$$\chi : \mathbb{C}^{2n} \to \mathbb{H}^n$$

by

$$\chi(\mathbf{u}, \mathbf{v}) = \mathbf{u} + \mathbf{j}\mathbf{v},$$

where $\mathbf{u}$, $\mathbf{v} \in \mathbb{C}^n$, and we regard $\mathbb{C}$ embedded in $\mathbb{H}$ by $x + iy \mapsto x + \mathbf{i}y$, for $x, y \in \mathbb{R}$. We have

$$\chi((\mathbf{u}, \mathbf{v})\alpha) = \chi(\mathbf{u}, \mathbf{v})\alpha$$

for all $\alpha \in \mathbb{C}$. So, again, we get only *right* linearity. The key property of χ is that it turns a left quaternionic matrix multiplication operator into a usual complex linear operator on $\mathbb{C}^{2n}$. Indeed, *if* $[t_{pr}]$ *is a quaternionic* $n \times n$ *matrix, then* $\chi^{-1}T\chi : \mathbb{C}^{2n} \to \mathbb{C}^{2n}$ *is complex linear*. To verify this, let $\alpha \in \mathbb{C}$, $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ and note that

$$\begin{aligned}\left(\chi^{-1}T\chi\right)(\alpha(\mathbf{u}, \mathbf{v})) &= \left(\chi^{-1}T\chi\right)((\mathbf{u}, \mathbf{v})\alpha) = \left(\chi^{-1}T\right)((\chi(\mathbf{u}, \mathbf{v}))\alpha) \\ &= \left(\chi^{-1}\right)((T\chi(\mathbf{u}, \mathbf{v}))\alpha) = \left(\chi^{-1}T\chi(\mathbf{u}, \mathbf{v})\right)\alpha \\ &= \alpha\left(\chi^{-1}T\chi(\mathbf{u}, \mathbf{v})\right).\end{aligned}$$

Let us determine, for example, the $2n \times 2n$ complex matrix J that corresponds to the right linear quaternionic map given by left multiplication with the diagonal map $\mathbf{jI}$. We have

$$\begin{aligned}J(\mathbf{u}, \mathbf{v}) &= (\chi^{-1}\mathbf{jI}\chi)(\mathbf{u}, \mathbf{v}) \\ &= (\chi^{-1}\mathbf{jI})(\mathbf{u} + \mathbf{j}\mathbf{v}) = \chi^{-1}(\mathbf{j}\mathbf{u} - \mathbf{v}) \\ &= (-\mathbf{v}, \mathbf{u}),\end{aligned}$$

that is,

$$J = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{bmatrix}$$

is the canonical symplectic structure on $\mathbb{C}^n \times \mathbb{C}^n = \mathbb{C}^{2n}$. Define the injective map between the space of right linear quaternionic maps on $\mathbb{H}^n$ defined by left multiplication by a matrix to the space of complex linear maps on $\mathbb{C}^{2n}$ by $T \mapsto T_\chi := \chi^{-1}T\chi$. Among all the complex linear maps $\mathbb{C}^{2n} \to \mathbb{C}^{2n}$ we want to characterize those that correspond to left matrix multiplication on $\mathbb{H}^n$. To achieve this, write

$$T = A + \mathbf{j}B$$

where A and B are complex $n \times n$ matrices. The relation

$$(A + \mathbf{j}B)(\mathbf{u} + \mathbf{j}\mathbf{v}) = A\mathbf{u} - \overline{B}\mathbf{v} + \mathbf{j}(B\mathbf{u} + \overline{A}\mathbf{v})$$

obtained by using the identity $\mathbf{j}\alpha = \overline{\alpha}\mathbf{j}$ for $\alpha \in \mathbb{H}$, shows that

$$\begin{aligned} T_\chi(\mathbf{u}, \mathbf{v}) &= \chi^{-1}T\,(\mathbf{u} + \mathbf{j}\mathbf{v}) \\ &= \left(A\mathbf{u} - \overline{B}\mathbf{v}, B\mathbf{u} + \overline{A}\mathbf{V}\right) \\ &= \begin{bmatrix} A & -\overline{B} \\ B & \overline{A} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \end{aligned}$$

Thus

$$\mathcal{H} : A + \mathbf{j}B \in \mathfrak{gl}(n, \mathbb{H}) \mapsto \begin{bmatrix} A & -\overline{B} \\ B & \overline{A} \end{bmatrix} \in \mathfrak{gl}(2n, \mathbb{C})$$

satisfies

$$\mathcal{H}\left((A + \mathbf{j}B)^\dagger\right) = \left[\mathcal{H}(A + \mathbf{j}B)\right]^\dagger,$$

where $(A + \mathbf{j}B)^\dagger = \overline{A}^T - \overline{B}^T\mathbf{j}$ and

$$\text{trace } \mathcal{H}(A + \mathbf{j}B) = 2 \text{ Re trace } (A + \mathbf{j}B).$$

Also, $\mathcal{H}$ is a homomorphism relative to multiplication. Therefore, *the Lie algebra* $\mathfrak{gl}(n, \mathbb{H})$ *is isomorphic over* $\mathbb{C}$ *to the complex Lie algebra*

$$\mathfrak{u}^*(2n) := \left\{ \begin{bmatrix} A & -\overline{B} \\ B & \overline{A} \end{bmatrix} \middle| A, B \in \mathfrak{gl}(n, \mathbb{C}) \right\}.$$

Define

$$\mathfrak{sl}(n, \mathbb{H}) = \{T \in \mathfrak{gl}(n, \mathbb{H}) \mid \text{Re trace } T = 0\}.$$

From the above considerations, it follows that $\mathfrak{sl}(n, \mathbb{H})$ *is isomorphic over* $\mathbb{C}$ *to the complex Lie algebra*

$$\mathfrak{su}^*(2n) = \{M \in \mathfrak{u}^*(2n) \mid \operatorname{trace} M = 0\} .$$

Since $\mathcal{H}$ is injective and preserves multiplication, it follows that

$$\begin{aligned}
&\mathcal{H}(\mathrm{GL}(n, \mathbb{H})) := U^*(2n) \\
&= \left\{ \left[\begin{array}{cc} A & -\overline{B} \\ B & \overline{A} \end{array} \right] \middle| A, B \in \mathfrak{gl}(n, \mathbb{C}),\ \det \left[\begin{array}{cc} A & -\overline{B} \\ B & \overline{A} \end{array} \right] \neq 0 \right\}
\end{aligned}$$

is a closed Lie subgroup of $\mathrm{GL}(2n, \mathbb{C})$. Realizing $\mathrm{GL}(n, \mathbb{H})$ as $U^*(2n)$ avoids the introduction of the concept of determinant of a square matrix with entries in $\mathbb{H}$, which is possible (this determinant is called the ***Dieudonné determinant***), but would take us too far afield. Thus one defines

$$\begin{aligned}
\mathrm{SU}^*(2n) = \Bigg\{ \left[\begin{array}{cc} A & -\overline{B} \\ B & \overline{A} \end{array} \right] \Bigg| & A, B \in \mathfrak{gl}(n, \mathbb{C}), \\
& \det \left[\begin{array}{cc} A & -\overline{B} \\ B & \overline{A} \end{array} \right] = 1 \Bigg\} .
\end{aligned}$$

and

$$\mathrm{SL}(n, \mathbb{H}) = \mathcal{H}^{-1} \left(\mathrm{SU}^*(2n) \right) .$$

The subgroups $\mathrm{SU}^*(2n)$ and $\mathrm{SL}(n, \mathbb{H})$ are closed in $\mathrm{U}^*(2n)$ and $\mathrm{GL}(n, \mathbb{H})$ respectively.

Proceeding as in the real and complex cases (and using the quaternionic inner product

$$\langle M, N \rangle = \operatorname{trace}(M N^\dagger)$$

for $M, N \in \mathfrak{gl}(n, \mathbb{H})$ one gets

Proposition N9.E.2. *The quaternionic general linear group* $\mathrm{GL}(n, \mathbb{H})$ *is isomorphic over* $\mathbb{C}$ *to* $\mathrm{U}^*(2n)$ *and has complex dimension* $2n^2$*. It is a noncompact connected Lie group. Its Lie algebra* $\mathfrak{gl}(n, \mathbb{H})$ *consisting of* $n \times n$ *quaternionic matrices is isomorphic over* $\mathbb{C}$ *to* $\mathfrak{u}^*(2n)$*. The quaternionic special linear group* $\mathrm{SL}(n, \mathbb{H})$ *is isomorphic over* $\mathbb{C}$ *to* $\mathrm{SU}^*(2n)$*. Its complex dimension is* $2n^2 - 1$ *and it is a noncompact connected closed Lie subgroup of* $\mathrm{GL}(n, \mathbb{H})$*. Its Lie algebra is* $\mathfrak{sl}(n, \mathbb{H})$ *which is isomorphic over* $\mathbb{C}$ *to* $\mathfrak{su}^*(2n)$*.*

As usual, the connectedness statements need some comments. We shall see in §9.3 that $\mathrm{SL}(n, \mathbb{H})$ is connected because $\mathrm{Sp}(2n)$, to be introduced below, is connected. The connectedness of $\mathrm{GL}(n, \mathbb{H})$ follows from the exact sequence

$$\{1\} \to \mathbb{H} \setminus \{0\} \to \mathrm{GL}(n, \mathbb{H}) \to \mathrm{SL}(n, \mathbb{H}) \to \{\mathbf{I}\}$$

and the theorem stating that if H is closed subgroup of G such that both H and G/H are connected, then G is connected (see Varadarajan [1974] or Abraham, Marsden, and Ratiu [1988] for the general case of bundles).

The Unitary Symplectic Group $\mathrm{Sp}(2n)$**.** We want to construct a group analogous to $\mathrm{O}(n)$ when we worked with $\mathbb{R}^n$, or to $\mathrm{U}(n)$ when we worked with $\mathbb{C}^n$.

For this, we introduce the quaternionic inner product

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{H}} = \sum_{p=1}^{n} a_p \overline{b}_p,$$

where $\mathbf{a}$, $\mathbf{b} \in \mathbb{H}^n$ and $\overline{b}_p$ is the quaternion conjugate to b_p, for $p = 1, \ldots, n$. Again, the usual axioms for the inner product are satisfied, by being careful in the scalar multiplication by quaternions, that is,

(i) $\langle \mathbf{a}_1 + \mathbf{a}_2, \mathbf{b} \rangle = \langle \mathbf{a}_1, \mathbf{b} \rangle + \langle \mathbf{a}_2, \mathbf{b} \rangle$,

(ii) $\langle \alpha \mathbf{a}, \mathbf{b} \rangle = \alpha \langle \mathbf{a}, \mathbf{b} \rangle$ and $\langle \mathbf{a}, \mathbf{b}\alpha \rangle = \langle \mathbf{a}, \mathbf{b} \rangle \overline{\alpha}$, for all $\alpha \in \mathbb{H}$,

(iii) $\langle \mathbf{a}, \mathbf{b} \rangle = \overline{\langle \mathbf{b}, \mathbf{a} \rangle}$,

(iv) $\langle \mathbf{a}, \mathbf{a} \rangle \geq 0$ and $\langle \mathbf{a}, \mathbf{a} \rangle = 0$ iff $\mathbf{a} = \mathbf{0}$.

Any quaternionic vector can be written as $\mathbf{u} + \mathbf{j}\mathbf{v} \in \mathbb{H}^n$, where $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$. A straightforward computation shows that

$$\begin{aligned} \langle \mathbf{u}_1 + \mathbf{j}\mathbf{v}_1, \mathbf{u}_2 + \mathbf{j}\mathbf{v}_2 \rangle &= \langle \mathbf{u}_1, \mathbf{u}_2 \rangle + \overline{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle} \\ &\qquad + \mathbf{j} \left(\langle \mathbf{v}_1, \mathbf{u}_2 \rangle - \overline{\langle \mathbf{u}_1, \mathbf{v}_2 \rangle} \right). \end{aligned}$$

If $T \in \mathrm{GL}(n, \mathbb{H})$, express it as $T = A + \mathbf{j}B$, with $A, B \in \mathfrak{gl}(n, \mathbb{C})$, use the homomorphism $\mathcal{H}$ and, defining

$$\mathrm{Sp}(2n) = \left\{ T \in \mathrm{GL}(n, \mathbb{H}) \mid T^\dagger T = T T^\dagger = I \right\}$$

express the defining condition in terms of $2n \times 2n$ complex matrices, that is, in $\mathrm{U}^*(2n)$. We get

$$\begin{aligned} \mathrm{Sp}(2n) &= \left\{ \begin{bmatrix} A & -\overline{B} \\ B & \overline{A} \end{bmatrix} \middle| AA^\dagger + \overline{B}\overline{B}^\dagger = I,\ AB^{-T} \text{ symmetric} \right\} \\ &= \left\{ \begin{bmatrix} A & -\overline{B} \\ B & \overline{A} \end{bmatrix} \middle| A^\dagger A + B^\dagger B = I,\ A^T B \text{ symmetric} \right\} \end{aligned}$$

whose Lie algebra is clearly

$$\begin{aligned} \mathfrak{sp}(2n) &= \left\{ \begin{bmatrix} A & -\overline{B} \\ B & \overline{A} \end{bmatrix} \in \mathfrak{gl}(2n, \mathbb{C}) \middle| A + A^\dagger = 0,\ B^T = B \right\} \\ &= \left\{ T \in \mathfrak{gl}(n, \mathbb{H}) \mid T^\dagger + T = 0 \right\}. \end{aligned}$$

Note that the trace of any element in $\mathfrak{sp}(2n)$ is necessarily zero and hence that any element in $\mathrm{Sp}(2n)$ has determinant equal to 1. Thus, unlike the case of real or complex matrices, where the isometry condition did *not* imply that the determinant is zero (and hence we distinguished between $\mathrm{O}(n)$ and $\mathrm{SO}(n), \mathrm{U}(n)$ and $\mathrm{SU}(n)$), in the case of quaternionic matrices there is only one group of isometries, namely $\mathrm{Sp}(2n)$ and the determinant equal to 1 condition is automatically satisfied.

Proposition N9.E.3. *The unitary symplectic group* $\mathrm{Sp}(2n)$ *is the group of isometries of* $\mathbb{H}^n$. *It is a compact connected subgroup* $\mathrm{SL}(n,\mathbb{H}) \cong \mathrm{SU}^*(2n)$ *of complex dimension* $2n^2+n$ *whose Lie algebra is* $\mathfrak{sp}(2n)$.

Compactness is proved exactly as in the real or complex case by showing that the norm of an element in $\mathrm{Sp}(2n)$ is equal to $\sqrt{2n}$ and the proof of connectedness is, as usual, deferred to §9.3. From our previous consideration it immediately follows that:

Proposition N9.E.4.

$$\mathrm{Sp}(2n) = \mathrm{SU}^*(2n) \cap \mathrm{U}(2n).$$

The Complex Symplectic Group $\mathrm{Sp}(2n,\mathbb{C})$ is defined exactly as in the real case by the condition

$$\mathrm{Sp}(2n,\mathbb{C}) = \left\{T \in \mathrm{GL}(2n,\mathbb{C}) \mid T^T \mathbb{J} T = \mathbb{J}\right\}.$$

It is a noncompact connected closed Lie subgroup of $\mathrm{GL}(2n,\mathbb{C})$ of complex dimension $2n^2+n$ and whose Lie algebra is

$$\mathfrak{sp}(2n,\mathbb{C}) = \left\{T \in \mathfrak{gl}(2n,\mathbb{C}) \mid T^T \mathbb{J} + \mathbb{J} T = 0\right\}.$$

Proposition N9.E.5.

$$\mathrm{Sp}(2n) = \mathrm{Sp}(2n,\mathbb{C}) \cap \mathrm{U}(2n).$$

Proof. Recall that

$$T = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \in \mathrm{Sp}(2n,\mathbb{C})$$

if and only if $A^T B$ and $C^T D$ are symmetric and $A^T D - B^T C = I$ (see 9.2.12). Also, $T \in \mathrm{U}(2n)$, if and only if

$$A^\dagger A + B^\dagger B = I,\ C^\dagger C + D^\dagger D = I \quad \text{and} \quad A^\dagger C + B^\dagger D = 0.$$

From the characterization (N9.E.7) it follows that all these conditions hold. Conversely, if these conditions hold, then

$$\begin{aligned}
\mathbb{J} &= T^T \mathbb{J} T = T^T \overline{T} \left(\overline{T}\right)^{-1} \mathbb{J} T \\
&= \overline{T}^\dagger \overline{T} \left(\overline{T}\right)^{-1} \mathbb{J} T = \overline{(T^\dagger T)} \left(\overline{T}\right)^{-1} \mathbb{J} T \\
&= \left(\overline{T}\right)^{-1} \mathbb{J} T
\end{aligned}$$

since $T^\dagger T = I$ because $T \in \mathrm{U}(2n)$. Therefore $\overline{T}\mathbb{J} = \mathbb{J} T$ which forces $C = -\overline{B}, D = \overline{A}$. But then, these conditions imply those in (N9.E.7). ■

N9.F Generic Coadjoint Isotropy Subalgebras are Abelian

The aim of this section is to prove a theorem of Duflo and Vergne [1969] showing that, generically, the isotropy algebras for the coadjoint action are Abelian. A very simple example is $G = \mathrm{SO}(3)$. Here $\mathfrak{g}^* \cong \mathbb{R}^3$ and $G_\mu = S^1$ for $\mu \in \mathfrak{g}^*$ and $\mu \neq 0$, and $G_0 = \mathrm{SO}(3)$. Thus, G_μ is abelian on the open dense set $\mathfrak{g}^* \backslash \{0\}$.

To prepare for the proof, we shall develop some tools.

If V is a finite-dimensional vector space, a subset $A \subset V$ is called ***algebraic*** if it is the common zero set of a finite number of polynomial functions on V. It is easy to see that if A_i is the zero set of a finite collection of polynomials C_i, for $i = 1, 2$, then $A_1 \cup A_2$ is the zero set of the collection $C_1 C_2$ formed by all products of an element in C_1 with an element in C_2. The whole space V is the zero set of the constant polynomial equal to 1. Finally, if A_α is the algebraic set given as the common zeros of some finite collection of polynomials C_α, where α ranges over some index set, then $\bigcap_\alpha A_\alpha$ is the zero set of the collection $\bigcup_\alpha C_\alpha$. This zero set can also be given as the common zeros of a *finite* collection of polynomials since the zero set of any collection of polynomials coincides with the zero set of the ideal in the polynomial ring generated by this collection and any ideal in the polynomial ring over $\mathbb{R}$ is finitely generated (we accept this from algebra). Thus, the collection of algebraic sets in V satisfies the axioms of the collection of closed sets of a topology which is called the ***Zariski topology*** of V.

Thus, the open sets of this topology are the complements of the algebraic sets. For example, the algebraic sets of $\mathbb{R}$ are just the finite sets, since every polynomial in $\mathbb{R}[X]$ has finitely many real roots (or none at all). Granting that we have a topology (the hard part), let us show that *any Zariski open set in V is open and dense in the usual topology.* Openness is clear, since algebraic sets are necessarily closed in the usual topology as inverse images of 0 by a continuous map. To show that a Zariski open set U is also dense, suppose the contrary, namely, that if $x \in V \backslash U$, then there is a neighborhood $U_1 \times U_2$ of x in the usual topology such that

$$(U_1 \times U_2) \cap U = \varnothing \quad \text{and} \quad U_1 \subset \mathbb{R}, U_2 \subset V_2$$

are open, where $V = \mathbb{R} \times V_2$, the splitting being achieved by the choice of a basis. Since $x \in V \backslash U$, there is a finite collection of polynomials

$$p_1, \ldots, p_N \in \mathbb{R}[X_1, \ldots, X_n], \qquad n = \dim V,$$

that vanishes identically on $U_1 \times U_2$. If $x = (x_1, \ldots, x_n) \in V$, then the polynomials

$$q_i(X_1) = p_i(X_1, x_2, \ldots, x_n) \in \mathbb{R}[X_1]$$

all vanish identically on the open set $U_1 \subset \mathbb{R}$, which is impossible since each q_i has at most a finite number of roots. Therefore, $(U_1 \times U_2) \cap U = \varnothing$ is absurd and hence U must be dense in V.

Theorem N9.F.1 (Duflo and Vergne)**.** *Let $\mathfrak{g}$ be a finite-dimensional Lie algebra with dual $\mathfrak{g}^*$ and let $r = \min\{\dim \mathfrak{g}_\mu \mid \mu \in \mathfrak{g}^*\}$. The set*

$$\mathfrak{g}^*_{reg} := \{\mu \in \mathfrak{g}^* \mid \dim \mathfrak{g}_\mu = r\}$$

is Zariski open and thus open and dense in the usual topology of $\mathfrak{g}^*$. If $\dim \mathfrak{g}_\mu = r$*, then $\mathfrak{g}_\mu$ is abelian.*

Proof (Due to J. Carmona, as presented in Rais [1972])**.** Define the map $\varphi_\mu : G \to \mathfrak{g}^*$ by $g \mapsto \mathrm{Ad}^*_{g^{-1}} \mu$. This is a smooth map whose range is the coadjoint orbit $\mathcal{O}_\mu$ through μ and whose tangent map at the identity is $T_e\varphi_\mu(\xi) = -\operatorname{ad}^*_\xi \mu$. Note that $\ker T_e\varphi_\mu = \mathfrak{g}_\mu$ and

$$\text{range } T_e\varphi_\mu = T_\mu\mathcal{O}_\mu.$$

Thus, if $n = \dim \mathfrak{g}$, we have

$$\text{rank } T_e\varphi_\mu = n - \dim \mathfrak{g}_\mu \leq n - r$$

since $\dim \mathfrak{g}_\mu \geq r$, for all $\mu \in \mathfrak{g}^*$. Therefore,

$$U = \{\mu \in \mathfrak{g}^* \mid \dim \mathfrak{g}_\mu = r\} = \{\mu \in \mathfrak{g}^* \mid \operatorname{rank}(T_e\varphi_\mu) = n - r\}$$

and $n - r$ is the maximal possible rank of all the linear maps

$$T_e\varphi_\mu : \mathfrak{g} \to \mathfrak{g}^*, \mu \in \mathfrak{g}^*.$$

Now choose a basis in $\mathfrak{g}$ and induce the natural bases on $\mathfrak{g}^*$ and

$$L(\mathfrak{g}, \mathfrak{g}^*).$$

Let

$$S_i = \{\mu \in \mathfrak{g}^* \mid \text{rank } T_e\varphi_\mu = n - r - i\}, 1 \leq i \leq n - r.$$

Then S_i is the zero set of the polynomials in μ obtained by taking all determinants of the $(n - r - i + 1)$-minors of the matrix representation of $T_e\varphi_\mu$ in these bases. Thus, S_i is an algebraic set. Since $\bigcup_{i=1}^{n-r} S_i$ is the complement of U, if follows that U is a Zariski open set in $\mathfrak{g}^*$, and hence open and dense in the usual topology of $\mathfrak{g}^*$.

Now let $\mu \in \mathfrak{g}^*$ be such that $\dim \mathfrak{g}_\mu = r$ and let V be a complement to $\mathfrak{g}_\mu$ in $\mathfrak{g}$, that is,

$$\mathfrak{g} = V \oplus \mathfrak{g}_\mu.$$

Then $T_e\varphi_\mu|V$ is injective. Fix $\nu \in \mathfrak{g}^*$ and define

$$S = \{t \in \mathbb{R} \mid T_e\varphi_{\mu+t\nu}|V \text{ is injective.}\}$$

Note that $0 \in S$ and that S is open in $\mathbb{R}$ because the set of injective linear maps is open in $L(\mathfrak{g}, \mathfrak{g}^*)$ and $\mu \mapsto T_e\varphi_\mu$ is continuous. Thus, S contains an open neighborhood of 0 in $\mathbb{R}$. Since the rank of a linear map can only increase by slight perturbations, we have rank

$$T_e\varphi_{\mu+t\nu}|V \geq \text{rank } T_e\varphi_\mu|V = n - r,$$

for $|t|$ small, and by maximality of $n - r$, this forces

$$\text{rank}\, T_e\varphi_{\mu+t\nu} = n - r$$

for t in a neighborhood of 0 contained in S. Thus, for $|t|$ small,

$$T_e\varphi_{\mu+t\nu}|V : V \to T_{\mu+t\nu}\mathcal{O}_{\mu+t\nu}$$

is an isomorphism. Hence, if $\xi \in \mathfrak{g}_\mu$, $\mathrm{ad}^*_\xi(\mu+t\nu) \in T_{\mu+t\nu}\mathcal{O}_{\mu+t\nu}$ is the image of a unique $\xi(t) \in V$ under $T_e\varphi_{\mu+t\nu}|V$, that is,

$$\xi(t) = (T_e\varphi_{\mu+t\nu}|V)^{-1}(\mathrm{ad}^*_\xi(\mu + t\nu)).$$

This formula shows that for $|t|$ small, $t \mapsto \xi(t)$ is a smooth curve in V and $\xi(0) = 0$. However, since

$$\mathrm{ad}^*_\xi(\mu + t\nu) = -T_e\varphi_{\mu+t\nu}(\xi),$$

the definition of $\xi(t)$ is equivalent to $T_e\varphi_{\mu+t\nu}(\xi(t) + \xi) = 0$, that is,

$$\xi(t) + \xi \in \mathfrak{g}_{\mu+t\nu}.$$

Similarly, given $\eta \in \mathfrak{g}_\mu$, there exists a unique $\eta(t) \in V$ such that

$$\eta(t) + \eta \in \mathfrak{g}_{\mu+t\nu}, \eta(0) = 0,$$

and $t \mapsto \eta(t)$ is smooth for small $|t|$. Therefore, the map

$$t \mapsto \langle \mu + t\nu, [\xi(t) + \xi, \eta(t) + \eta] \rangle$$

is identically zero for small $|t|$. In particular, its derivative at $t = 0$ is also zero. But this derivative equals

$$\begin{aligned}&\langle \nu, [\xi, \eta] \rangle + \langle \mu, [\xi'(0), \eta] \rangle + \langle \mu, [\xi, \eta'(0)] \rangle \\ &\qquad = \langle \nu, [\xi, \eta] \rangle - \left\langle \mathrm{ad}^*_\eta \mu, \xi'(0) \right\rangle + \left\langle \mathrm{ad}^*_\xi \mu, \eta'(0) \right\rangle = \langle \nu, [\xi, \eta] \rangle,\end{aligned}$$

since $\xi, \eta \in \mathfrak{g}_\mu$. Thus, $\langle \nu, [\xi, \eta] \rangle = 0$ for any $\nu \in \mathfrak{g}^*$, that is,

$$[\xi, \eta] = 0.$$

Since $\xi, \eta \in \mathfrak{g}_\mu$ are arbitrary, it follows that $\mathfrak{g}_\mu$ is Abelian. ■

We close this section with a different proof (due to R. Fillippini) that for $\mu \in \mathfrak{g}^*_{\mathrm{reg}}$, $\mathfrak{g}_\mu$ is Abelian. It uses concepts from later in the book on momentum maps and collective Hamiltonians, so its proof can be deferred. Another proof due to Weinstein [1983], page 535 is also instructive.

Proof that $\mathfrak{g}_\mu$ is Abelian for $\mu \in \mathfrak{g}^*_{\mathbf{reg}}$. The momentum map for lifted left action on G on T^*G is given by right translation to the identity $J_\lambda = \rho$. The momentum map for the lifted right action of G on T^*G is given by left translation to the identity, $J_\rho = \lambda$. Thus,

$$X_{\hat{J}_\lambda(\xi)}(p) = \xi \cdot p \quad \text{and} \quad X_{\hat{J}_\rho(\xi)}(p) = p \cdot \xi$$

for all p in T^*G and ξ in $\mathfrak{g}$.

Given $F : \mathfrak{g}^* \to \mathbb{R}$, a straightforward calculation shows that

$$X_{F \circ J_\lambda}(p) = \left(\frac{\delta F}{\delta J_\lambda(p)} \right) \cdot p \text{ and } X_{F \circ J_\rho}(p) = p \cdot \left(\frac{\delta F}{\delta J_\rho(p)} \right)$$

for all $p \in T^*G$ and $\xi \in \mathfrak{g}$.

If F is constant on the coadjoint orbits, then $F \circ J_\lambda = F \circ J_\rho$, hence

$$\left(\frac{\delta F}{\delta J_\lambda(p)} \right) \cdot p = p \cdot \left(\frac{\delta F}{\delta J_\rho(p)} \right) F$$

for all $p \in T^*G$. In particular, if $\mu \in \mathfrak{g}^*$ and $g \in G\mu$, so that $g \cdot \mu = \mu \cdot g$, we deduce that

$$\frac{\delta F}{\delta \mu} \cdot (g \cdot \mu) = (g \cdot \mu) \cdot \frac{\delta F}{\delta \mu}.$$

We know that $\mathfrak{g}^*_{\mathrm{reg}}$ is an open subset of $\mathfrak{g}^*$. Fix $\mu \in \mathfrak{g}^*_{\mathrm{reg}}$. There is then a neighborhood U of μ and a surjective submersion $\pi : U \to U/G$. If $F : U \to \mathbb{R}$ factors through $\pi : U \to U/G$, then a straightforward calculation shows that the preceding equation remains valid. A straightforward calculation shows that $\delta F/\delta \mu \in \mathfrak{g}_\mu$ for all $\mu \in \mathfrak{g}$. We now show conversely that given any $\xi \in \mathfrak{g}_\mu$, there exists a smooth function $F : U \to \mathbb{R}$ that factors through $\pi : U \to U/G$ such that $\delta F/\delta \mu = \xi$.

Let

$$[\mu, \mathfrak{g}] = \{ \mathrm{ad}^*_\eta \mu \mid \eta \in \mathfrak{g} \}.$$

Note that $[\mu, \mathfrak{g}]$ can be identified with $T_\mu O$, O being the coadjoint orbit through μ. It follows that we may identify

$$T_{\pi(\mu)}(U/G) \cong \mathfrak{g}^*/[\mu, \mathfrak{g}].$$

Since the linear map $\hat{\xi} : \mathfrak{g}^* \to \mathbb{R}$ factors through $\mathfrak{g}^*/[\mu, \mathfrak{g}]$, it follows that there exists a smooth map $\varphi : U/G \to \mathbb{R}$ for which

$$(\mathbf{d}\varphi)_{\pi(\mu)} \cdot (T_\mu \pi \cdot \nu) = \nu\,(\xi)$$

for all $\nu \in \mathfrak{g}^*$. Let $F = \varphi \circ \pi$. Then $\delta F/\delta\mu = \xi$.

It now follows that, for μ regular, $\xi \cdot (g \cdot \mu) = (g \cdot \mu) \cdot \xi$ for all $g \in G_\mu$ and $\xi \in \mathfrak{g}_\mu$. Taking $g = e$, we see that $\xi \cdot \mu = \mu \cdot \xi$; this is of course already clear, since $\xi \in \mathfrak{g}_\mu$. It follows that

$$(\xi \cdot g) \cdot \mu = \xi \cdot (g \cdot \mu) = (g \cdot \mu) \cdot \xi = g \cdot (\mu \cdot \xi) = g \cdot (\xi \cdot \mu) = (g \cdot \xi) \cdot \mu.$$

Since G acts freely on T^*G, it follows that

$$\xi \cdot g = g \cdot \xi$$

for all $g \in G_\mu$ and $\xi \in \mathfrak{g}_\mu$. By differentiating this relation in g at $g = e$, we see that g_μ is Abelian. ∎

Exercise

◇ **N9.F-1.** Prove the following generalization of the Duflo–Vergne Theorem due to Guillemin and Sternberg [1984]. Let S be an infinitesimally invariant submanifold of $\mathfrak{g}^*$, that is, $\mathrm{ad}^*_\xi \mu \in S$, whenever $\mu \in S$ and $\xi \in \mathfrak{g}$. Let $r = \min\{\dim \mathfrak{g}_\mu | \mu \in S\}$. Then $\dim \mathfrak{g}_\mu = r$ implies

$$[\mathfrak{g}_\mu, \mathfrak{g}_\mu] \subset (T_\mu S)^0 = \{\xi \in \mathfrak{g} \mid \langle u, \xi \rangle = 0, \quad \text{for all } u \in T_\mu S\}.$$

In particular $\mathfrak{g}_\mu/(T_\mu S)^0$ is abelian. Note that the Duflo–Vergne Theorem is the case for which $S = \mathfrak{g}^*$. [The solution to this exercise is given at the end of this internet supplement].

N9.G Some Infinite Dimensional Lie Groups

Infinite dimensional groups often arise as configuration spaces, symmetry groups, or gauge groups of physical systems with an infinite number of degrees of freedom. These groups often consist of functions, operators, or diffeomorphisms. Here we present without proof a number of facts about these infinite dimensional groups. To make the details of this infinite dimensional geometry precise, lengthy proofs involving delicate functional analytic and topological issues would need to be addressed. We shall discuss some of these issues, but shall not present all the detailed proofs. (These may be found in Palais [1968], Ebin and Marsden [1970], Ratiu and Schmid [1981], and Adams, Ratiu, and Schmid [1986].) In particular, we will find that some of the examples we discuss in this section are not Lie groups in the strict sense of our previous definitions, and caution will be required when applying the general theory to them. Fortunately, one can understand many of the main ideas and techniques without needing all the technicalities.

Examples

A. Consider a compact manifold M (possibly with boundary) and the infinite dimensional vector space $C^\infty(M)$ of all smooth real value functions on M. Evidently, $C^\infty(M)$ is an abelian group with pointwise defined group operations; that is,

$$\mu : C^\infty(M) \times C^\infty(M) \to C^\infty(M),\ (f, g) \mapsto f + g$$

is defined by $(f + g)(x) = f(x) + g(x)$, $x \in M$, and

$$I : C^\infty(M) \to C^\infty(M),\ f \mapsto -f$$

is defined by $(-f)(x) = -(f(x))$. The unit element $e = 0$ is given by $e(x) = 0 \in \mathbb{R}$ for all $x \in M$. As for any vector space, we formally have $T_eC^\infty(M) = C^\infty(M)$, so the Lie algebra of $C^\infty(M)$ coincides with $C^\infty(M)$, and the bracket is trivial: $[f, g] = 0$.

The space $C^\infty(M)$ is not a Banach Lie group since spaces of C^∞ functions do not form a Banach space. To get a Banach Lie group we can complete $C^\infty(M)$ to $C^k(M)$, $0 \le k < \infty$ or to $H^s(M)$, $s \ge 0$.

> Here H^s denotes a Sobolev space, whose definition and properties are summarized in a separate section below.

Thus $C^k(M)$ and $H^s(M)$ are Banach Lie groups. For noncompact M, it is sometimes useful to consider weighted Sobolev spaces for technical reasons involving elliptic equations. In fact it is usually necessary to use Sobolev spaces to prove facts about elliptic and hyperbolic equations. This is why C^∞, while formally nicer, is not suitable for many facts about partial differential equations.

The vector group $C^\infty(\mathbb{R}^3)$ and the Banach Lie groups $C^k(\mathbb{R}^3)$ and $H^s(\mathbb{R}^3)$ are closely related to the gauge group for electromagnetism: Maxwell's equations are invariant under the gauge transformation of the vector potential $A \mapsto A + \nabla\varphi$, for all $\varphi \in C^\infty(\mathbb{R}^3)$. The gauge group for a general Yang-Mills field also forms a Banach Lie group. ◆

B. Next consider a manifold M and the space

$$C^\infty(M, \mathbb{R}\backslash\{0\}) = \{f : M \to \mathbb{R}\backslash\{0\} \mid f \text{ is smooth } \}$$

of real valued, nowhere vanishing functions on M. Thus, $C^\infty(M, \mathbb{R}\backslash\{0\})$ is a group with pointwise defined group multiplication

$$\mu : C^\infty(M, \mathbb{R}\backslash\{0\}) \times C^\infty(M, \mathbb{R}\backslash\{0\}) \to C^\infty(M, \mathbb{R}\backslash\{0\})$$

defined by $(f, h) \mapsto fh$ where $(fh)(x) = f(x)h(x)$, $x \in M$, and inversion

$$I : C^\infty(M, \mathbb{R}\backslash\{0\}) \to C^\infty(M, \mathbb{R}\backslash\{0\})$$

defined by $f \mapsto f^{-1}$ where $f^{-1}(x) = 1/f(x)$, and $x \in M$. The unit element $e = 1$ satisfies $e(x) = 1 \in \mathbb{R}$, for all $x \in M$. This group is abelian since $\mathbb{R}\backslash\{0\}$ is abelian, and so its Lie algebra is $C^\infty(M) = C^\infty(M, \mathbb{R})$, with the trivial bracket $[f, g] = 0$.

The spaces $C^k(M, \mathbb{R}\backslash\{0\})$ and $H^s(M, \mathbb{R}\backslash\{0\})$ are Banach Lie groups for compact M and certain values of k and s given shortly. Note that $C^k(M, \mathbb{R}\backslash\{0\})$ is a subset of the Banach space $C^k(M, \mathbb{R})$, but in general it is not an open subset nor a submanifold. However, it is open if M is compact. Thus, for compact M, $C^k(M, \mathbb{R}\backslash\{0\})$, $k \geq 0$, is a Banach Lie group. For $H^s(M, \mathbb{R}\backslash\{0\})$ we need M compact as well, but even then, H^s need not be closed under pointwise multiplication. This requires, in addition, $s > n/2$, $n = \dim M$, as is discussed in the supplement below. For non-compact M one can use different topologies or replace $\mathbb{R}\backslash\{0\}$ by a compact Lie group G. In the latter case, $C^k(M, G)$ and $H^s(M, G)$, for $s > (1/2)\dim(M)$, are Banach Lie groups under pointwise multiplication

$$\mu(f, g)(x) = f(x)g(x),\ x \in M,$$

the latter product taken in G and inversion

$$I(f)(x) = (f(x))^{-1}.$$

The Lie algebra is $C^k(M, \mathfrak{g})$ and $H^s(M, \mathfrak{g})$ respectively with bracket

$$[\xi, \eta](x) = [\xi(x), \eta(x)],\ x \in M,$$

the bracket on the right-hand side being taken in $\mathfrak{g}$. For example, if $s > 3/2$, then $H^s(\mathbb{R}^3, S^1)$ is an abelian Banach Lie group under pointwise multiplication, with Lie algebra $H^s(\mathbb{R}^3)$, $[\cdot\,,\cdot] = 0$. If G is a compact Lie group with Lie algebra $\mathfrak{g}$ and $s > n/2$, then $H^s(\mathbb{R}^n, G)$ has Lie algebra $H^s(\mathbb{R}^n, \mathfrak{g})$ and bracket $[f, g](x) = [f(x), g(x)]$, the latter bracket being in $\mathfrak{g}$. ♦

Diffeomorphism Groups Among the most important "classical" examples of infinite dimensional groups are the diffeomorphism groups of manifolds. Let M be a compact boundaryless manifold and denote by H^s Diff (M) the set of all H^s diffeomorphisms of M to M, $s > n/2 + 1$.

We will now outline the sense in which H^s Diff (M) is a Lie group with Lie algebra H^s-$\mathfrak{X}(M)$, the H^s vector fields on M. Similar results will be valid for C^k Diff (M) [resp. $W^{s,p}$ Diff (M)], the group of C^k-diffeomorphisms of M [resp. $W^{s,p}$-diffeomorphisms, of M].

The set H^s Diff (M) is a smooth Banach manifold and is a group under composition; explicitly, the group operations are

$$\mu : H^s \text{ - Diff } (M) \times H^s \text{ -Diff } (M) \to H^s \text{ - Diff } (M);\ \mu(f, g) = f \circ g$$

and

$$I : H^s \text{ - Diff } (M) \to H^s\text{ - } D\,\mathrm{Diff}(M);\ I(f) = f^{-1}.$$

The unit element e is the identity map. For $s > (n/2) + 1$, $H^s \operatorname{Diff}(M)$ is a Banach manifold and in fact is an open submanifold of the Banach-manifold $H^s(M, M)$. The condition $s > (n/2)+1$ guarantees that elements of $H^s \operatorname{Diff}(M)$ are C^1 and a map C^1 close to a diffeomorphism is a diffeomorphism by the inverse function theorem (plus an additional argument to guarantee it is globally one to one and onto; see Marsden, Ebin, and Fischer [1972]). The manifold $H^s \operatorname{Diff}(M)$ is not, however, a Banach Lie group, since group multiplication is differentiable only in the following restricted sense. Right multiplication

$$R_g : H^s \operatorname{Diff}(M) \to H^s \operatorname{Diff}(M); R_g(f) = f \circ g$$

is smooth (C^∞) for each $g \in H^s \operatorname{Diff}(M)$, and if $g \in H^{s+k} \operatorname{Diff}(M)$, left multiplication

$$L_g : H^s \operatorname{Diff}(M) \to H^s \operatorname{Diff}(M), L_g(f) = g \circ f$$

is differentiable of class C^k. Hence if $g \in H^s \operatorname{Diff}(M)$, L_g is only C^0, that is, continuous. More generally, composition

$$(f, g) \in H^{s+k} \operatorname{Diff}(M) \times H^s \operatorname{Diff}(M) \mapsto f \circ g \in H^s \operatorname{Diff}(M)$$

is a C^k-map. Therefore,

group multiplication is continuous, but is not smooth.

The inversion map $I : f \mapsto f^{-1}$ is continuous when regarded as a map of H^s to H^s, but is C^k if regarded as a map of $H^{s+k} \operatorname{Diff}(M)$ to $H^s \operatorname{Diff}(M)$.

The tangent space $T_f(H^s \operatorname{Diff}(M))$ at $f \in H^s \operatorname{Diff}(M)$ is the space of H^s vector fields along f; explicitly, elements of $T_f(H^s \operatorname{Diff}(M))$ are given by

$$T_f(H^s\text{-}\operatorname{Diff}(M)) = \{X_f : M \to TM \mid X_f \text{ is } H^s \text{ and } \tau_M \circ X_f = f\},$$

where $\tau_M : TM \to M$ denotes the tangent bundle projection. In particular, for $f = e = \mathrm{id}_M$,

$$\begin{aligned} T_e(H^s\text{-}\operatorname{Diff}(M)) &= \{X : M \to TM \mid X \text{ is } H^s \text{ and } \tau_M \circ X = \mathrm{id}_M\} \\ &= H^s(TM). \end{aligned}$$

The idea behind the above assertions is as follows. An element of the tangent space at the point f, namely, $T_f(H^s \operatorname{Diff}(M))$ is the tangent vector to a curve $f(t) \in H^s \operatorname{Diff}(M)$ at $t = 0$ where $f(0) = f$. But each $f(t)$ maps M to M, so for $x \in M$, $f(t)(x)$ is a curve in M. Thus $(d/dt) f(t)(x)|_{t=0} \in T_{f(x)}M$ and so we get a map X_f of M to TM taking x to an element of $T_{f(x)}M$. In particular, the tangent space at the identity of e is the space of H^s vector fields on M. The tangent manifold $T(H^s \operatorname{Diff}(M))$ can be identified with the set of all mappings from M to TM that cover diffeomorphisms and it is again an infinite dimensional Banach manifold. It can be shown that the map

$$TR : H^{s+k}(TM) \times H^s \text{ Diff } (M) \to T(H^s \text{ Diff } (M)),$$

defined by $TR(f, X) = T_e R_f(X)$ is a C^k map.

A vector field X on $H^s \operatorname{Diff}(M)$ is a map

$$X : H^s \operatorname{Diff}(M) \to T(H^s \operatorname{Diff}(M))$$

such that

$$X(f) \in T_f(H^s \operatorname{Diff}(M)) \quad \text{for all} \quad f \in H^s \operatorname{Diff}(M).$$

It is ***right invariant*** if $(R_g)_* X = X$, for any $g \in H^s \operatorname{Diff}(M)$; similarly it is ***left invariant*** if $(L_g)_* X = X$. It is not hard to see that $TR_g = R_g$ and that $TL_g = L_{Tg}$; explicitly, TR_g and

$$TL_g : T(H^s \operatorname{Diff}(M) \to (H^s \operatorname{Diff}(M))$$

at $X_f \in T_f(H^s \operatorname{Diff}(M))$ are given by

$$TR_g(X_f) = X_f \circ g \in T_{f\circ g}(H^s \operatorname{Diff}(M))$$

and

$$TL_g(X_f) = Tg \circ X_f \in T_{g\circ f}(H^s \operatorname{Diff}(M)).$$

The diagrams in Figure N9.G.1 may help to clarify the situation: in them, $TR_g(X_f) = X_f \circ g$ is a vector field along $f \circ g$ and $TL_g(X_f) = Tg \circ X_f$ is a vector field along $g \circ f$.

$$\begin{tikzcd}
& TM \arrow[dr, "\tau_M"] & \\
M \arrow[ur, "TR_g(X_f)"] \arrow[r, "g"] & M \arrow[u, "X_f"] \arrow[r, "f"] & M
\end{tikzcd}
\qquad
\begin{tikzcd}
TM \arrow[rr, "Tg"] & & TM \arrow[d, "\tau_M"] \\
M \arrow[u, "X_f"] \arrow[urr, "TL_g(X_f)"] \arrow[r, "f"'] & M \arrow[r, "g"'] & M
\end{tikzcd}$$

FIGURE N9.G.1. Mapping vector fields over maps.

As in finite dimensions, the spaces $\mathfrak{X}_R(H^s \operatorname{Diff}(M))$ and $\mathfrak{X}_L(H^s \operatorname{Diff}(M))$ of right and left invariant vector fields on $H^s \operatorname{Diff}(M)$ are isomorphic as vector spaces to $T_e(H^s \operatorname{Diff}(M)) = H^s(TM)$. The isomorphisms are given by $\xi \in H^s(TM) \mapsto Y_\xi \in \mathfrak{X}_R(H^s \operatorname{Diff}(M)$, where

$$Y_\xi(f) := \xi \circ f \in T_f(H^s\text{-}\operatorname{Diff}(M)),$$

and $\xi \in H^s(TM) \mapsto X_\xi \in \mathfrak{X}_L(H^s \operatorname{Diff}(M))$, where

$$X_\xi(f) := Tf \circ \xi \in T_f(H^s\text{-}\operatorname{Diff}(M)).$$

Note that $X_\xi(e) = \xi$ and $Y_\xi(e) = \xi$. See Figure N9.G.2.

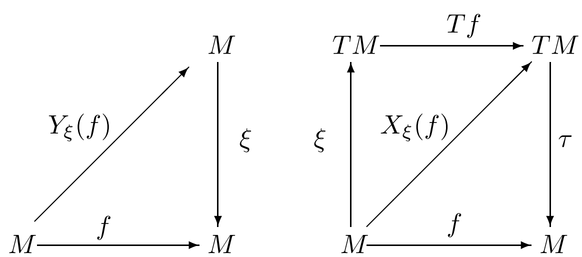

Figure N9.G.2. Left and right invariant vector fields on Diff.

Proposition N9.G.1. *Let $\xi_1, \xi_2 \in T_e(H^s \operatorname{Diff}(M)) = H^s(TM)$. Then for the corresponding right invariant vector fields Y_{ξ_1}, Y_{ξ_2} on $H^s \operatorname{Diff}(M)$ we have*

$$[Y_{\xi_1}, Y_{\xi_2}](e) = [\xi_1, \xi_2], \quad i.e., \quad [Y_{\xi_1}, Y_{\xi_2}] = Y_{[\xi_1, \xi_2]}.$$

Proof. Recall that the Lie bracket $[X, Y]$ of two vector fields X and Y on $H^s \operatorname{Diff}(M)$ is given by

$$[X, Y] = \frac{d}{dt}(F_t^* Y)\bigg|_{t=0}$$

where F_t is the flow of X. One checks that the flow of Y_{ξ_1} is given by $F_t(\eta) = \varphi_t \circ \eta = L_{\varphi_t}(\eta)$ where φ_t is the flow of ξ_1 on M. Then

$$\begin{aligned}
\frac{d}{dt}(F_t^* Y_{\xi_2})(\eta)\bigg|_{t=0} &= \frac{d}{dt}(TF_{-t} \circ Y_{\xi_2} \circ F_t)(\eta)\bigg|_{t=0} \\
&= \frac{d}{dt}(T\varphi_{-t} \circ \xi_2 \circ \varphi_t \circ \eta)\bigg|_{t=0} \\
&= \frac{d}{dt}(\varphi_t^* \xi_2) \circ \eta\bigg|_{t=0} = [\xi_1, \xi_2] \circ \eta = Y_{[\xi_1, \xi_2]}(\eta).
\end{aligned}$$

■

For the corresponding *left* invariant vector fields X_{ξ_1}, X_{ξ_2} on $H^s \operatorname{Diff}(M)$, there is a sign change:

$$[X_{\xi_1}, X_{\xi_2}](e) = -[\xi_1, \xi_2], \quad i.e., \quad [X_{\xi_1}, X_{\xi_2}] = -X_{[\xi_1,\xi_2]}.$$

Note that the bracket $[\xi_1, \xi_2]$ is the ordinary Lie bracket of the vector fields ξ_1, ξ_2 on the manifold M. Due to this fact, we define the right *"Lie" algebra* of the "Lie" group $H^s \operatorname{Diff}(M)$ to be the space of *right* invariant vector fields on $H^s \operatorname{Diff}(M)$.

> *Thus the usual bracket of vector fields on M is the right "Lie" algebra bracket. The Lie algebra bracket associated to the conventional left invariant definition is the negative of the usual Jacobi-Lie bracket of vector fields.*

Note that for $\xi_1, \xi_2 \in H^s(TM)$, $[\xi_1, \xi_2] \in H^{s-1}(TM)$ so one derivative is lost. Hence $[Y_{\xi_1}, Y_{\xi_2}]$ is an H^{s-1}-vector field, and the "Lie" algebra is not closed under the bracket. This corresponds to the fact that $H^s \operatorname{Diff}(M)$, and likewise $C^k \operatorname{Diff}(M)$ are *not* Banach Lie groups.

For $\xi \in H^s(TM)$ let $\varphi_t \in (H^s \operatorname{Diff}(M))$ be its flow. (That φ_t is H^s if ξ is H^s is proved in Ebin and Marsden [1970]; cf. Abraham, Marsden, and Ratiu [1988], Supplement 4.1C.) The curve $c : \mathbb{R} \to H^s \operatorname{Diff}(M)$, $c(t) = \varphi_t$, is an integral curve of the *right* invariant vector field Y_ξ on $H^s \operatorname{Diff}(M)$. Indeed, $c(0) = \varphi_0 = e$, and for $x \in M$,

$$\frac{d}{dt} c(t)(x) = \frac{d}{dt} \varphi_t(x) = \xi(\varphi_t(x)) = (\xi \circ c(t))(x) = Y_\xi(c(t))(x).$$

In particular, note that $c'(0) = \xi$. Thus the *exponential map*

$$\exp : T_e(H^s\text{-}\operatorname{Diff}(M)) = H^s(TM) \to H^s \operatorname{Diff}(M)$$

is given by $\exp(\xi) = \varphi_1$. The map exp is continuous, but unlike the case for Banach Lie groups, it is *not* C^1; in fact, there is no neighborhood of the identity onto which it maps surjectively (Kopell [1970]). As a result in this direction, let us prove that *if a diffeomorphism of S^1 has no fixed points and is in the image of the exponential map, then it must be conjugate to a rotation* (Hamilton [1982]). Let $\eta \in \operatorname{Diff}(S^1)$ have no fixed points. If there is $\xi \in \mathfrak{X}(S^1)$ such that $\exp(\xi) = \eta$, then ξ is nowhere vanishing. Write $\xi = f(t)(d/dt)$, where $t \in \mathbb{R}$ is a parameterization of S^1 modulo 2π. Now reparametrize the circle by

$$\theta = c \int \frac{dt}{f(t)} \quad \text{where} \quad c = \frac{1}{\int_0^{2\pi} \frac{dt}{f(t)}}$$

and note that $\xi = c(d/d\theta)$, *i.e.*, ξ is constant in the parameterization θ and therefore its exponential is given by $\theta \mapsto \theta + c$, *i.e.*, it is a rotation. Since the parameter change $t \mapsto \theta$ is given by a diffeomorphism of S^1, it follows that $\exp(\xi)$ is conjugate to the rotation $\theta \mapsto \theta + c$, which is what we claimed. These facts are important pathologies to keep in mind, but fortunately they will not impair our main development. In our applications to fluid dynamics and plasma physics, various subgroups of $\mathrm{Diff}(M)$ will appear, which we will consider later. In view of the pathologies mentioned above, we cannot invoke Proposition 9.1.14 to prove that they are Lie subgroups — other special arguments are needed.

Subgroups. In the same sense as $H^s \,\mathrm{Diff}(M)$, there are other diffeomorphism groups that are Lie groups. We shall review a few of them here, following Ebin and Marsden [1970], §6 (see Marsden, Ratiu, and Shkoller [1999] for additional examples). Let $s > \dim(M)/2 + 1$. If M is a compact boundaryless manifold and $N \subset M$ is a closed submanifold (possibly zero dimensional) without boundary, let

$$H^s \,\mathrm{Diff}_N(M) = \{ f \in H^s \text{ Diff } (M) \mid f(N) \subset N \} ,$$

i.e., the diffeomorphisms keeping N *setwise* fixed. Arguing as in the case of $H^s \,\mathrm{Diff}(M)$, one sees that the Lie algebra of $H^s \,\mathrm{Diff}_N(M)$ is

$$H^s_N(TM) = \{ X \in H^s(TM) \mid X(n) \in T_n N \text{ for all } n \in N \} ,$$

the H^s vector fields on M tangent to N. Indeed, if $f(t) \in H^s \,\mathrm{Diff}_N(M)$ is a curve with $f(0) = e$, $f(t)(n) \in N$ and so $(d/dt)|_{t=0} f(t)(n) \in T_n N$.

Similarly, one can consider the group of H^s-diffeomorphisms keeping the submanifold N *pointwise* fixed, *i.e.*,

$$H^s\text{-}\mathrm{Diff}_{N,p}(M) = \{ f \in H^s \text{-Diff } (M) \mid f(n) = n \quad \text{for all } n \in N \} .$$

As before, it is easy to see that the Lie algebra of this groups is

$$H^s_{N,p}(TM) = \{ X \in H^s(TM) \mid X(n) = 0 \text{ for all } n \in N \} .$$

Now let M be a compact manifold with boundary and consider the group $H^s \,\mathrm{Diff}(M)$. Since $f(\partial M) = \partial M$ for every diffeomorphism f, the previous argument used for the Lie algebra $H^s_N(TM)$ shows that the Lie algebra here is

$$H^s_\partial(TM) = \{ X \in H^s(TM) \mid X(x) \in T_x(\partial M) \text{ for all} x \in \partial M \} ,$$

the H^s vector fields tangent to the boundary. This group and its Lie algebra are useful in the continuum mechanics of compressible fluids. If $N \cap \partial M = \varnothing$, the groups $H^s \,\mathrm{Diff}_N(M)$ and $H^s \,\mathrm{Diff}_{N,p}(M)$ have Lie algebras equal to

$$H^s_N(TM) \cap H^s_\partial(TM) \quad \text{and} \quad H^s_{N,p}(TM) \cap H^s_\partial(TM),$$

respectively. Similarly

$$H^s \,\mathrm{Diff}_p(M) = \{ f \in H^s \text{ Diff } (M) \mid f(x) = x \text{ for all } x \in \partial M \}$$

has Lie algebra

$$H^s_p(TM) = \{x \in H^s(TM) \mid X(x) = 0 \text{ for all } x \in \partial M\}\,.$$

For a manifold with boundary ∂M, the ***boundaryless double*** $\tilde{M}$ is obtained by gluing together two copies of M along the boundary. Then $\tilde{M}$ is a boundaryless smooth manifold, $\dim(\tilde{M}) = \dim(M)$, and $\tilde{M}$ is compact if M is. One checks that $H^s\operatorname{Diff}(M)$ is a submanifold of $H^s(M, \tilde{M})$ and $H^s\operatorname{Diff}_p(M)$ is a submanifold of both $H^s\operatorname{Diff}(M)$ and $H^s(M, \tilde{M})$. Using $\tilde{M}$ there is yet another group that often shows up in the literature. A diffeomorphism $f \in H^s\operatorname{Diff}(M)$ is said to have ***support*** in M if and only if f can be extended to $\tilde{f} \in H^s\operatorname{Diff}(\tilde{M})$ with $\tilde{f}|(\tilde{M}\backslash M) =$ identity. Let $H^s\operatorname{Diff}_0(M)$ denote the H^s-diffeomorphisms with support in M. Then for $s > (\dim M)/2 + 1$, the embedding

$$f \in H^s\operatorname{Diff}_0(M) \mapsto \tilde{f} \in H^s\operatorname{Diff}(\tilde{M})$$

makes $H^s\operatorname{Diff}_0(M)$ a closed submanifold of $H^s\operatorname{Diff}(\tilde{M})$. The Lie algebra of $H^s\operatorname{Diff}_0(M)$ is

$$H^s_0(TM) = \Big\{ X \in H^s(TM) \text{ there exists an } H^s\text{-extension} \\ \tilde{X} \in H^s(\tilde{T}M) \text{ with } X \text{ zero on } \tilde{M}\backslash M \Big\}.$$

N9.G.1 Basic Facts about Sobolev Spaces and Manifolds.

In our discussions of $H^s(M)$ and $H^s\operatorname{Diff}(M)$, we implicitly used some basic properties of Sobolev spaces. We summarize these here. More details may be found in Adams [1975], Palais [1965, 1968], Ebin and Marsden [1970] and in Marsden and Hughes [1983] (Ch. 6).

A fundamental point is that functional analysis requirements necessitate the use of Sobolev spaces where C^k spaces will *not do*. For example if the Laplacian of f, namely, Δf is C^k, it does *not* follow that f is C^{k+2}. However, its Sobolev analogue *is* true.

These spaces will first be defined over a domain $\Omega \subset \mathbb{R}^n$ and are vector subspaces of various L^p spaces. For any integer $k \geq 0$ and $1 \leq p < \infty$ we define a norm $\|\cdot\|_{k,p}$ on real or complex valued functions on Ω as follows

$$\|F\|_{k,p} = \left[\sum_{|r|=0}^{k} \|D^k f\|_p^p\right]^{1/p}$$

for any function f for which the right side makes sense, $\|\cdot\|_p$ being the $L^p(\Omega)$-norm. It is clear that $\|\cdot\|_{k,p}$ defines a norm on any vector space of functions on which the right side takes finite values, provided that two functions are identified in the space if they are equal almost everywhere in Ω. Let $H^{k,p}(\Omega)$ be the completion of

$$\left\{ f \in C^k(\Omega) \mid \|f\|_{k,p} < \infty \right\}$$

with respect to the norm $\|\cdot\|_{k,p}$. Then $H^{k,p}(\Omega)$ is a Banach space called a ***Sobolev space***. For $p = 2$ we denote $H^{k,2}(\Omega)$ by $H^k(\Omega)$.

For the reader familiar with distributional derivatives, another definition may be helpful. Let

$$W^{k,p}(\Omega) = \{ f \in L^p(\Omega) \mid D^r f \in L^p(\Omega) \text{ for } 0 \leq |r| \leq k \}$$

where $D^r f$ is the distributional rth derivative. Equipped with the norm above, these are Banach spaces. Clearly $W^{0,p}(\Omega) = L^p(\Omega)$. The ***Meyers-Serrin theorem*** (Meyers and Serrin [1960]) states that $H^{k,p}(\Omega) = W^{k,p}(\Omega)$.

The definition of $W^{s,p}(\Omega)$ for an arbitrary $s \geq 0$, *i.e.*, s not an integer is more complicated. Let $s = k + \sigma$ where k is an integer and $0 < \sigma < 1$. Then the norm $\|\cdot\|_{s,p}$ is defined by

$$\|f\|_{s,p} = \left\{ \|f\|_{k,p}^p + \sum_{|r|=k} \int_\Omega \int_\Omega \frac{|D^r f(x) - D^r f(h)|^p}{|x-y|^{n+\sigma p}} \right\}^{1/p} .$$

Let $W^{s,p}(\Omega)$ denote the completion of $\{ f \in C^\infty(\Omega) \mid \|f\|_{s,p} < \infty \}$ and set $H^s(\Omega) \equiv W^{s,2}(\Omega)$. One treats $W^{s,p}(\Omega, \mathbb{R}^m)$ in a similar way. Also, by completing the space of functions of Ω that extend to C^∞ functions in an open neighborhood of the closure $\bar{\Omega}$ one similarly defines $W^{s,p}(\bar{\Omega})$ and $W^{s,p}(\bar{\Omega}, \mathbb{R}^m)$.

Now we turn to Sobolev spaces on manifolds. Let M be a compact C^∞ manifold, possibly with boundary, and let $C^\infty(M)$ be the set of real-valued C^∞ functions on M. For s and p a pair of positive integers, a Sobolev norm on $f \in C^\infty(M)$ is defined as follows: let $\{(U_i, \varphi_i)\}_{0 \leq i \leq N}$ be a finite atlas of M, $f_i = f \circ \varphi_i^{-1} : \varphi_i(U_i) \subset \mathbb{R}^n \to \mathbb{R}$ be the local representatives of f, and set

$$\|f\|_{s,p} := \max_{0 \leq i \leq N} \|f_i\|_{s,p}$$

where $\|\cdot\|_{s,p}$ are the Sobolev norms on $\varphi_i(U_i) \subset \mathbb{R}^n$ defined as above (respectively on the closure $\varphi_i(U_i)$ in $\mathbb{R}^n_+$ at the boundary). The Sobolev space $W^{s,p}(M)$ is then defined to be the (Cauchy) completion of $C^\infty(M)$ with respect to $\|\cdot\|_{s,p}$. One shows that the resulting *space* is independent of the choice of atlas on M, but, of course, its *norm* is not. To get an intrinsically defined norm one requires additional structure on M, such as a Riemannian metric. The generalization of this definition to $\mathbb{R}^n$-valued functions on M, and further to sections of a vector bundle $E \to M$ should be clear (presuming the existence of a vector bundle norm on E).

To obtain useful information concerning the Sobolev spaces $W^{k,p}$, we need to establish certain fundamental relationships between these spaces. To do this, one uses the following fundamental inequality of Sobolev, as generalized by Nirenberg and Gagliardo.

Theorem N9.G.2 (SNG Inequality). *Let $1 \leq q \leq \infty$, $0 \leq r \leq \infty$, $0 \leq j < m$, $j/m \leq a \leq 1$, $0 < p < \infty$, with j, m integers ≥ 0; assume that*

$$\frac{1}{p} = \frac{j}{n} + a\left(\frac{1}{r} - \frac{m}{n}\right) + (1-a)\,\frac{1}{q} \tag{N9.G.1}$$

(if $1 < r < \infty$ and $m - j - (n/r)$ is an integer ≥ 0, assume $(j/m) \leq a < 1$). Then there is a constant C such that for any smooth $u : \mathbb{R}^n \to \mathbb{R}^1$, we have

$$\|D^j u\|_{L^p} \leq C\|D^m u\|_{L^r}^a \|u\|_{L^q}^{1-a}. \tag{N9.G.2}$$

(If $j = 0$, $rm < n$, and $q = \infty$, assume $u \to 0$ at ∞ or u lies in L^σ for some finite $\sigma > 0$.)

Below we shall prove some special cases of this result. (The arguments given by Nirenberg [1959] are geometric in flavor in contrast to the usual Fourier transform proofs and therefore are more suitable for generalization to manifolds; cf. Cantor [1975] and Aubin [1976].)

The above theorem remains valid for u defined on a region with piecewise smooth boundary, or more generally if the boundary satisfies a certain "cone condition."

If one knows an inequality of the *form* (N9.G.2) exists, one can infer that (N9.G.1) must hold by the following scaling argument: Replace $u(x)$ by $u(tx)$ for a real $t > 0$. Then writing $u_t(x) = u(tx)$, one has

$$\begin{aligned}
\|D^j u_t\|_{L^p} &= t^{j-n/p}\|D^j u\|_{L^p}, \\
\|D^m u_t\|_{L^r}^a &= t^{a(m-n/r)}\|D^m u\|_{L^r}^a, \\
\|u_t\|_{L^q}^{1-a} &= t^{-n(1-a)/q}\|u\|_{L^q}^{1-a}.
\end{aligned}$$

Thus if (N9.G.2) is to hold for u_t (with the constant independent of t), we must have

$$j - \left(\frac{n}{p}\right) = a\left(m - \left(\frac{n}{r}\right)\right) - n\left(\frac{(1-a)}{q}\right),$$

which is the relation (N9.G.1).

The following corollary is useful in a number of applications:

Corollary N9.G.3. *With the same relations as in the SNG Inequality, for any $\varepsilon > 0$ there is a constant K_ε such that*

$$\|D^j u\|_{L^p} \leq \varepsilon\|D^m u\|_{L^r} + K_\varepsilon\|u\|_{L^q}$$

for all (smooth) functions u.

Proof. This follows from the SNG inequality (N9.G.2) and ***Young's inequality***:

$$x^a y^{1-a} \leq ax + (1-a)y,$$

which implies that

$$x^a y^{1-a} = (\varepsilon x)^a (K_\varepsilon y)^{1-a} \leq a\varepsilon x + (1-a)K_\varepsilon y$$

where $K_\varepsilon = 1/\varepsilon^{a/(1-a)}$. ■

Let us *illustrate* how Fourier transform techniques can be used to directly prove the special case of the preceding Corollary in which $n = 3$, $j = 0$, $p = \infty$, $m = 2$, $r = 2$, and $q = 2$.

Proposition N9.G.4. *There is a constant $c > 0$ such that for any $\varepsilon > 0$ and function $f : \mathbb{R}^3 \to \mathbb{R}$ smooth with compact support, we have*

$$\|f\|_\infty \leq c\left(\varepsilon^{3/2}\|f\|_{L^2} + \varepsilon^{-1/2}\|\Delta f\|_{L^2}\right).$$

(It follows that if $f \in H^2(\mathbb{R}^3)$, then f is uniformly continuous and the above inequality holds.)

Proof. Let

$$\hat{f}(k) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} e^{-ik\cdot x} f(x)dx$$

denote the Fourier transform. Recall that $(\Delta \hat{f})(k) = -\|k\|^2 \hat{f}(k)$. From Schwarz' inequality, we have

$$\begin{aligned}\left(\int \left|\hat{f}(k)\right| dk\right)^2 &\leq \left(\int \frac{dk}{\left(\varepsilon^2 + \|k\|^2\right)^2}\right)\left(\int \left(\varepsilon^2 + \|k\|^2\right)^2 \left|\hat{f}(k)\right|^2 dk\right)\\ &= \frac{c_1}{\varepsilon}\|\left(\varepsilon^2 - \Delta\right) f\|_{L^2}^2,\end{aligned}$$

where

$$c_1 = \int_{\mathbb{R}^3} \frac{d\xi}{\left(1 + \|\xi\|^2\right)^2} < \infty.$$

Here we have used the fact that $h \mapsto \hat{h}$ is an isometry in the L^2-norm (Plancherel's theorem). Thus, from $f(x) = 1/(2\pi)^{3/2} \int_{\mathbb{R}^3} e^{ik\cdot x} \hat{f}(k)dk$, we get

$$\begin{aligned}(2\pi)^{3/2}\|f\|_\infty \leq \|\hat{f}\|_{L^1} &\leq \frac{c_2}{\sqrt{\varepsilon}}\|\left(\varepsilon^2 - \Delta\right) f\|_{L^2}\\ &\leq c_2\left(\varepsilon^{3/2}\|f\|_{L^2} + \varepsilon^{-1/2}\|\Delta f\|_{L^2}\right).\end{aligned}$$ ■

Thus we have shown that $H^2(\mathbb{R}^3) \subset C^0(\mathbb{R}^3)$ and that the inclusion is continuous. More generally, one can show by similar arguments that $H^2(\Omega) \subset C^k(\Omega)$ provided $s > (n/2) + k$ and

$$W^{s,p}(\Omega) \subset C^k(\Omega) \text{ if } s > (n/p) + k.$$

This is one of the ***Sobolev embedding theorems***.

For Ω bounded, the inclusion $W^{s,p}(\Omega) \to C^k(\Omega)$, $s > (n/p) + k$ is compact; that is, the ball in $W^{s,p}(\Omega)$ is compact in $C^k(\Omega)$ (***Rellich's theorem***). This is proved in a manner similar to the classical Arzela-Ascoli theorem, one version of which states that the inclusion $C^1(\Omega) \subset C^0(\Omega)$ is compact (see Marsden and Hoffman [1993], for instance). Also, $W^{s,p}(\Omega) \subset W^{s',p'}(\Omega)$ is compact if $s > s'$ and $p = p'$ or if $s = s'$ and $p > p'$. (See Friedman [1969] for the proofs.)

One application of Rellich's theorem is to the proof of the Fredholm alternative. (See, for example, Marsden and Hughes [1983] (Chapter 6).) It is often used in this way in existence theorems, using compactness to extract convergent sequences. Compactness is also used in existence theory in another crucial way when one seeks weak solutions. This is through the fact that the unit ball in a Banach space is weakly compact — that is, compact in the weak topology. See, for example, Yosida [1980] for the proof (and for refinements, involving weak *sequential* compactness).

We shall give another illustration of the SNG Inequality through a special case that is useful in the study of, amongst other things, nonlinear wave equations. This is the following inequality in $\mathbb{R}^3$:

$$\|u\|_{L^6} \leq C\|\operatorname{grad} u\|_{L^2}.$$

Proposition N9.G.5. *Let $u : \mathbb{R}^3 \to \mathbb{R}$ be smooth and have compact support. Then*

$$\int_{\mathbb{R}^3} u^6 dx \leq 48 \left(\int_{\mathbb{R}^3} \|\operatorname{grad} u\|^2 dx \right)^3$$

so $C = \sqrt[6]{48}$.

Proof. (Following Ladyzhenskaya [1969].) From

$$u^3(x,y,z) = 3 \int_{-\infty}^{x} u^2 \frac{\partial u}{\partial x} dx$$

one gets

$$\sup_x \left| u^3(x,y,z) \right| \leq 3 \int_{-\infty}^{\infty} \left| u^2 \frac{\partial u}{\partial x} \right| dx.$$

Set $I = \int_{\mathbb{R}^3} u^6 dx$ and write

$$\begin{aligned}
I &= \int_{-\infty}^{\infty} \left(\iint \left|u^3\right| \left|u^3\right| dydz \right) dx \\
&\leq \int_{-\infty}^{\infty} \left[\left(\sup_y \int_{-\infty}^{\infty} \left|u^3\right| dz \right) \left(\int_{-\infty}^{\infty} \sup_x \left|u^3\right| dy \right) \right] dx \\
&\leq 9 \int_{-\infty}^{\infty} \left[\left(\iint \left| u^2 \frac{\partial u}{\partial y} \right| dydz \right) \left(\iint \left| u^2 \frac{\partial u}{\partial z} \right| dydz \right) \right] dx.
\end{aligned}$$

Using Schwarz's inequality gives

$$\begin{aligned}
I &\leq 9 \int_{-\infty}^{\infty} \left[\left(\iint u^4 dydz \right) \left(\iint_{-\infty}^{\infty} \left(\frac{\partial u}{\partial y} \right)^2 dydz \right)^{1/2} \right. \\
&\quad \times \left. \left(\iint_{-\infty}^{\infty} \left(\frac{\partial u}{\partial z} \right)^2 dydz \right)^{1/2} \right] dx \\
&\leq 9 \max_x \left[\left(\iint_{-\infty}^{\infty} u^4 dydz \right) \left(\int_{\mathbb{R}^3} \left(\frac{\partial u}{\partial y} \right)^2 dxdydz \right)^{1/2} \right. \\
&\quad \times \left. \left(\int_{\mathbb{R}^3} \left(\frac{\partial u}{\partial z} \right)^2 dxdydz \right)^{1/2} \right] \\
&\leq 36 \left(\int_{\mathbb{R}^3} \left| u^3 \frac{\partial u}{\partial x} \right| dxdydz \right) \left(\int_{\mathbb{R}^3} \left(\frac{\partial u}{\partial y} \right)^2 dxdydz \right)^{1/2} \\
&\quad \times \left(\int_{\mathbb{R}^3} \left(\frac{\partial u}{\partial z} \right)^2 dxdydz \right) \\
&\leq 36\sqrt{I} \left(\int_{\mathbb{R}^3} \left(\frac{\partial u}{\partial x} \right)^2 dxdydz \right)^{1/2} \left(\int_{\mathbb{R}^3} \left(\frac{\partial u}{\partial y} \right)^2 dxdydz \right)^{1/2} \\
&\quad \times \left(\int_{\mathbb{R}^3} \left(\frac{\partial u}{\partial z} \right)^2 dxdydz \right)^{1/2}.
\end{aligned}$$

Using the ***arithmetic–geometric mean inequality***

$$\sqrt[3]{a}\sqrt[3]{b}\sqrt[3]{c} < (a+b+c)/3$$

gives

$$I \leq 36\sqrt{I} \left(\int_{\mathbb{R}^3} \|\operatorname{grad} u\|^2 \right)^{3/2} \frac{1}{3^{3/2}};$$

i.e.,

$$I \leq \frac{(36)^2}{3^3}\left(\int_{\mathbb{R}^3} \|\operatorname{grad} u\|^2\right)^3. \qquad \blacksquare$$

Another important corollary of the SNG Inequality can be used to determine to which $W^{s,p}$ space a product belongs.

Corollary N9.G.6. *For $s > n/2$, $H^s(\mathbb{R}^n)$ is a Banach algebra (under pointwise multiplication). That is, there is a constant $K > 0$ such that for $u, v \in H^s(\mathbb{R}^n)$,*

$$\|u \cdot v\|_{H^s} \leq K\|u\|_{H^s}\|v\|_{H^s}.$$

This is an important property of H^s not satisfied for low s; it is *not* true that L^2 forms an algebra under multiplication.

Proof. Choose in the SNG inequality, $a = j/s$, $r = 2$, $q = \infty$, $p = 2s/j$, $m = s$ $(0 \leq j \leq s)$ to obtain

$$\|D^j u\|_{L^{2s/j}} \leq \text{const.}\, \|D^s u\|_{L^2}^{j/s}\|u\|_{\infty}^{1-j/s} \leq \text{const.}\, \|u\|_{H^s}.$$

Let $j + k = s$. From Hölder's inequality we have

$$\|D^j u \cdot D^k v\|_{L^2}^2 \leq \text{const.}\, \|D^j u\|_{L^{2s/j}}^2 \|D^k v\|_{L^{2s/k}}^2 \leq \text{const.}\, \|u\|_{H^s}^2 \|v\|_{H^s}^2.$$

Now $D^s(uv)$ consists of terms like $D^j u \cdot D^k v$, so we obtain

$$\|D^s(uv)\|_{L^2} \leq \text{const.}\, \|u\|_{H^s}\|v\|_{H^s}.$$

Similarly for the lower-order terms. Summing gives the result. $\blacksquare$

The ***trace theorems*** state that the restriction map from Ω to a submanifold $M \subset \Omega$ of codimension m induces a bounded operator from $W^{s,p}(\Omega)$ to $W^{s-(1/mp),p}(M)$. Adams [1975] and Morrey [1966] are good references; the latter contains some useful refinements.

There are also basic extension theorems that are right inverses of restriction maps. For example, the ***Calderon extension theorem*** asserts that there is an extension map $T : W^{s,p}(\Omega) \to W^{s,p}(\mathbb{R}^n)$ that is a bounded operator and "restriction to Ω"$\circ T =$ Identity. This is related to a classical C^k theorem due to Whitney. See, for example, Abraham and Robbin [1967], Stein [1970], and Marsden [1973a]. Finally, we mention that these $\mathbb{R}^n$ results carry over to manifolds in a straightforward way using local charts, as in for example, Palais [1965].

Consider the set of all strictly positive functions in $W^{s,p}(M)$. Clearly this set (which we shall call $W_+^{s,p}(M)$) is not a Sobolev space since it fails to be closed under multiplication by scalars. Elementary properties of continuous functions guarantee that $C_+^k(M)$ is open in $C^k(M)$. This is not generally true for $W_+^{s,p}(M)$ in $W^{s,p}(M)$. However, if $s > n/p$ then $W^{s,p}(M)$ is continuously embedded in $C^0(M)$. Then

$$W_+^{s,p}(M) = W^{s,p}(M) \cap C_+^0(M),$$

from which it follows that $W_+^{s,p}(M)$ is an open subset of $W^{s,p}(M)$. Note that the differentiability, the embedding, and the multiplication properties discussed above all hold (in appropriate form) for open subsets of Sobolev spaces.

We make some additional comments about the group of diffeomorphisms of M. We define the collection of diffeomorphisms of M as open subsets of Sobolev spaces, we discuss (in terms of Sobolev spaces) the composition of diffeomorphisms with scalar functions, and finally we consider the diffeomorphisms as a topological group.

The space of diffeomorphisms of M. A map $\eta : M \to M$ is locally (in a coordinate chart of M) $\mathbb{R}^n$-valued and hence we can define a Sobolev space $W^{s,p}(M, M)$ of such maps. One can check that this space is chart-independent if $s > n/p$ and hence is well-defined. The diffeomorphisms are the invertible elements of $W^{s,p}(M, M)$. Now, the continuous diffeomorphisms, *i.e.*, the homeomorphisms, aren't open in the set of continuous maps $C^0(M, M)$. However, in $C^1(M, M)$, the inverse function theorem is available and one can use it to verify openness. So one finds that if $s > n/p + 1$, the set of diffeomorphisms $W^{s,p}\,\mathrm{Diff}(M)$ is open in the Sobolev space $W^{s,p}(M, M)$.

Composing scalar functions with diffeomorphisms. For $s > (n/p)+1$ let

$$F : (W^{s',p}\text{ -Diff }(M)) \times W^{s'',p}(M) \to W^{s,p}(M),$$

for $s', s'' \geq s$, be defined by

$$F(\eta, f) = f \circ \eta.$$

Then F is C^∞.

Sobolev inverse functions. If $\eta \in H^s(M, M)$, $s > (n/p) + 1$ and η has a C^1 inverse, then the inverse is $W^{s,p}$ so $\eta \in W^{s,p}\,\mathrm{Diff}(M)$.

Now consider two maps related to the composition map F: let

$$F_\eta : W^{s,p}(M) \to W^{s,p}(M),$$

be defined by

$$F_\eta(f) := f \circ \eta$$

for fixed $\eta \in (W^{s',p}\text{ -Diff }(M))$, with $s' \geq s + 1)$ and let

$$O_f : (W^{s',p}\text{ -Diff }(M)) \to W^{s,p}(M),$$

be defined by

$$O_f(\eta) := f \circ \eta$$

for fixed $f \in W^{s'',p}(M)$ with $s' \geq s$ and $s'' \geq s$).

The first of these, F_η, is the pullback map. It is linear since

$$F_\eta(f + g) = (f + g) \circ \eta = f \circ \eta + g \circ \eta = F_\eta(f) + F_\eta(g).$$

Since one verifies that it is also C^0, F_η is a smooth map.

The second map, O_f, is the ***orbit map***; its image of $W^{s,p}(M)$ is called the ***orbit*** of $W^{s,p}\,\text{Diff}(M)$ through f. Since O_f is not linear, smoothness is not so elementary. One does, however, find that if $f \in W^{s+k,p}(M)$, then

$$O_f : W^{s',p}\,\text{Diff}(M) \to W^{s,p}(M)$$

is a C^k map as long as $s' \geq s$. This result is proved using the "ω-lemma" (Abraham, Marsden, and Ratiu [1988], Supplement 2.4B).

$W^{s,p}$-Diff(M) as a topological group. $W^{s,p}$-Diff (M) is a group using composition for the group multiplication. For $\eta \in W^{s,p}\,\text{Diff}(M)$, we set

$$R_\eta : W^{s,p}\text{ -Diff }(M) \to W^{s,p}\text{ -Diff }(M);\quad R_\eta(\mu) = \mu \circ \eta$$

and

$$L_\eta : W^{s,p}\text{ -Diff }(M) \to W^{s,p}\text{ -Diff }(M);\quad L_\eta(\mu) = \eta \circ \mu.$$

To ensure that $W^{s,p}\,\text{Diff}(M)$ is a group, one requires that $s > (n/p) + 1$ so that an inverse exists. One can use the results on composition above to show that while right multiplication is smooth, left multiplication and inversion are only C^0. Hence $W^{s,p}\,\text{Diff}(M)$ is a topological group and *not* a Banach Lie group.

Nested Groups. In view of the important technical properties of group multiplication in $H^s\,\text{Diff}(M)$, we introduce an abstract context for this phenomenon developed by Omori [1974] and Adams, Ratiu, and Schmid [1986].

Definition N9.G.7. *A collection of groups $\{G^\infty, G^s \mid s \geq s_0\}$ is called a* ***nest*** *if*

- **i** *each G^s is a Hilbert manifold of class $C^{k(s)}$, modeled on the Hilbert space E^s, where the order of differentiability $k(s)$ tends to ∞ as $s \to \infty$;*

ii *for each $s \geq s_0$, there are linear continuous, dense inclusions $E^{s+1} \to E^s$ and dense inclusions of class $C^{k(s)}$, $G^{s+1} \to G^s$;*

iii *each G^s is a topological group and $G^\infty = \varprojlim G^s$ is a topological group with the inverse limit topology;*

iv *if (U^s, φ^s) is a chart on G^s, then $(U^s \cap G^t,\ \varphi^s \mid U^s \cap G^t)$ is a chart on G^t for every $t \geq s$;*

v *group multiplication $\mu : G^\infty \times G^\infty \to G^\infty$ can be extended to a C^k-map $\mu : G^{s+k} \times G^s \to G^s$ for any s such that $k \leq k(s)$;*

vi *inversion $I : G^\infty \to G^\infty$ can be extended to a C^k-map $I : G^{s+k} \to G^s$, for any s satisfying $k \leq k(s)$;*

vii *right multiplication R_g by $g \in G^s$ is a $C^{k(s)}$-map $R_g : G^s \to G^s$.*

*If the manifolds are Banach manifolds rather than Hilbert manifolds then $\{G^\infty, G^s \mid s \geq s_0\}$ is a **Banach nest**.*

*A collection of vector spaces $\{\mathfrak{g}^\infty, \mathfrak{g}^s \mid s \geq s_0\}$ is called a **Lie algebra nest** if*

i *each $\mathfrak{g}^s$ is a Hilbert (Banach)-space and for each $s \geq s_0$ there are linear, continuous, dense inclusions $\mathfrak{g}^{s+1} \to \mathfrak{g}^s$ and $\mathfrak{g}^\infty = \varprojlim \mathfrak{g}^s$ is a Fréchet space with the inverse limit topology;*

ii *there are bilinear, continuous, antisymmetric maps $[\,,] : \mathfrak{g}^{s+2} \times \mathfrak{g}^{t+2} \to \mathfrak{g}^{\min(s,t)}$, for all $s, t \geq s_0$, which satisfy the Jacobi identity on $\mathfrak{g}^{\min(s,t,r)}$ for elements in $\mathfrak{g}^{s+4} \times \mathfrak{g}^{t+4}$.*

If $\{G^\infty, G^s \mid s \geq s_0\}$ is a Lie group nest, put $\mathfrak{g}^s \equiv T_e G^s$ and $\mathfrak{g}^\infty = \varprojlim \mathfrak{g}^s$. Then it is easy to see that $\{\mathfrak{g}^\infty, \mathfrak{g}^s \mid s \geq s_0\}$ is the Lie algebra nest of the Lie group nest $\{G^\infty, G^s \mid s \geq s_0\}$.

Examples

A. The classical examples of Lie group nests are the diffeomorphism groups

$$\{\mathrm{Diff}(M), H^s\text{ -Diff }(M) \mid s > (\dim M)/2\}$$

with Lie algebra nests

$$\{\mathfrak{X}(M), H^s\text{-}\,\mathfrak{X}(M) \mid s > (\dim M)/2\}$$

for M a compact manifold. Previously, we stated the facts, proved in Ebin [1970], Ebin and Marsden [1970], Omori [1974] and Marsden, Ebin, and Fischer [1972], which verify the preceding definition. ♦

B. The group of homogeneous symplectomorphisms of $T^*M\backslash\{0\}$.

The symplectic manifold most widely used is the cotangent bundle T^*M of a manifold M. The canonical symplectic form on T^*M is exact, *i.e.*, it is the differential of a canonical one-form θ on T^*M. Thus one can ask about the structure of diffeomorphisms of T^*M that preserve θ. A diffeomorphism $\varphi : T^*M \to T^*M$ satisfying $\varphi^*\theta = \theta$ is necessarily a lift, *i.e.*, $\varphi = T^*\eta$, for $\eta \in \mathrm{Diff}^\infty(M)$ (this is proved in the main text). However, in some cases one must consider $T^*M\backslash\{0\}$, and its diffeomorphisms preserving θ. Then $\varphi^*\theta = \theta$ if and only if φ is symplectic and homogeneous of degree one, *i.e.*, $\varphi(\tau\alpha_m) = \tau\varphi(\alpha_m)$ for all $\tau > 0$ and $\alpha_m \in T^*_m M$. Now we can consider the group $H^s\,\mathrm{Diff}_\theta(T^*M\backslash\{0\})$ of homogeneous symplectomorphisms of $T^*M\backslash\{0\}$. But right away we are faced with the problem of non-compactness of $T^*M\backslash\{0\}$. We sketch below, following Ratiu and Schmid [1981] how

$$\{\mathrm{Diff}_\theta(T^*M\backslash\{0\}), H^s \text{ Diff }_\theta(T^*M\backslash\{0\}) \mid s \geq \dim M + 1/2\}$$

is a Lie group nest with Lie algebra nest

$$\left\{S^\infty(T^*M\backslash\{0\}), S^{s+1}(T^*M\backslash\{0\}) \mid s > \dim M + 1/2\right\},$$

where

$$\begin{aligned} S^s(T^*M\backslash\{0\}) = \{H : T^*M\backslash\{0\} \to \mathbb{R} \mid H \text{ is of}\\ \text{class } H^s \text{ and homogeneous of degree one}\} \end{aligned}$$

with the Poisson bracket as Lie algebra bracket. Note the gain in one derivative at the Lie algebra level. The basic idea of the ensuing discussion is that $H^s\,\mathrm{Diff}_\theta(T^*M\backslash\{0\})$ is algebraically isomorphic to the group of all H^s contact transformations of the cosphere bundle of M, which is a compact manifold if M is. We start by recalling the relevant facts.

The multiplicative group of strictly positive reals $\mathbb{R}_+$ acts smoothly on $T^*M\backslash\{0\}$ by $\alpha_x \mapsto \tau\alpha_x$, $\tau > 0$, $\alpha_x \in T^*_x M$, $\alpha_x \neq 0$. This action is free and proper and therefore $\pi : T^*M\backslash\{0\} \to Q \equiv T^*M\backslash\{0\}/\mathbb{R}_+$ is a smooth principal fiber bundle over Q, the ***cosphere bundle*** of M. Note that Q is compact (supposing M is) and odd-dimensional. Q carries no canonical contact one-form but for each global section $\sigma : Q \to T^*M\backslash\{0\}$ we can define an exact contact one-form θ_σ on Q by $\theta_\sigma = \sigma^*\theta$. Such global sections exist in abundance; for example, any Riemannian metric on M identifies T^*M with TM and Q with the unit sphere bundle. Then the usual inclusion of the sphere bundle into TM gives a section σ. The section σ is uniquely determined by a smooth function $f_\sigma : T^*M\backslash\{0\} \to \mathbb{R}_+$ defined by $\sigma(\pi(\alpha_x)) = f_\sigma(\alpha_x)\alpha_x$. In other words, f_σ measures how far from the section σ an element $\alpha_x \in T^*M\backslash\{0\}$ lies. The function f_σ is homogeneous of degree -1 and $\pi^*\theta_\sigma = f_\sigma\theta$.

An H^{s+1} ***contact transformation*** on Q is a diffeomorphism $\varphi \in H^*\,\mathrm{Diff}_\theta(Q)$ such that for any two sections $\sigma, \zeta : Q \to T^*M\backslash\{0\}$, there exists an H^{s+1} function $h_{\sigma\zeta} : Q \to \mathbb{R}_+$ satisfying $\varphi^*\theta_\sigma = h_{\sigma\zeta}\theta_\zeta$. Equivalently, $\varphi \in H^{s+1}\operatorname{Diff}(Q)$ is an H^{s+1} contact transformation if and only if for each global section σ there exists an H^{s+1} function $h_\sigma : Q \to \mathbb{R}_+$ such that $\varphi^*\theta_\sigma = h_\sigma\theta_\sigma$. The function h_σ is uniquely determined by σ, namely $h_\sigma = \langle \varphi^*\theta_\sigma, E_\sigma\rangle$, where E_σ is the Reeb vector field on Q determined by the contact structure θ_σ and $\langle\,,\rangle$ denotes the pairing between vector fields and one-forms. (E_σ is the unique vector field satisfying $\langle \theta_\sigma, E_\sigma\rangle = 1$ and $\mathbf{i}_{E_\sigma}(\mathbf{d}\theta_\sigma) = 0$, where $\mathbf{i}_{E_\sigma}(\mathbf{d}\theta_\sigma)$ denotes the interior product of E_σ with $\mathbf{d}\theta_\sigma$; in local coordinates $\left(x^1, \ldots, x^{n-1}, y^1, \ldots, y^{n-1}, t\right)$ on Q, where

$$\theta_\sigma = \sum_{i=1}^{n-1} y^i dx^i + dt,$$

we have $E_\sigma = \partial/\partial t$. Therefore the group of H^{s+1}-contact transformations on Q is isomorphic to the group

$$\begin{aligned}&\operatorname{Con}_\sigma^{s+1}(Q)\\ &\quad = \left\{(\varphi, h) \in H^{s+1}\text{ -Diff }(Q) \circledS H^{s+1}(Q, \mathbb{R}\backslash\{0\}) \mid \varphi^*\theta_\sigma = h\theta_\sigma\right\}\end{aligned}$$

for any fixed but arbitrary global section σ, where

$$H^{s+1}\text{-Diff}(Q) \circledS H^{s+p}(Q, \mathbb{R}\backslash\{0\})$$

is the semidirect product of the two Lie groups

$$H^{s+1}\operatorname{Diff}(Q) \quad \text{and} \quad H^{s+1}(Q, \mathbb{R}\backslash\{0\})$$

($H^{s+1}(Q, \mathbb{R}\backslash\{0\})$ regarded as a multiplicative group) with composition law

$$(\varphi_1, h_1)(\varphi_1, h_2) = ((\varphi_1 \circ \varphi_2), h_2(h_1 \circ \varphi_2)).$$

Omori [1974] has shown that $\operatorname{Con}_\sigma^{s+1}(Q)$ is a closed Lie subgroup of the semidirect product Lie group

$$H^{s+1}\operatorname{Diff}(Q) \circledS H^{s+1}(Q, \mathbb{R}\backslash\{0\}).$$

The Lie algebra of this semidirect product group is the semidirect product Lie algebra

$$H^{s+1}\text{-}\mathfrak{X}(Q) \circledS H^{s+1}(Q, \mathbb{R})$$

of H^{s+1}-vector fields and H^{s+1}–functions, with bracket

$$[(X, f), (Y, g)] = ([X, Y], X(g) - Y(f)).$$

The Lie algebra of $\operatorname{Con}_\sigma^{s+1}(Q)$ is

$$\operatorname{con}_\sigma^{s+1}(Q) = \{(Y, g) \in H^{s+1}\text{-}\mathfrak{X}(Q) \circledS H^{s+1}(Q, \mathbb{R}) \mid L_Y\theta_\sigma = g\theta_\sigma\}.$$

In Ratiu and Schmid [1981] (Theorem 4.1) it is shown that the group

$$H^{s+1}\,\mathrm{Diff}_\theta(T^*M\backslash\{0\})$$

is isomorphic (as a group) to the Lie group $\mathrm{Con}_\theta^{s+1}(Q)$. The isomorphism is given by

$$\Phi : H^{s+1}\,\mathrm{Diff}_\theta(T^*M\backslash\{0\}) \to \mathrm{Con}_\theta^{s+1}(Q)\Phi(\eta) = (\varphi, h)$$

where φ is defined by $\varphi \circ \pi = \pi \circ \eta$ and h by

$$h \circ \pi = (f_\sigma \circ \eta)/f_\sigma, \sigma(\pi(\alpha_x)) = f_\sigma(\alpha_x)\alpha_x.$$

The inverse of Φ is given by

$$\Phi^{-1}(\varphi, h) = (\sigma \circ \varphi \circ \pi)/(h \circ x) \cdot f_\sigma.$$

Since $\mathrm{Con}_\sigma^{s+1}(Q)$ and $\mathrm{Con}_\zeta^{s+1}(Q)$ are isomorphic as nested Lie groups, for any two global sections σ and ζ, the isomorphism Φ determines a nested Lie group structure on $H^{s+1}\,\mathrm{Diff}_\theta(T^*M\backslash\{0\})$ which is independent of σ (or independent of the Riemannian metric if σ is induced from such). Furthermore, the Lie algebra

$$H^s\text{-}\mathfrak{X}_\theta(T^*M\backslash\{0\}) = \{Y \in H^s\text{-}\mathfrak{X}(T^*M\backslash\{0\}) \mid \pounds_Y\theta = 0\}$$

of $H^{s+1}\,\mathrm{Diff}_\theta(T^*M\backslash\{0\})$ is isomorphic to

$$\begin{aligned} S^{s+2}(T^*M\backslash\{0\}) &= \{H \in H^{s+2}(T^*M\backslash\{0\}) \\ &= \{H \in H^{s+2}(T^*M\backslash\{0\}, \mathbb{R}) \mid \\ &\qquad\qquad H \text{ homogeneous of degree one }\}. \end{aligned}$$

This is because $\pounds_Y\theta = 0$ if and only if Y is globally Hamiltonian, homogeneous of degree zero, with Hamiltonian function $H = \theta(Y)$, homogeneous of degree one. The Lie algebras $H^{s+1}\text{-}\mathfrak{X}_\theta(T^*M\backslash\{0\})$ and $\mathrm{con}_\sigma^{s+1}(Q)$ are isomorphic via

$$T_e\Phi : H^{s+1}\text{-}\mathfrak{X}_\theta(T^*M\backslash\{0\}) \to \mathrm{con}_\sigma^{s+1}(Q).$$

Explicitly, $T_e\Phi(X_H) = (X, k)$, where X is uniquely defined by $T\pi \circ X_H = X \circ \pi$ and k by $k \circ \pi = \{f_\sigma, H\}/f_\sigma$, ($\{\,,\}$ is the canonical Poisson bracket on T^*M). The map $H \mapsto H \circ \sigma$ is an isomorphism from $S^{s+2}(T^*M\backslash\{0\})$ onto $H^{s+2}(Q, \mathbb{R})$, with inverse

$$j : H^{s+2}(Q, \mathbb{R}) \to S^{s+2}(T^*M\backslash\{0\}),$$

where $j(f)$ is the extension to $T^*M\backslash\{0\}$ by homogeneity of degree one of

$$f \circ \sigma^{-1} : \sigma(Q) \subset (T^*M\backslash\{0\}) \to \mathbb{R},$$

for $f \in H^{s+2}(Q, \mathbb{R})$. The composition of these two isomorphisms with $T_e\Phi^{-1}$ gives an isomorphism

$$F : \mathrm{con}_\sigma^{s+1}(Q) \xrightarrow{T_e\Phi^{-1}} H^{s+1}\text{-}\mathfrak{X}_\theta(T^*M\backslash\{0\}) \to S^{s+2}(T^*M\backslash\{0\}) \xrightarrow{j^{-1}} H^{s+2}(Q, \mathbb{R}),$$

defined by $F(X,k) = \theta_\sigma(X)$. In the condition $\pounds_X\theta_\sigma = k\theta_\sigma$, the function k is uniquely determined by X, namely $k = E_\sigma(\theta_\sigma(X))$. From this it follows that F is continuous and hence an isomorphism between $\mathrm{con}_\sigma^{s+1}(Q)$ and $H^{s+2}(Q, \mathbb{R})$ (note the gain of one derivative). We see thus once again that $\mathrm{Con}_\sigma^{s+1}(Q)$ and $\mathrm{Con}_\zeta^{s+1}(Q)$ are isomorphic as nested Lie groups for any two global sections $\sigma, \zeta : Q \to T^*M\backslash\{0\}$, since both are modeled on $H^{s+2}(Q, \mathbb{R})$.

Defining the Hilbert space structure of $S^{s+2}(T^*M\backslash\{0\})$ as the one induced by the isomorphism

$$j : H^{s+2}(Q, \mathbb{R}) \to S^{s+2}(T^*M\backslash\{0\}),$$

it follows that

$$H^{s+2}(Q, \mathbb{R}), S^{s+2}(T^*M\backslash\{0\}), H^{s+1} - \mathfrak{X}_\sigma(T^*M\backslash\{0\}) \,\text{and}\, \mathrm{con}_\sigma^{s+1}(Q)$$

are all isomorphic as Hilbert spaces. It is desirable to compare the topology of $S^{s+2}(T^*M\backslash\{0\})$ with the strong C^1-Whitney topology. Since all elements of $S^{s+2}(T^*M\backslash\{0\})$ are C^2 by the Sobolev embedding theorem, we can define a new topology on $S^{s+2}(T^*M\backslash\{0\})$ in the following way: a neighborhood of zero consists of all those functions $H \in S^{s+2}(T^*M\backslash\{0\})$ for which

$$\mathbf{d}H : (T^*M\backslash\{0\}) \to T^*(T^*M\backslash\{0\})$$

is C^1-close to zero in the strong C^1 -Whitney topology. Let S_W^{s+2} $(T^*M\backslash\{0\})$ be the set $S^{s+2}(T^*M\backslash\{0\})$ equipped with this new topology. One checks that

$$j : H^{s+2}(Q, \mathbb{R}) \to S_W^{s+2}(T^*M\backslash\{0\}),$$

or equivalently the identity $S^{s+2}(T^*M\backslash\{0\}) \to S_W^{s+2}(T^*M\backslash\{0\})$ is continuous with discontinuous inverse, *i.e.*, the new topology is strictly coarser than the original one on $S^{s+2}(T^*M\backslash\{0\})$. This remark is useful in the construction of an explicit chart at e in $H^{s+1}\,\mathrm{Diff}_\theta(T^*M\backslash\{0\})$. ♦

Remarks.

1. The gain of one derivative at the Lie algebra level has a corresponding statement in $H^{s+1}\operatorname{Diff}_\theta(T^*M\backslash\{0\})$: for $\eta \in H^{s+1}\operatorname{Diff}_\theta(T^*M\backslash\{0\})$,

$$\tau^* \circ \eta : (T^*M\backslash\{0\}) \to M$$

is of class H^{s+1}, where $\tau^* : T^*M \to M$ is the cotangent bundle projection. Locally, this means that if $\eta(x,\alpha) = (y(x,\alpha), \beta(x,\alpha))$, then y is H^{s+2} jointly in x and α. To prove this, note that $\eta^*\theta = \theta$ is equivalent locally to

$$\sum_{i=1}^n \beta_i \frac{\partial y^i}{\partial x^k} = \alpha_k, \quad \sum_{i=1}^n \beta_i \frac{\partial y^i}{\partial \alpha_k} = 0, \quad k = 1, \ldots, n.$$

Since η is a diffeomorphism of class H^{s+1}, for fixed x, there exists a unique α such that $\beta = (0, \ldots, 1, \ldots, 0)$, the i-th basis vector. For this choice of α, the first relation shows that the i-th column of the matrix $\left[\partial y^i/\partial x^k\right]$ is H^{s+1}. This says that $y(x,\alpha)$ has all derivatives of order at most $s+2$ square integrable except the derivatives involving only α_k's. The second relation is an elliptic equation with H^{s+1} coefficients of first order in y^i regarded as a function of α only (its symbol maps $(\xi^i) \in \mathbb{R}^n$ to $\left(\beta_1 + \ldots + \xi^i\beta_i + \ldots + \beta_n\right) \in \mathbb{R}^n$) and thus its solution is of class H^{s+2}, *i.e.*, the $(s+2)$-nd derivative of y with respect to α is square integrable and thus y is of class H^{s+2}.

2. Let $\eta \in H^{s+1}\operatorname{Diff}_\theta(T^*M\backslash\{0\})$ be fiber preserving, *i.e.*, $\pi^*\eta(\alpha_x) = \pi^*\eta(\alpha_x')$ for all $\alpha_x, \alpha_x' \in T_x^*M\backslash\{0\}$. Then η can be extended H^{s+1}-smoothly to the zero section by $\eta(0_x) = 0_y$ for $y = \pi(\eta(\alpha_x))$, $\alpha_x \in T^*M\backslash\{0\}$. Thus, $\eta : T^*M \to T^*M$, $\eta^*\theta = \theta$ and hence $\eta = T^*g$ for an H^{s+2} diffeomorphism $g : M \to M$. In particular, if $\pi(\eta(\alpha_x)) = x$ for all $\alpha_x \in T^*M\backslash\{0\}$, then $\eta = e$. From this it follows that the effect of η on base points uniquely determines η, *i.e.*, if $\eta, \overline{\eta} \in H^{s+1}\operatorname{Diff}_\theta(T^*M\backslash\{0\})$ satisfy $\pi \circ \eta = \pi \circ \overline{\eta}$, then $\eta = \overline{\eta}$ (since $\pi(\overline{\eta} \circ \eta^{-1})(\alpha_x) = x$). ♦

Another interesting group considered by Adams, Ratiu, and Schmid [1986] is the group of ***invertible Fourier integral operators***, of interest in the KdV equation. We refer to these papers for details.

N10
Poisson Manifolds

N10.A Proof of the Symplectic Stratification Theorem

We proceed in a series of technical propositions.[1]

Proposition N10.A.1. *Let P be a finite dimensional Poisson manifold with $B^\sharp_z : T^*_z P \to T_z P$ the Poisson tensor. Take $z \in P$ and functions $f_1, \ldots, f_k$ defined on P such that $\{B^\sharp_z df_j\}_{1 \le j \le k}$ is a basis of the range of $B^\sharp_z$. Let $\Phi_{j,t}$ be the local flow defined in a neighborhood of z generated by the Hamiltonian vector field $X_{f_j} = B^\sharp \, df_j$. Let*

$$\Psi^z_{f_1, \ldots, f_k}(t_1, \ldots, t_k) = (\Phi_{1,t_1} \circ \cdots \circ \Phi_{k,t_k})(z)$$

for small enough $t_1, \ldots, t_k$. Then:

(i) *There is an open neighborhood U_δ of $0 \in \mathbb{R}^k$ such that:*

$$\Psi^z_{f_1, \ldots, f_k} : U_\delta \to P$$

is an embedding.

(ii) *The ranges of $(T\Psi^z_{f_1, \ldots, f_k})(t)$ and $B^\sharp_{\Psi^z_{f_1, \ldots, f_k}(t)}$ are equal for $t \in U_\delta$.*

(iii) $\Psi^z_{f_1, \ldots, f_k}(U_\delta) \subset \Sigma_z$.

[1]This proof was kindly supplied by O. Popp

(iv) *If*

$$\Psi^y_{g_1,\dots,g_k} : U_\eta \to P$$

is another map constructed as above and $y \in \Psi^z_{f_1,\dots,f_k}(U_\delta)$*, then there is an open subset,* $U_\epsilon \subset U_\eta$*, such that* $\Psi^y_{g_1,\dots,g_k}$ *is a diffeomorphism from* U_ϵ *to an open subset in* $\Psi^z_{f_1,\dots,f_k}(U_\delta)$*.*

Proof. (i) The smoothness of $\Psi^z_{f_1,\dots,f_k}$ follows from the smoothness of $\Phi_{j,t}$ in both the flow parameter and manifold variables. Then

$$T_0\Psi^z_{f_1,\dots,f_k}(\partial/\partial t_j) = X_{f_j}(z) = B^\sharp_z df_j,$$

which shows that $T_0\Psi^z_{f_1,\dots,f_k}$ is injective. It follows that $\Psi^z_{f_1,\dots,f_k}$ is an embedding on a sufficiently small neighborhood of 0, say U_δ. Notice also that the ranges of $T_0\Psi^z_{f_1,\dots,f_k}$ and of $B^\sharp_z$ coincide.

(ii) Recall from the main text that for any invertible Poisson map Φ on P, we have

$$T\Phi \cdot X_f = X_{f\circ\Phi^{-1}} \circ \Phi$$

and also recall that that Hamiltonian flows are Poisson maps. Therefore, if $t = (t_1, \dots, t_k)$,

$$\begin{aligned} &T_t\Psi^z_{f_1,\dots,f_k}(\partial/\partial t_j) \\ &\quad = (T\Phi_{1,t_1} \circ \dots \circ T\Phi_{j-1,t_{j-1}} \circ X_{f_j} \circ \Phi_{j+1,t_{j+1}} \circ \dots \circ \Phi_{k,t_k})(z) \\ &\quad = (X_{h_j} \circ \Psi^z_{f_1,\dots,f_k})(t), \end{aligned}$$

where

$$h_j = f_j \circ \left(\Phi_{1,t_1} \circ \dots \circ \Phi_{j-1,t_{j-1}}\right)^{-1}.$$

This shows that

$$\text{range } T_t\Psi^x_{f_1,\dots,f_k} \subset \text{ range } B^\sharp_{\Psi^x_{f_1,\dots,f_k}(t)}$$

if $t \in U_\delta$. Since $B^\sharp$ is invariant under Hamiltonian flows, it follows that

$$\dim \text{ range } B^\sharp_{\Psi^z_{f_1,\dots,f_k}(t)} = \dim \text{ range } B^\sharp_z.$$

This last equality, the previous inclusion, and the last remark in the proof of (i) above conclude (ii).

(iii) This is obvious since $\Psi^z_{f_1,\dots,f_k}$ is built from piecewise Hamiltonian curves starting from z.

(iv) Note that $X_g(z) \in$ range $B^\sharp_z$ for any $z \in P$ and any smooth function g. Using (ii), we see that X_g is tangent to the image of $\Psi^z_{f_1,\dots,f_k}$.

Therefore, the integral curves of X_g remain tangent to $\Psi^z_{f_1,\ldots,f_k}(U_\delta)$ if they start from that set. To get $\Psi^y_{g_1,\ldots,g_k}$ we just have to find Hamiltonian curves which start from y. Therefore, we can restrict ourselves to the submanifold $\Psi^z_{f_1,\ldots,f_k}(U_\delta)$ when computing the flows along the Hamiltonian vector fields X_{g_j}; therefore we can consider that the image of $\Psi^y_{g_1,\ldots,g_k}$ is in $\Psi^z_{f_1,\ldots,f_k}(U_\delta)$. The derivative at $0 \in \mathbb{R}^k$ of $\Psi^y_{g_1,\ldots,g_k}$ is an isomorphism to the tangent space of $\Psi^z_{f_1,\ldots,f_k}(U_\delta)$ at y (that is, range $B^\sharp_y$), using (ii) above. Thus, the existence of the neighborhood U_ϵ follows from the inverse function theorem. ■

Proposition N10.A.2. *Let P be a Poisson manifold and B its Poisson tensor. Then for each symplectic leaf $\Sigma \subset P$, the family of charts satisfying* (i) *in the previous proposition, namely,*

$$\left\{\Psi^z_{f_1,\ldots,f_k} \mid z \in \Sigma,\ \{B^\sharp_z\, df_j\}_{1\le j\le k} \quad \textit{a basis for range } B^\sharp_z \right\},$$

gives Σ the structure of a differentiable manifold such that the inclusion is an immersion. Then $T_z\Sigma = \text{range}\, B^\sharp_z$ (so $\dim \Sigma = \operatorname{rank} B^\sharp_z$), for all $z \in \Sigma$. Moreover, Σ has a unique symplectic structure such that the inclusion is a Poisson map.

Proof. Let

$$w \in \Psi^z_{f_1,\ldots,f_k}(U_\delta) \cap \Psi^y_{g_1,\ldots,g_k}(U_\epsilon)$$

and consider $\Psi^w_{h_1,\ldots,h_k} : U_\gamma \to P$. Using (iv) in the proposition above, we can choose U_γ small enough so that

$$\Psi^w_{h_1,\ldots,h_k}(U_\gamma) \subset \Psi^z_{f_1,\ldots,f_k}(U_\delta) \cap \Psi^y_{g_1,\ldots,g_k}(U_\epsilon)$$

is a diffeomorphic embedding in both $\Psi^z_{f_1,\ldots,f_k}(U_\delta)$ and $\Psi^y_{g_1,\ldots,g_k}(U_\epsilon)$. This shows that the transition maps for the given charts are diffeomorphisms and so define the structure of a differentiable manifold on Σ. The fact that the inclusion is an immersion follows from (i) of the above proposition. We get the tangent space of Σ using (i), (ii) of the previous proposition; then the equality of dimensions follows.

It follows from the definition of an immersed Poisson submanifold that Σ is such a submanifold of P. Thus, if $i : \Sigma \to P$ is the inclusion,

$$\{f \circ i, g \circ i\}_\Sigma = \{f, g\} \circ i.$$

Hence if $\{f \circ i, g \circ i\}_\Sigma(z) = 0$ for all functions g then $\{f, g\}(z) = 0$ for all g, that is, $X_g[f](z) = 0$ for all g. This implies that $\mathbf{d}f | T_z\Sigma = 0$ since the vectors $X_g(z)$ span $T_z\Sigma$. Therefore, $i^*\mathbf{d}f = \mathbf{d}(f \circ i) = 0$, which shows that the Poisson tensor on Σ is nondegenerate and thus Σ is a symplectic manifold. This proves the proposition and also completes the proof of the symplectic stratification theorem. ■

There is another proof of the symplectic stratification theorem (using the same idea as for the Darboux coordinates) in Weinstein [1983] (see Libermann and Marle [1987] as well.) The proof given above is along the Frobenius integrability idea. Actually it can be used to produce a proof of the generalized Frobenius theorem.

Theorem N10.A.3 (Singular Frobenius Theorem). *Let D be a distribution of subspaces of the tangent bundle of a finite dimensional manifold M, that is, $D_x \subset T_xM$ as x varies in M. Suppose it is smooth in the sense that for each x there are smooth vector fields X_i defined on some open neighborhood of x and with values in D such that $X_i(x)$ give a basis of D_x. Then D is integrable, that is, for each $x \in M$ there is an immersed submanifold $\Sigma_x \subset M$ with $T_x\Sigma_x = D_x$, if and only if the distribution D is invariant under the (local) flows along vector fields with values in D.*

Proof. The "only if" part follows easily. For the "if" part we remark that the proof of the theorem above can be reproduced here replacing the range of $B_z^\sharp$ by D_x and the Hamiltonian vector fields with vector fields in D. The crucial property needed to prove (ii) in the above proposition (i.e. Hamiltonian fields remain Hamiltonian under Hamiltonian flows) is replaced by the invariance of D given in the hypothesis. ■

Remarks.

1. The conclusion of the above theorem is the same as the Frobenius integrability theorem but it is not assumed that the dimension of D_x is constant.

2. Analogous to the symplectic leaves of a Poisson manifold, we can define the *maximal integral manifolds* of the integrable distribution D using curves along vector fields in D instead of Hamiltonian vector fields. They are also injectively immersed submanifolds in M.

3. The condition that (local) flows of the vector fields with values in D leave D invariant implies the involution property of D, that is, $[X, Y]$ is a vector field with values in D if both X and Y are vector fields with values in D (see Chapter 4 of the text). But the involution property alone is not enough to guarantee that D is integrable (if the dimension of D is not constant).

4. This generalization of the Frobenius integrability theorem is due to Hermann [1962], Stefan [1974], Sussman [1973], and it has proved quite useful in control theory; see also Libermann and Marle [1987]. ◆

N11

Momentum Maps

N11.A Another Example of a Momentum Map

Here is an interesting example of a momentum map that illustrates some of the *convexity properties* of momentum maps of torus actions.

As in Example (a) of §3.3 and Exercise 5.5-4, one checks that the momentum map of the standard $\mathbb{T}^{n+1}$ action

$$(\theta_0, \ldots, \theta_n) \cdot (z_0, z_1, \ldots, z_n) = \left(e^{i\theta_0} z_0, \ldots, e^{i\theta_n} z_n\right)$$

on $\mathbb{C}^{n+1}$ has the expression

$$\mathbf{J}_{\mathbb{C}^{n+1}} (z_0, \ldots, z_n) = \frac{1}{2} \left(|z_0|^2, \ldots, |z_n|^2\right).$$

Since $\mathbf{J}_{\mathbb{C}^{n+1}}$ is invariant under the circle action

$$\theta \cdot (z_0, \ldots, z_n) = \left(e^{i\theta} z_0, \ldots, e^{i\theta} z_n\right)$$

on the unit sphere $S^{2n-1} = \mathbf{J}^{-1}_{\mathbb{C}^{n+1}} (1/2)$, it follows that $\mathbf{J}_{\mathbb{C}^{n+1}}$ induces a map $\mathbf{J}_{\mathbb{CP}^n} : \mathbb{CP}^n \to \mathbb{R}^{n+1}$ given by

$$\mathbf{J}_{\mathbb{CP}^n} ([z_0 : \cdots : z_n]) = \frac{1}{2} \left(|z_0|^2, \ldots, |z_n|^2\right)$$

where

$$[z_0 : z_1 : \cdots : z_n] = [(z_0, \cdots, z_n)]$$

denotes the equivalence class of $(z_0, \ldots, z_n)$ in $\mathbb{CP}^n$. It is easily verified that $\mathbf{J}_{\mathbb{CP}^n}$ is a momentum map of the $\mathbb{T}^{n+1}$ action

$$(\theta_0, \ldots, \theta_n) \cdot [z_0 : \cdots : z_n] = \left[e^{i\theta_0} z_0 : \cdots : e^{i\theta_n} z_n\right]$$

on $\mathbb{CP}^n$. The image of $\mathbb{CP}^n$ under $\mathbf{J}_{\mathbb{CP}^n}$ clearly coincides with

$$\begin{aligned} \mathbf{J}_{\mathbb{C}^{n+1}}\left(S^{2n-1}\right) = \big\{ & (t_0, \ldots, t_n) \in \mathbb{R}^{n+1} \mid t_0 + \cdots + t_n = 1, \\ & t_i \geq 0 \text{ for all } i = 0, \ldots, n \big\}, \end{aligned}$$

which is the standard n simplex in $\mathbb{R}^{n+1}$ spanned by the vertices

$$(1, 0, \ldots, 0),\ (0, 1, \ldots, 0), \ldots, (0, \ldots 1).$$

This is an instance of the Atiyah–Guillemin–Sternberg curvexity theorem (Atiyah [1982], Guillemin and Sternberg [1982]), which states that if a torus T acts on a compact connected symplectic manifold P in a Hamiltonian fashion with invariant momentum map $\mathbf{J} : P \to \mathfrak{t}^*$, then $\mathbf{J}(P)$ is a convex compact polytope whose vertices are given by $\mathbf{J}(P^T)$, where P^T is the fixed point set of the T–action on P. These fixed point sets are, interestingly, the bifurcation points of the momentum mapping, according to Arms, Marsden, and Moncrief [1981].

In our example,

$$(\mathbb{CP}^n)^{\mathbb{T}^{n+1}} = \{[1 : 0 : \cdots : 0], \ldots, [0 : \cdots : 1]\},$$

whose image clearly consists of the vertices

$$(1/2, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1/2).$$

If $n = 1$, the Hopf fibration

$$(z_0, z_1) \in S^3 \subset \mathbb{C}^2 \mapsto \left(2 z_0 \overline{z}_1 |z_1|^2 - |z_0|^2\right) \in S^2$$

identifies $\mathbb{CP}^1$ with S^2 and the $\mathbb{T}^2$ action on S^2 is given by rotations about the vertical axis:

$$\begin{aligned} (\theta_0, \theta_1) \cdot \left(x^1, x^2, x^3\right) = \big(& x^1 \cos(\theta_0 - \theta_1) - x^2 \sin(\theta_0 - \theta_1)\, x^1 \sin(\theta_0 - \theta_1) \\ & + x^2 \cos(\theta_0 - \theta_1), x^3 \big). \end{aligned}$$

In terms of $\left(x^1, x^2, x^3\right) \in S^2$ the momentum map $\mathbf{J}_{\mathbb{CP}^1}$ becomes

$$\mathbf{J}_{S^2}\left(x^1, x^2, x^3\right) = \frac{1}{4}(1 + x_3, 1 - x_3)$$

whose image is in the line segment joining $(0, 1/2)$ and $(1/2, 0)$ in the plane.

N13

Lie-Poisson Reduction

N13.A Proof of the Lie–Poisson Reduction Theorem for $\mathrm{Diff}_{\mathrm{vol}}(M)$

An interesting special case of the Lie–Poisson reduction theorem is $G = \mathrm{Diff}_{\mathrm{vol}}(\Omega)$, the subgroup of the group of diffeomorphisms $\mathrm{Diff}(\Omega)$ of a region $\Omega \subset \mathbb{R}^3$, consisting of the volume-preserving diffeomorphisms. We shall treat $\mathrm{Diff}(\Omega)$ and $\mathrm{Diff}_{\mathrm{vol}}(\Omega)$ formally, although it is known how to handle the functional analysis issues involved (see Ebin and Marsden [1970] and Adams, Ratiu, and Schmid [1986] and references therein). We shall prove the Lie–Poisson reduction theorem for this special case.

The Lie Algebra of Diff. For $\eta \in \mathrm{Diff}(\Omega)$, the tangent space at η is given by the set of maps $V : \Omega \to T\Omega$ satisfying $V(X) \in T_{\eta(X)}\Omega$, that is, vector fields over η. We think of V as a material velocity field. Thus, the tangent space at the identity is the space of vector fields on Ω (tangent to $\partial\Omega$). Given two such vector fields, their left Lie algebra bracket is related to the Jacobi–Lie bracket by (see Chapter 9):

$$[V, W]_{LA} = -\,[V, W]_{JL}\,,$$

that is,

$$[V, W]_{LA} = (W \cdot \nabla)V - (V \cdot \nabla)W, \tag{N13.A.1}$$

as one finds using the definitions.

Right Translation. We will be computing the *right* Lie–Poisson bracket on $\mathfrak{g}^*$. Right translation by φ on G is given by

$$R_\varphi \eta = \eta \circ \varphi. \tag{N13.A.2}$$

Differentiating (N13.A.2) with respect to η gives

$$TR_\varphi \cdot V = V \circ \varphi. \tag{N13.A.3}$$

Identify $T_\eta G$ with those V's such that the vector field on $\mathbb{R}^3$ given by $\mathbf{v} = V \circ \eta^{-1}$, is divergence-free and identify $T^*_\eta G$ with $T_\eta G$ via the pairing

$$\langle \pi, V \rangle = \int_\Omega \pi \cdot V \, dx \, dy \, dz, \tag{N13.A.4}$$

where $\pi \cdot V$ is the dot product on $\mathbb{R}^3$. By the change of variables formula, and the fact that $\varphi \in G$ has unit Jacobian,

$$\begin{aligned}\langle T^* R_\varphi \cdot \pi, V \rangle &= \langle \pi, TR_\varphi \cdot V \rangle \\ &= \int_\Omega \pi \cdot (V \circ \varphi) \, dx \, dy \, dz = \int_\Omega (\pi \circ \varphi^{-1}) \cdot V \, dx \, dy \, dz,\end{aligned}$$

so

$$T^* R_\varphi \cdot \pi = \pi \circ \varphi^{-1}. \tag{N13.A.5}$$

Derivatives of Right Invariant Extensions. If $F : \mathfrak{g}^* \to \mathbb{R}$ is given, its right invariant extension is

$$F_R(\eta, \pi) = F(\pi \circ \eta^{-1}). \tag{N13.A.6}$$

Let us denote elements of $\mathfrak{g}^*$ by $\mathbf{M}$, so we are investigating the relation between the canonical bracket of F_R and H_R and the Lie–Poisson bracket of F and H via the relation

$$\mathbf{M} \circ \eta = \pi.$$

From (N13.A.6) and the chain rule, we get

$$\begin{aligned}\mathbf{D}_\eta F_R(\mathrm{Id}, \pi) \cdot \mathbf{v} &= -\mathbf{D}_\mathbf{M} F(\mathbf{M}) \cdot \mathbf{D}_\eta \pi(\mathrm{Id}) \cdot \mathbf{v} \\ &= -\int_\Omega ((\mathbf{v} \cdot \nabla)\mathbf{M}) \cdot \frac{\delta F}{\delta \mathbf{M}} \, dx \, dy \, dz,\end{aligned} \tag{N13.A.7}$$

where $\delta F / \delta \mathbf{M}$ is a divergence-free vector field parallel to the boundary. Since T^*G is not given as a product space, one has to worry about what it means to hold π constant in (N13.A.7). We leave it to the ambitious reader to justify this formal calculation.

Computation of Brackets. Thus, the canonical bracket at the identity becomes

$$
\begin{aligned}
\{F_R, H_R\}(\mathrm{Id}, \pi) &= \int_\Omega \left(\frac{\delta F_R}{\delta \eta} \frac{\delta H_R}{\delta \pi} - \frac{\delta H_R}{\delta \eta} \frac{\delta F_R}{\delta \pi} \right) dx\, dy\, dz \\
&= \mathbf{D}_\eta F_R(\mathrm{Id}, \pi) \cdot \frac{\delta H_R}{\delta \pi} - \mathbf{D}_\eta H_R(\mathrm{Id}, \pi) \cdot \frac{\delta F_R}{\delta \pi}. \quad \text{(N13.A.8)}
\end{aligned}
$$

At the identity, $\pi = \mathbf{M}$ and $\delta F_R/\delta\pi = \delta F/\delta\mathbf{M}$, so substituting this and (N13.A.7) into (N13.A.8), we get

$$
\begin{aligned}
&\{F_R, H_R\}(\mathrm{Id}, \mathbf{M}) \\
&\quad = -\int_\Omega \left[\left(\frac{\delta H}{\delta \mathbf{M}} \cdot \nabla \right) \mathbf{M} \cdot \frac{\delta F}{\delta \mathbf{M}} - \left(\frac{\delta F}{\delta \mathbf{M}} \cdot \nabla \right) \mathbf{M} \cdot \frac{\delta H}{\delta \mathbf{M}} \right] dx\, dy\, dz. \\
&\qquad \text{(N13.A.9)}
\end{aligned}
$$

Equation (N13.A.9) may be integrated by parts to give

$$
\begin{aligned}
&\{F_R, H_R\}(\mathrm{Id}, \mathbf{M}) \\
&\quad = \int \mathbf{M} \cdot \left[\left(\frac{\delta H}{\delta \mathbf{M}} \cdot \nabla \right) \frac{\delta F}{\delta \mathbf{M}} - \left(\frac{\delta F}{\delta \mathbf{M}} \cdot \nabla \right) \frac{\delta H}{\delta \mathbf{M}} \right] dx\, dy\, dz \\
&\quad = \int \mathbf{M} \cdot \left[\frac{\delta F}{\delta \mathbf{M}}, \frac{\delta H}{\delta \mathbf{M}} \right]_{LA} dx\, dy\, dz, \qquad \text{(N13.A.10)}
\end{aligned}
$$

which is the "+" Lie–Poisson bracket. In doing this step note $\mathrm{div}(\delta H/\delta\mathbf{M}) = 0$ and since $\delta H/\delta\mathbf{M}$ and $\delta F/\delta\mathbf{M}$ are parallel to the boundary, no boundary term appears. When doing free boundary problems, these boundary terms are essential to retain (see Lewis, Marsden, Montgomery, and Ratiu [1986]).

For other diffeomorphism groups, it may be convenient to treat $\mathbf{M}$ as a one-form density rather than a vector field.

N13.B Proof of the Lie–Poisson Reduction Theorem for $\mathrm{Diff}_{\mathrm{can}}(P)$

This section discusses the Lie–Poisson reduction theorem for the special case $G = \mathrm{Diff}_{\mathrm{can}}(P)$, the group of canonical transformations of a boundaryless symplectic (or Poisson) manifold P. The Lie algebra of $\mathrm{Diff}_{\mathrm{can}}(P)$ is the algebra of infinitesimal Poisson automorphisms, or Poisson derivations, that is, vector fields X on P for which

$$
X[\{f, h\}] = \{X[f], h\} + \{f, X[h]\}
$$

for any $f, h \in \mathcal{F}(P)$. To avoid complications, we work with the globally Hamiltonian vector fields by suitably restricting P or $\mathrm{Diff}_{\mathrm{can}}(P)$. Each

Hamiltonian vector field can be identified with its generator (modulo additive constants being understood). From the formula

$$[X_k, X_h]_{LA} = -[X_k, X_h]_{JL} = X_{\{k,h\}} \tag{N13.B.1}$$

we see that $\mathfrak{g}$ may be identified with $\mathcal{F}(P)$ with the Lie bracket given by the Poisson bracket. One could then identify $\mathfrak{g}^*$ with functions f on P via the pairing

$$\langle f, h \rangle = \int_P fh \, d\mu, \tag{N13.B.2}$$

where $d\mu$ is the Liouville measure. If P is only a Poisson manifold, identify $\mathfrak{g}^*$ with the densities on P. As in the last section, $T_\eta G$ consists of vector fields of the form $X_k \circ \eta$.

To identify the dual space of $T_\eta G$, we need objects to pair with $T_\eta G$ in a nondegenerate way. Since $X_k \circ \eta = T\eta \circ X_{k \circ \eta}$, we cannot simply use the pairing (N13.B.2) to identify $T^*_\eta G$ with $\mathcal{F}(P)$; such a procedure would not account for the extra factor $T\eta$. Instead, regard $\pi \in \mathfrak{g}^*$ as a one-form on P and pair it with $X_k \in \mathfrak{g}$ by

$$\langle \pi, X_k \rangle = \int_P \pi \cdot X_k \, d\mu. \tag{N13.B.3}$$

This pairing is degenerate; for example, if $\pi = \mathbf{d}f$, then $\langle \pi, X_k \rangle = 0$ by Stokes' theorem. To simplify matters, let us work in coordinates, and write

$$\pi = \pi_i \, dq^i + \pi^i \, dp_i$$

so, integrating by parts,

$$\begin{aligned} \int_P \pi \cdot X_k \, d\mu &= \int_P \left(\pi_i \frac{\partial k}{\partial p_i} - \pi^i \frac{\partial k}{\partial q^i} \right) dq \, dp \\ &= \int_P \left(-\frac{\partial \pi_i}{\partial p_i} + \frac{\partial \pi^i}{\partial q^i} \right) k \, dq \, dp. \end{aligned} \tag{N13.B.4}$$

Thus if we work modulo π's satisfying the divergence-like condition

$$\frac{\partial \pi^i}{\partial q^i} - \frac{\partial \pi_i}{\partial p_i} = 0, \tag{N13.B.5}$$

then the pairing (N13.B.3) is nondegenerate. Now let $f \in \mathcal{F}(P)$ be given and define π by requiring

$$\int_P \pi \cdot X_k \, d\mu = \int_P fk \, d\mu \tag{N13.B.6}$$

for all $k \in \mathcal{F}(P)$. Thus, from (N13.B.4), we need

$$\frac{\partial \pi^i}{\partial q^i} - \frac{\partial \pi_i}{\partial p_i} = f. \tag{N13.B.7}$$

Note that if $\pi = (\partial h/\partial q^i)\, dq^i + (\partial h/\partial p_i)\, dp_i$, the left side of (N13.B.7) is identically zero since $\partial^2 h/\partial q^i\, \partial p_i = \partial^2 h/\partial p_i\, \partial q^i$. If we take $\pi = (\partial \psi/\partial p_i)\, dq^i - (\partial \psi/\partial q^i)\, dp_i$, then ψ is determined by $-\Delta\psi = f$, so ψ is now uniquely determined modulo π's satisfying (N13.B.5). In two-dimensional incompressible flow, which corresponds to the special case $\dim P = 2$, ψ is the stream function and f the vorticity.

Identify $T^*_\eta G$ with one-forms modulo exact one-forms over η ; that is, objects of the form $\pi_\eta = \pi \circ \eta$. Given $F \in \mathcal{F}(P)$, define F on $\mathfrak{g}^*$ by $F(\pi) = F(f)$, where π and f are related by (N13.B.6) and extend it to be right invariant by

$$F_R(\eta, \pi_\eta) = F(\pi_\eta \circ \eta^{-1}).$$

As in the preceding section and using vector analysis notation,

$$\begin{aligned} \mathbf{D}_\eta F_R(\mathrm{Id}, \pi) \cdot X_k &= -\mathbf{D}F(\pi) \cdot \mathbf{D}_\eta \pi_\eta(\mathrm{Id}) \cdot X_k \\ &= -\int \frac{\delta F}{\delta \pi} \cdot (X_k \cdot \nabla \pi) d\mu. \end{aligned} \tag{N13.B.8}$$

Also, $\delta F_R/\delta\pi = \delta F/\delta\pi$ at $\eta = \mathrm{Id}$, as before. Thus the canonical bracket at $\eta = \mathrm{Id}$ is

$$\begin{aligned} \{F_R, H_R\}(\mathrm{Id}, \pi) &= \int_P \left(\frac{\delta F_R}{\delta \eta} \frac{\delta H_R}{\delta \pi} - \frac{\delta H_R}{\delta \eta} \frac{\delta F_R}{\delta \pi} \right) d\mu \\ &= -\int_P \left[\frac{\delta F}{\delta \pi} \cdot \left(\frac{\delta H}{\delta \pi} \cdot \nabla \pi \right) - \frac{\delta H}{\delta \pi} \cdot \left(\frac{\delta F}{\delta \pi} \cdot \nabla \pi \right) \right] d\mu, \end{aligned}$$

which may be integrated by parts to give

$$\begin{aligned} \{F_R, H_R\}(\mathrm{Id}, \pi) &= \int_P \pi \cdot \left\{ \left(\frac{\delta H}{\delta \pi} \cdot \nabla \right) \frac{\delta F}{\delta \pi} - \left(\frac{\delta F}{\delta \pi} \cdot \nabla \right) \frac{\delta F}{\delta \pi} \right) d\mu \\ &= \int_P \pi \cdot \left[\frac{\delta F}{\delta \pi}, \frac{\delta H}{\delta \pi} \right]_{LA} d\mu. \end{aligned} \tag{N13.B.9}$$

To write this in terms of $\mathcal{F}(P)$ we use (N13.B.6) to write

$$\left\langle \delta\pi, \frac{\delta F}{\delta \pi} \right\rangle = \int_P \delta\pi \cdot X_k \, d\mu$$

for some k to be determined. By the chain rule,

$$\begin{aligned} \left\langle \delta\pi, \frac{\delta F}{\delta \pi} \right\rangle &= \mathbf{D}_\pi F \cdot \delta\pi = \mathbf{D}_f F \cdot (\mathbf{D}_\pi f \cdot \delta\pi) \\ &= \int_P \frac{\delta F}{\delta f} \left(\mathbf{D}_\pi f \cdot \delta\pi \right) d\mu. \end{aligned} \tag{N13.B.10}$$

Differentiating (N13.B.6) implicitly relative, π we get

$$\int_P \delta\pi \cdot X_k \, d\mu = \int_P (\mathbf{D}_\pi f \cdot \delta\pi) \, k \, d\mu,$$

so by (N13.B.10)

$$\left\langle \delta\pi, \frac{\delta F}{\delta \pi} \right\rangle = \int_P \delta\pi \cdot X_{\delta F/\delta f} \, d\mu, \tag{N13.B.11}$$

i.e.,

$$\frac{\delta F}{\delta \pi} = X_{\delta F/\delta f}. \tag{N13.B.12}$$

Thus (N13.B.9) becomes, with the aid of (N13.B.1) and (N13.B.6),

$$\begin{aligned} \{F_R, H_R\}(\mathrm{Id}, \pi) &= \int_P \pi \cdot \left[X_{\delta F/\delta f}, X_{\delta H/\delta f} \right]_{LA} d\mu \\ &= \int_P \pi \cdot X_{\{\delta F/\delta f, \delta H/\delta f\}} \, d\mu \\ &= \int_P f \left\{ \frac{\delta F}{\delta f}, \frac{\delta H}{\delta f} \right\} d\mu \end{aligned} \tag{N13.B.13}$$

which is the "+" Lie–Poisson bracket on $\mathfrak{g}^*$ identified with $\mathcal{F}(P)$.

Remarks. 1. This derivation is related to one given by Kaufman and Dewar [1984].

2. The bracket (N13.B.13) can be understood as a limit of the canonical bracket for a larger number of particles moving in P by taking f to be a sum of delta functions at the particle positions. This derivation is due to Bialynicki-Birula, Hubbard, and Turski [1984]; see also Kaufman [1982] and Marsden, Morrison, and Weinstein [1984]. ♦

N13.C The Linearized Lie–Poisson Equations

Here we show that the Lie–Poisson equations linearized about an equilibrium solution (such as the rigid body or the ideal fluid equations) are Hamiltonian with respect to a "constant coefficient" Lie–Poisson bracket. The Hamiltonian for these linearized equations is $\frac{1}{2}\delta^2 (H + C)|_e$, the quadratic functional obtained by taking one-half of the second variation of the Hamiltonian plus conserved quantities and evaluating it at the equilibrium solution where the conserved quantity C (often a Casimir) is chosen so that the first variation $\delta(H + C)$ vanishes at the equilibrium. A consequence is that the linearized dynamics preserves $\frac{1}{2}\delta^2 (H + C)|_e$. This is useful for studying stability of the linearized equations.

For a Lie algebra $\mathfrak{g}$, recall that the Lie–Poisson bracket is defined on $\mathfrak{g}^*$, the dual of $\mathfrak{g}$ with respect to (a weakly nondegenerate) pairing $\langle\, , \rangle$ between $\mathfrak{g}^*$ and $\mathfrak{g}$ by the usual formula

$$\{F, G\}(\mu) = \left\langle \mu, \left[\frac{\delta F}{\delta \mu}, \frac{\delta G}{\delta \mu}\right]\right\rangle, \tag{N13.C.1}$$

where $\delta F/\delta\mu \in \mathfrak{g}$ is determined by

$$\mathbf{D}F(\mu) \cdot \delta\mu = \left\langle \delta\mu, \frac{\delta F}{\delta\mu}\right\rangle \tag{N13.C.2}$$

when such an element $\delta F/\delta\mu$ exists, for any μ, $\delta\mu \in \mathfrak{g}^*$. The equations of motion are

$$\frac{d\mu}{dt} = -\operatorname{ad}^*_{\delta H/\delta\mu}\mu, \tag{N13.C.3}$$

where $H : \mathfrak{g}^* \to \mathbb{R}$ is the Hamiltonian, $\operatorname{ad}_\xi : \mathfrak{g} \to \mathfrak{g}$ is the adjoint action, $\operatorname{ad}_\xi \cdot\eta = [\xi, \eta]$ for ξ, $\eta \in \mathfrak{g}$, and $\operatorname{ad}^*_\xi : \mathfrak{g}^* \to \mathfrak{g}^*$ is its dual. Let $\mu_e \in \mathfrak{g}^*$ be an equilibrium solution of (N13.C.3). The linearized equations of (N13.C.3) at μ_e are obtained by expanding in a Taylor expansion with small parameter ε using $\mu = \mu_e + \varepsilon\delta\mu$, and taking $(d/d\varepsilon)|_{\varepsilon=0}$ of the resulting equations. This gives

$$\frac{\delta H}{\delta\mu} = \frac{\delta H}{\delta\mu_e} + \varepsilon\mathbf{D}\left(\frac{\delta H}{\delta\mu}\right)(\mu_e)\cdot\delta\mu + O(\varepsilon^2), \tag{N13.C.4}$$

where $\langle \delta H/\delta\mu_e, \delta\mu\rangle := \mathbf{D}H(\mu_e)\cdot\delta\mu$, and the derivative $\mathbf{D}(\delta H/\delta\mu)(\mu_e)\cdot\delta\mu$ is the linear functional

$$\nu \in \mathfrak{g}^* \mapsto \mathbf{D}^2 H(\mu_e)\cdot(\delta\mu, \nu) \in \mathbb{R} \tag{N13.C.5}$$

by using the definition (N13.C.2). Since

$$\delta^2 H(\delta\mu) := \mathbf{D}^2 H(\mu_e)\cdot(\delta\mu, \delta\mu),$$

it follows that the functional (N13.C.5) equals

$$\frac{1}{2}\frac{\delta(\delta^2 H)}{\delta(\delta\mu)}.$$

Consequently, (N13.C.4) becomes

$$\frac{\delta H}{\delta\mu} = \frac{\delta H}{\delta\mu_e} + \frac{1}{2}\varepsilon\frac{\delta(\delta^2 H)}{\delta(\delta\mu)} + O(\varepsilon^2) \tag{N13.C.6}$$

and the Lie–Poisson equations (N13.C.3) yield

$$\frac{d\mu_e}{dt} + \varepsilon \frac{d(\delta\mu)}{dt} = -\operatorname{ad}^*_{\delta H/\delta\mu_e} \mu_e \\ - \frac{1}{2}\varepsilon \left[\operatorname{ad}^*_{\delta(\delta^2 H)/\delta(\delta\mu)} \mu_e - \operatorname{ad}^*_{\delta H/\delta\mu_e} \delta\mu\right] + O(\varepsilon^2).$$

Thus, the linearized equations are

$$\frac{d(\delta\mu)}{dt} = -\frac{1}{2} \operatorname{ad}^*_{\delta(\delta^2 H)/\delta(\delta\mu)} \mu_e - \operatorname{ad}^*_{\delta H/\delta\mu_e} \delta\mu. \tag{N13.C.7}$$

If H is replaced by $H_C := H + C$, with the Casimir function C chosen to satisfy $\delta H_C/\delta\mu_e = 0$, we get $\operatorname{ad}^*_{\delta H_C/\delta\mu_e} \delta\mu = 0$, and so

$$\frac{d(\delta\mu)}{dt} = -\frac{1}{2} \operatorname{ad}^*_{\delta(\delta^2 H_C)/\delta(\delta\mu)} \mu_e. \tag{N13.C.8}$$

Equation (N13.C.8) is Hamiltonian with respect to the linearized Poisson bracket (see Example (f) of §10.1):

$$\{F, G\}(\mu) = \left\langle \mu_e, \left[\frac{\delta F}{\delta\mu}, \frac{\delta G}{\delta\mu}\right]\right\rangle. \tag{N13.C.9}$$

Ratiu [1982] interprets this bracket in terms of a Lie–Poisson structure of a loop extension of $\mathfrak{g}$. The Poisson bracket (N13.C.9) differs from the Lie–Poisson bracket (N13.C.1) in that it is *constant* in μ. With respect to the Poisson bracket (N13.C.9), Hamilton's equations given by $\delta^2 H_C$ are (N13.C.8), as an easy verification shows. Note that the critical points of $\delta^2 H_C$ are stationary solutions of the linearized equation (N13.C.8), that is, they are *neutral modes* for (N13.C.8).

If $\delta^2 H_C$ is definite, then either $\delta^2 H_C$ or $-\delta^2 H_C$ is positive-definite and hence defines a norm on the space of perturbations $\delta\mu$ (which is $\mathfrak{g}^*$). Being twice the Hamiltonian function for (N13.C.8), $\delta^2 H_C$ is conserved. So, any solution of (N13.C.8) starting on an energy surface of $\delta^2 H_C$ (i.e., on a sphere in this norm) stays on it and hence the zero solution of (N13.C.8) is (Liapunov) stable. Thus, *formal stability*, i.e., definiteness of $\delta^2 H_C$, implies *linearized stability*. It should be noted, however, that the conditions for definiteness of $\delta^2 H_C$ are entirely different from the conditions for "normal mode stability," that is, that the operator acting on $\delta\mu$ given by (N13.C.8) have a purely imaginary spectrum. In particular, having a purely imaginary spectrum for the linearized equation does *not* produce Liapunov stability of the linearized equations.

The difference between $\delta^2 H_C$ and the operator in (N13.C.8) can be made explicit, as follows. Assume that there is a weak Ad-invariant metric $\langle\!\langle\, , \rangle\!\rangle$ on $\mathfrak{g}$ and a linear operator $L : \mathfrak{g} \to \mathfrak{g}$ such that

$$\delta^2 H_C = \langle\!\langle \delta\mu, L\delta\mu \rangle\!\rangle; \tag{N13.C.10}$$

L is symmetric with respect to the metric $\langle\!\langle\, , \rangle\!\rangle$, that is, $\langle\!\langle \xi, L\eta \rangle\!\rangle = \langle\!\langle L\xi, \eta \rangle\!\rangle$ for all $\xi, \eta \in \mathfrak{g}$. Then the linear operator in (N13.C.8) becomes

$$\delta\mu \mapsto [L\delta\mu, \mu_e] \tag{N13.C.11}$$

which, of course, differs from L, in general. However, note that the kernel of L is included in the kernel of the linear operator (N13.C.11), that is, the zero eigenvalues of L give rise to "neutral modes" in the spectral analysis of (N13.C.11). There is a remarkable coincidence of the zero-eigenvalue equations for these operators in fluid mechanics: for the Rayleigh equation describing plane-parallel shear flow in an inviscid homogeneous fluid, taking normal modes makes the zero-eigenvalue equations corresponding to L and to (N13.C.11) coincide (see Abarbanel, Holm, Marsden, and Ratiu [1986]).

For additional applications of the stability method, see Holm, Marsden, Ratiu, and Weinstein [1985], Abarbanel and Holm [1987], Simo, Posbergh, and Marsden [1990, 1991], and Simo, Lewis, and Marsden [1991]. For a more general treatment of the linearization process, see Marsden, Ratiu, and Raugel [1991].

Exercises

◇ **N13.C-1.** Write out the linearized rigid body equations about an equilibrium explicitly.

◇ **N13.C-2.** Let $\mathfrak{g}$ be finite dimensional. Let $e_1, \ldots, e_n$ be a basis for $\mathfrak{g}$ and $e^1, \ldots, e^n$ a dual basis for $\mathfrak{g}^*$. Let $\mu = \mu_a e^a \in \mathfrak{g}^*$ and $H(\mu) = H(\mu_1, \ldots, \mu_n) : \mathfrak{g}^* \to \mathbb{R}$. Let $[\mu_a, \mu_b] = C^d_{ab}\mu_d$. Derive a coordinate expression for the linearized equations (N13.C.7):

$$\frac{d(\delta\mu)}{dt} = -\frac{1}{2}\,\mathrm{ad}^*_{\delta(\delta^2 H)/\delta\mu}\,\mu_e - \mathrm{ad}^*_{\delta H/\delta\mu_e}\,\delta\mu.$$

N14

Coadjoint Orbits

N14.A Casimir Functions do not Determine Orbits

The purpose of this section is to use Corollary 14.4.3 to determine all Casimir functions for the Lie algebra in Example (f) of §14.1. If

$$\mu = \begin{bmatrix} iu & 0 & 0 \\ 0 & i\alpha u & 0 \\ a & b & 0 \end{bmatrix} \in \mathfrak{g}^*, \quad \xi = \begin{bmatrix} is & 0 & x \\ 0 & i\alpha s & y \\ 0 & 0 & 0 \end{bmatrix} \in \mathfrak{g},$$

for $a, b, x, y \in \mathbb{C}, u, s \in \mathbb{R}$, then it is straightforward to check that

$$\operatorname{ad}_\xi^* \mu = \begin{bmatrix} iu'' & 0 & 0 \\ 0 & i\alpha u'' & 0 \\ -isa & -i\alpha sb & 0 \end{bmatrix},$$

where

$$u'' = -\frac{1}{1+\alpha^2} \operatorname{Im}(ax + \alpha by).$$

Thus, if at least one of a, b is not zero, then

$$\mathfrak{g}_\mu = \left\{ \begin{bmatrix} 0 & 0 & x \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} \middle| \operatorname{Im}(ax + \alpha by) = 0 \right\},$$

whereas if $a = b = 0$, then $\mathfrak{g}_\mu = \mathfrak{g}$. For $C : \mathfrak{g}^* \to \mathbb{R}$, denote by

$$\frac{\delta C}{\delta \mu} = \begin{bmatrix} iC_u & 0 & C_a \\ 0 & i\alpha C_u & C_b \\ 0 & 0 & 0 \end{bmatrix},$$

where $C_u \in \mathbb{R}, C_a, C_b \in \mathbb{C}$ are the partial derivatives of C relative to the variables u, a, b. Thus, the condition $\delta C/\delta\mu \in \mathfrak{g}_\mu$ for all μ implies that $C_u = 0$, that is, C is independent of u and

$$\operatorname{Im}(aC_a + \alpha b C_b) = 0. \tag{N14.A.1}$$

The same condition could have been obtained by lengthier direct calculations involving the Lie–Poisson bracket. Here are the highlights. The commutator bracket on $\mathfrak{g}$ is given by

$$\left[\begin{bmatrix} is & 0 & x \\ 0 & i\alpha s & y \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} iu & 0 & z \\ 0 & i\alpha u & w \\ 0 & 0 & 0 \end{bmatrix}\right] = \begin{bmatrix} 0 & 0 & i(sz - ux) \\ 0 & 0 & i\alpha(sw - uy) \\ 0 & 0 & 0 \end{bmatrix},$$

so that for $\mu \in \mathfrak{g}^*$ parameterized by $u \in \mathbb{R}, a, b, \in \mathbb{C}$, we have

$$\begin{aligned} \{F, H\}_-(\mu) &= -\operatorname{Re}\left[\operatorname{Trace}\left(\mu\left[\frac{\delta F}{\delta\mu}, \frac{\delta H}{\delta\mu}\right]\right)\right] \\ &= \operatorname{Im}[a(F_u H_a - H_u F_a) + \alpha b(F_u H_b - H_u F_b)]. \end{aligned} \tag{N14.A.2}$$

Taking $F_u = F_b = 0$ in $\{F, C\}_- = 0$, forces $C_u = 0$. Then the remaining condition reduces to (N14.A.1).

To solve (N14.A.1) we need first to convert it into a real equation. Regard C as being defined on $\mathbb{C}^2$ with values in $\mathbb{C}$ and write $C = A + iB$, with A and B real-valued functions. We start by searching for holomorphic Casimir functions.

Write $a = p + iq$, $b = v + iw$ so that by the Cauchy-Riemann equations we have

$$A_p = B_q, \quad A_q = -B_p, \quad A_v = B_w, \quad A_w = -B_v$$

and also, since C is holomorphic

$$\begin{aligned} C_a &= A_p + iB_p = B_q - iA_q = C_p = -iC_q \\ C_b &= A_v + iB_v = B_w - iA_w = C_v = -iC_w. \end{aligned}$$

Therefore,

$$\begin{aligned} 0 &= \operatorname{Im}((p + iq)(A_p + iB_p) + \alpha(v + iw)(A_v + iB_v)) \\ &= qA_p + pB_p + \alpha(wA_v + vB_v) \\ &= qA_p - pA_q + \alpha w A_v - \alpha v A_w \end{aligned}$$

by the Cauchy–Riemann equations. We solve this partial differential equation by the method of characteristics. The flow of the vector field with components $(q, -p, \alpha w, -\alpha v)$ is given by

$$\begin{aligned} F_t(p, q, v, w) = (&p\cos t + q\sin t, -p\sin t + q\cos t, \\ &v\cos\alpha t + w\sin\alpha t, -v\sin\alpha t + w\cos\alpha t) \end{aligned}$$

and thus any solution is a rotationally invariant function. An argument (using a theorem of Whitney [1943]) shows that solutions have the form

$$A = f(p^2 + q^2, v^2 + w^2)$$

for a real valued function $f : \mathbb{R}^2 \to \mathbb{R}$ is the general solution of this equation. Thus, any Casimir function is a functional of $p^2 + q^2$ and $v^2 + w^2$. Note that

$$\begin{aligned} C_a &= A_p + iB_p = A_p - iA_q, \quad \text{and} \\ C_b &= A_v + iB_v = A_v - iA_w. \end{aligned}$$

In particular, if $f(x, y) = x$, that is, $A = p^2+q^2$, we have $C_a = 2(p-iq)$ and $C_b = 0$. One can then verify directly that p^2+q^2 is a Casimir function using formula (N14.A.2). Similarly, one sees directly that $v^2 + w^2$ is a Casimir function.

Since the generic leaf of $\mathfrak{g}^*$ is two-dimensional (see Example 14.1(f)) and the dimension of $\mathfrak{g}$ is five, it follows that the Casimir functions do not characterize the generic coadjoint orbits. This is in agreement with the observation made in Example 14.1(f) that the generic coadjoint orbits have as closure the three-dimensional submanifolds of $\mathfrak{g}^*$, which are the product of the torus of radii $|a|$ and $|b|$ and the u'-line, if one expresses the orbit through

$$\mu = \begin{bmatrix} iu & 0 & 0 \\ 0 & i\alpha u & 0 \\ a & b & 0 \end{bmatrix}$$

as

$$\left\{ \begin{bmatrix} iu' & 0 & 0 \\ 0 & iu' & 0 \\ ae^{-it} & be^{-i\alpha t} & 0 \end{bmatrix} \middle| \, u' = u + \operatorname{Im}(ae^{-it}z + be^{-i\alpha t}\alpha w),\ t \in \mathbb{R}, z, w \in \mathbb{C} \right\}.$$

Note that this is consistent with these two Casimir functions preserving $|ae^{-it}| = |a|$ and $|be^{-i\alpha t}| = |b|$. ♦

Another illuminating example of a similar phenomenon (due to Juan Simo) is the semidirect product SL(3) Ⓢ SM(3), where SM(3) is the space of symmetric 3×3 matrices and the action of SL(3) on it is by similarity, $A \times A^T$. The Lie algebra and its dual are 14 dimensional and the generic coadjoint orbit is 12 dimensional. But there are no nontrivial Casimir functions because the closure of any orbit contains the origin, so one cannot separate two orbits by continuous functions.

Solution to Exercise N9F-1. Proceed as in the proof of the Duflo-Vergne Theorem, replacing the curve $\mu + t\nu$ by a curve $\mu(t) \in S$, with $\mu(0) = \mu$ and $\mu(1) = \nu \in S$. We assume that S is connected; if not, work on connected components. The proof remains unchanged till the end when the conclusion is that $\langle \mu'(0), [\xi, \eta] \rangle = 0$ for all $\xi, \eta \in \mathfrak{g}_\mu$. Since $\mu'(0)$ is an arbitrary vector in $T_\mu S$, this implies that $[\mathfrak{g}_\mu, \mathfrak{g}_\mu] \in (T_\mu S)^0$.

References

References

Abarbanel, H. D. I. and D. D. Holm [1987] Nonlinear stability analysis of inviscid flows in three dimensions: incompressible fluids and barotropic fluids, *Phys. Fluids* **30**, 3369–3382.

Abarbanel, H. D. I., D. D. Holm, J. E. Marsden, and T. S. Ratiu [1986] Nonlinear stability analysis of stratified fluid equilibria. *Phil. Trans. Roy. Soc. London A* **318**, 349–409; also Richardson number criterion for the nonlinear stability of three-dimensional stratified flow. *Phys. Rev. Lett.* **52** [1984], 2552–2555.

Abraham, R. and J. E. Marsden [1978] *Foundations of Mechanics.* Second Edition, Addison-Wesley.

Abraham, R., J. E. Marsden, and T. S. Ratiu [1988] *Manifolds, Tensor Analysis, and Applications.* Second Edition, Applied Mathematical Sciences **75**, Springer-Verlag.

Abraham, R. and Robbin, J. [1967] *Transversal mappings and flows*. Benjamin-Cummings, Reading, Mass.

Adams, J. F. [1969] *Lectures on Lie groups.* Benjamin-Cummings, Reading, Mass.

Adams, R.A. [1975] *Sobolev Spaces.* Academic Press.

Adams, M. R., T. S. Ratiu, and R. Schmid [1986] A Lie group structure for pseudodifferential operators. *Math. Ann.* **273**, 529–551 and A Lie group structure for Fourier integral operators. *Math. Ann.* **276**, 19–41.

Arms, J.M., J.E. Marsden, and V. Moncrief [1981] Symmetry and bifurcations of momentum mappings, *Comm. Math. Phys.* **78**, 455–478.

Arms, J. M., J. E. Marsden, and V. Moncrief [1982] The structure of the space solutions of Einstein's equations: II Several Killings fields and the Einstein-Yang-Mills equations. *Ann. of Phys.* **144**, 81–106.

Arnold, V. I. [1967] Characteristic class entering in conditions of quantization. *Funct. Anal. Appl.* **1**, 1–13.

Atiyah, M. [1982] Convexity and commuting Hamiltonians. *Bull. London Math. Soc.* **14**, 1–15.

Aubin, T. [1976] Espaces de Sobolev sur les varièt̀es riemanniennes. *Bull. Sci. Math.* **100**, 149–173.

Bates, S. and A. Weinstein [1997] *Lectures on the Geometry of Quantization*, CPAM/UCB, Am. Math. Soc.

Bialynicki-Birula, I., J. C. Hubbard, and L. A. Turski [1984] Gauge-independent canonical formulation of relativistic plasma theory. *Physica A* **128**, 509–519.

Cantor, M. [1975] Perfect fluid flows over $\mathbb{R}^n$ with asymptotic conditions. *J. Func. Anal.* **18**, 73–84.

Chernoff, P. R. and J. E. Marsden [1974] *Properties of Infinite Dimensional Hamiltonian systems.* Springer Lect. Notes in Math. **425**.

de Gosson, M. [1997] Maslov classes, metaplectic representation and Lagrangian quantization. *Mathematical Research*, **95**. Akademie-Verlag, Berlin.

Duflo, M. and M. Vergne [1969] Une proprieté de la représentation coadjointe d'une algébre de Lie. *C.R. Acad. Sci. Paris* **268**, 583–585.

Ebin, D. G. [1970] On the space of Riemannian metrics. *Symp. Pure Math., Am. Math. Soc.* **15**, 11–40.

Ebin, D. G. and J. E. Marsden [1970] Groups of diffeomorphisms and the motion of an incompressible fluid. *Ann. Math.* **92**, 102–163.

Eckmann J.-P. and R. Seneor [1976] The Maslov-WKB method for the anharmonic oscillator, *Arch. Rat. Mech. Anal.* **61** 153–173.

Fischer, A. E., J. E. Marsden, and V. Moncrief [1980] The structure of the space of solutions of Einstein's equations, I: One Killing field. *Ann. Inst. H. Poincaré* **33**, 147–194.

Friedman, A. [1969] *Partial differential equations.* Holt, Rinehart and Winston.

Gotay, M. J., H. B. Grundling, G. M. Tuynman [1996] Obstruction results in quantization theory. *J. Nonlinear Sci.* **6**, 469–498.

Guillemin, V. and S. Sternberg [1977] *Geometric Asymptotics.* Amer. Math. Soc. Surveys, **14**. (Revised edition, 1990.)

Guillemin, V. and S. Sternberg [1982] Convexity properties of the moment map. *Inv. Math.* **67**, 491-513, **77** (1984) 533–546.

Guillemin, V. and S. Sternberg [1984] *Symplectic Techniques in Physics.* Cambridge University Press.

Hamilton, R. [1982] The Inverse function theorem of Nash and Moser. *Bull. Am. Math. Soc.* **7**, 65–222.

Hermann, R. [1962] The differential geometry of foliations, *J. Math. Mech.* **11**, 303–315.

Holm, D. D., J. E. Marsden, T. S. Ratiu, and A. Weinstein [1985] Nonlinear stability of fluid and plasma equilibria. *Phys. Rep.* **123**, 1–116.

Kaufman, A. [1982] Elementary derivation of Poisson structures for fluid dynamics and electrodynamics. *Phys. Fluids* **25**, 1993–1994.

Kaufman, A. and R. L. Dewar [1984] Canonical derivation of the Vlasov-Coulomb noncanonical Poisson structure. *Contemp. Math.* **28**, 51–54.

Kopell, N. [1970] Commuting diffeomorphisms. *Proc. Sympos. Pure Math., AMS*, **14**, 165–184.

Ladyzhenskaya, O. A. [1969] *The mathematical theory of viscous incomprehensible flow.* Second English edition, revised and enlarged. Math. and its Appl. **2** Gordon and Breach.

Lewis, D., J. E. Marsden, R. Montgomery, and T. S. Ratiu [1986] The Hamiltonian structure for dynamic free boundary problems, *Physica D* **18**, 391–404.

Libermann, P. and C. M. Marle [1987] *Symplectic Geometry and Analytical Mechanics.* Kluwer Academic Publishers.

Littlejohn, R. G. [1988] Cyclic evolution in quantum mechanics and the phases of Bohr-Sommerfeld and Maslov. *Phys. Rev. Lett.* **61**, 2159–2162.

Marsden, J. E. [1973a] A proof of the Calderon extension theorem. *Can. Math. Bull.* **16**, 133-136.

Marsden, J. E., D. G. Ebin, and A. Fischer [1972] Diffeomorphism groups, hydrodynamics and relativity. *Proceedings of the 13th Biennial Seminar on Canadian Mathematics Congress*, pp. 135–279.

Marsden, J.E. and M.J. Hoffman [1993] *Elementary Classical Analysis.* Second Edition, W.H. Freeman and Co, NY.

Marsden, J. E. and T. J. R. Hughes [1983] *Mathematical Foundations of Elasticity.* Prentice-Hall. Dover edition [1994].

Marsden, J. E., P. J. Morrison, and A. Weinstein [1984] The Hamiltonian structure of the BBGKY hierarchy equations. *Cont. Math. AMS* **28**, 115–124.

Marsden, J. E., T. S. Ratiu, and G. Raugel [1991] Symplectic connections and the linearization of Hamiltonian systems. *Proc. Roy. Soc. Edinburgh A* **117**, 329-380

Marsden, J. E., T. S. Ratiu and S. Shkoller [1999] *The geometry and analysis of the averaged Euler equations with normal boundary conditions*, in preparation.

Marsden, J.E. and A. Weinstein [1979], Review of *Geometric Asymptotics* and *Symplectic Geometry and Fourier Analysis*, *Bull. Amer. Math. Soc.* **1**, 545-553.

Maslov, V. P. [1965] *Theory of Perturbations and Asymptotic Methods.* Moscow State University.

Meyers, N. and J. Serrin [1960] The exterior Dirichlet problem for second order elliptic partial differential equations. *J. Math. Mech.* **9**, 513–538.

Miller, S. C. and R. H. Good [1953] A WKB-type approximation to the Schrödinger equation. *Phys. Rev.* **91**, 174–179.

Milnor, J. [1965] *Topology from the Differential Viewpoint.* University of Virginia Press.

Morrey, Jr. C.B. [1966] *Multiple Integrals in the Calculus of Variations.* Springer–Verlag.

Nirenberg, L. [1959] On elliptic partial differential equations. *Ann. Scuola. Norm. Sup. Pisa* **13**(3), 115–162.

Omori, H. [1974] *Infinite dimensional Lie transformation groups.* Lecture Notes in Mathematics, **427**. Springer-Verlag.

Palais, R. S. [1965] Seminar on the the Atiyah - Singer Index Theorem. Princeton University Press, Princeton, N.J.

Palais, R. S. [1968] *Foundations of Global Non-Linear Analysis.* Benjamin.

Rais, M. [1972] Orbites de la représentation coadjointe d'un groupe de Lie, *Représentations des Groupes de Lie Résolubles*, P. Bernat, N. Conze, Mj. Duflo, M. Lévy-Nahas, M. Aais, P. Renoreard, Mj. Vergne, eds. *Monographies de la Société Mathématique de France, Dunod, Paris* **4**, 15–27.

Ratiu, T. S. [1982] Euler-Poisson equations on Lie algebras and the N-dimensional heavy rigid body, *Am. J. Math.* **104**, 409–448, 1337.

Ratiu, T. S. and R. Schmid [1981] The differentiable structure of three remarkable diffeomorphism groups, *Math. Zeitschrift* **177**, 81–100.

Sánchez de Alvarez, G. [1986] *Thesis*. University of California at Berkeley.

Sánchez de Alvarez, G. [1989] Controllability of Poisson control systems with symmetry. *Cont. Math. AMS* **97**, 399–412.

Simo, J. C., D. R. Lewis, and J. E. Marsden [1991] Stability of relative equilibria I: The reduced energy momentum method, *Arch. Rat. Mech. Anal.* **115**, 15-59.

Simo, J. C., T. A. Posbergh, and J. E. Marsden [1990] Stability of coupled rigid body and geometrically exact rods: block diagonalization and the energy-momentum method, *Physics Reports* **193**, 280–360.

Simo, J. C., T. A. Posbergh, and J. E. Marsden [1991] Stability of relative equilibria II: Three dimensional elasticity, *Arch. Rat. Mech. Anal.* **115**, 61–100.

Slawianowski, J. J. [1971] Quantum relations remaining valid on the classical level. *Rep. Math. Phys.* **2**, 11–34.

Stefan, P. [1974] Accessible sets, orbits and foliations with singularities, *Proc. Lond. Math. Soc.* **29**, 699–713.

Stein, E. M. [1970] *Singular integrals and differentiability properties of functions.* Princeton Mathematical Series,**30** Princeton University Press.

Sussman, H. [1973] Orbits of families of vector fields and integrability of distributions. *Trans. Am. Math. Soc.* **180**, 171–188.

Tulczyjew, W. M. [1977] The Legendre transformation. *Ann. Inst. Poincaré* **27**, 101–114.

Varadarajan, V. S. [1974] *Lie Groups, Lie Algebras and Their Representations.* Prentice Hall. (Reprinted in Graduate Texts in Mathematics, Springer-Verlag.)

Weinstein, A. [1983] The local structure of Poisson manifolds. *J. Diff. Geom.* **18**, 523–557

H. Whitney [1943] Differentiable even functions. *Duke Math. J.* **10**, 159–160.

Woodhouse, N. M. J. [1992] *Geometric Quantization.*, Second Edition, Clarendon Press, Oxford University Press.

Yosida, K. [1980] *Functional analysis.* Sixth edition. Grundlehren **123**. Springer-Verlag.