LIE TRANSFORM TUTORIAL - II

KENNETH R. MEYER

1. Introduction.

This survey paper is an extension of Meyer(1988)[1] since it contains complete proofs of the main theorems and some generalizations of Lie transform theory. However, the first part of this paper deals with the applications of Lie transforms to various perturbation problems leaving the technical proofs to the later sections.

Over the years many different techniques have been developed for handing various perturbation problems. Some are suited for a few special problems while others are quite general, but almost all were developed before the computer age. To our knowledge only one general technique was developed specifically to be used in conjunction with a computer algebra system, namely the method of Lie transforms. It is truly an algorithm in the sense of modern computer science: a clearly defined iterative procedure.

The method was first given in Deprit(1969) for Hamiltonian systems of differential equations, then generalized to arbitrary systems of differential equations by Kamel(1970) and Henrard(1970). The predecessor of this method was a limited set of formulas given in Hori(1966). All these papers appeared in astronomy journals far from the usual journals of perturbation analysis. Through the seventies only a few papers on this subject appeared outside the astronomy literature. Recently, several books have presented the method but only in the limited context in which it was initially developed.

In this paper we would like to indicate the great generality of the method by illustrating how it can be used to solve perturbation problems that are typically solved by other methods, often special ad hoc methods. In most cases we have chosen the simplest standard examples. There are many topics of current research that are not considered here since this is to be a tutorial, not a summary of new results.

Below we will indicate how the method of Lie transforms can be used to: calculate the function given by the implicit function theorem; calculate the coordinates given in the splitting lemma of catastrophe theory; calculate the center and stable manifolds of a critical point; calculate a limit cycle or invariant torus; calculate the Poincare normal form for a center; do classical averaging to arbitrary order; calculate Floquet exponents; calculate the Darboux coordinates of symplectic geometry. All these seemingly distinct calculations can be done with one simple algorithm – the method of Lie transforms.

Most of the first part of the paper consists of examples of problems that can be solved by Lie transforms, without spending too much time on the derivation or the theory. One main

Date: June 3, 2006.

This research was supported by a grant from ACMP/DARPA administered by NIST..

This article appeared in *Computer Aided Proofs in Analysis* (eds. K. R. Meyer and D. S. Schmidt), IMA Volumes in Mathematics and its Applications 28, Springer-Verlag, New York, 1991.

[1]Actually the preceedings in which the orginal survey was to appear was never published.

theorem summarizes the power of the method and it is given in Section 2. The proof of this general theorem is postponed until Section 8. The middle sections are all examples.

Section 8 is written independently of most of the paper so if you are interested in the proof itself you can skip the examples and go directly from Section 2 to Section 8. On a first reading this might be the best approach.

2. **The Main Theorem of the Theory:**

In the traditional setting of perturbation theory one is given a differential equation depending on a small parameter ϵ. When $\epsilon = 0$ the differential equation is simple and well understood, say for example a harmonic oscillator. The problem is to understand the solutions of the equations when ϵ is non-zero but small. To gain generality think of any smooth tensor field defined on some open set $D \subset \mathbb{R}^n$ depending on a small parameter. The tensor field might be a function; a contravariant vector field, i.e. an ordinary differential equation; a covariant vector field, i.e. a differential form; a Riemannian metric; a symplectic structure; or any of the other classical tensors of differential geometry. The important thing about these objects is that there is a Lie derivative defined for them.

Let F be a smooth tensor field defined on an open set $D \in \mathbb{R}^n$ that is for each point $x \in D$ there is assigned a unique tensor, $F(x)$, of a fixed type say p-covariant and q-contravariant. Let W be a smooth autonomous ordinary differential equation defined on D, i.e. a contravariant vector field on D, and let $\phi(\tau, \xi)$ be the solution of the equation which satisfies $\phi(0, \xi) = \xi$. The Lie derivative, $\pounds_W F$, is simply the directional derivative of F in the direction of W and is a tensor field of the same type as F itself. The general definition is given in any non-elementary book on differential geometry and in Section 8. For now we shall simply give examples.

Differential geometry has used many different notations which still persist today making a general presentation difficult. For example the object W given above might be called an autonomous differential equation on D and so W is thought of as a smooth function from D into $\mathbb{R}^n$ and is denoted by

$$\frac{dx}{d\tau} = W(x). \tag{2.1}$$

Then W is considered as a column vector with components $W^1, \ldots, W^n$. In classical tensor terminology W is 1-contravariant and we write W^i where i is a free index ranging from 1 to n – here the superscript tells you it is contravariant. More recent notation is

$$\sum_{i=1}^{n} W^i(x) \frac{\partial}{\partial x^i} = W^1(x) \frac{\partial}{\partial x^1} + \cdots + W^n(x) \frac{\partial}{\partial x^n}. \tag{2.2}$$

In any case let $\phi(\tau, \xi)$ be the solution satisfying the initial condition $\phi(0, \xi) = \xi$. The simplest tensor field is a smooth function $f : D \to \mathbb{R}^1$, i.e. to each point of D you assign a scalar. The Lie derivative of f along W, $\pounds_W f$, is a smooth function from D to $\mathbb{R}^1$ also and is defined by

$$\pounds_W f(x) = \frac{\partial}{\partial \tau} f(\phi(\tau, \xi)) \bigg|_{\tau=0} = \nabla f(x) \cdot W(x), \tag{2.3}$$

the dot product of the gradient of f and W.

The next simplest tensor field is a vector field, either covariant or contravariant. First let χ be a contravariant vector field or differential equation on D. Using differential equation notation for χ we write

$$\chi : \dot{x} = F(x) \tag{2.4}$$

where $\dot{} = \dfrac{d}{dt}$. The column vector F is a representation of the contravariant vector field χ in the x coordinates. Do not confuse t and τ they are different parameters for different vector fields. Changing variables in (2.4) from x to ξ by $x = \phi(\tau, \xi)$ where τ is simply a parameter gives

$$\dot{\xi} = \left(\frac{\partial \phi}{\partial \xi}(\tau, \xi)\right)^{-1} F(\phi(\tau, \xi)) = G(\tau, \xi). \tag{2.5}$$

G is the representation of χ in the new coordinates system ξ. The Lie derivative, $\pounds_W \chi$, is defined by

$$\pounds_W \chi(x) = \frac{\partial}{\partial \tau} G(\tau, \xi)\bigg|_{\tau=0} = \frac{\partial F}{\partial x}(x) W(x) - \frac{\partial W}{\partial x}(x) F(x). \tag{2.6}$$

Note that x and ξ are the same when $\tau = 0$. $\pounds_W \chi$ is a smooth contravariant vector field on D. We usually abuse the notation and confuse the vector field χ with its representation F in a coordinate system by writing $\pounds_W F$ for (2.6).

Let η be a 1-covariant vector field on D, i.e. a differential form, so

$$\eta = \sum_{i=1}^{n} h_i(x) dx^i. \tag{2.7}$$

Think of h as the column vector $(h_1, \ldots, h_n)^T$ and change variables from x to ξ by $x = \phi(\tau, \xi)$ to get

$$\eta = \sum_{i=1}^{n} k_i(\xi) d\xi^i \tag{2.8}$$

where k is a column vector related to h by

$$k(\tau, \xi) = \frac{\partial \phi}{\partial \xi}(\tau, \xi)^T h(\phi(\tau, \xi)) \tag{2.9}$$

The vector k is the components of the differential form η in the new coordinates ξ. The Lie derivative of η in the direction of W, $\pounds_W \eta$, is a one form whose component vector is given by

$$\pounds_W \eta(x) = \frac{\partial}{\partial \tau} k(\tau, \xi)\bigg|_{\tau=0} = \frac{\partial h}{\partial x}(x)^T W + \frac{\partial W}{\partial x}(x)^T h(x). \tag{2.10}$$

The Lie derivative of other tensor fields in the direction W are defined in the same way and the reader can find a complete discussion in any book on differential geometry.

Let $\mathcal{T}_{pq} = \mathcal{T}_{pq}(D)$ denote the vector space of all smooth p-covariant and q-contravariant tensor fields D. A symmetric notation for $\pounds_W K$ is $[K, W]$, the Lie bracket of K and W. For

fixed W the map $\pounds_W = [\cdot, W]$ is a linear operator from $\mathcal{T}_{pq}$ into itself. The set, $\mathcal{V} = \mathcal{V}(D) = \mathcal{T}_{01}(D)$, of all smooth contravariant vector fields on D is a vector space and $[K, \cdot]$, for fixed K, is a linear from $\mathcal{V}$ into $\mathcal{T}_{pq}$. Thus $[\cdot, \cdot] : \mathcal{T}_{pq} x \mathcal{V} \to \mathcal{T}_{pq}$ is bilinear.

Suppose that the perturbation problem is given as a tensor field Z_* on D which has a formal expansion in a small parameter ϵ. In many cases ϵ is simply a scale parameter. Consider

$$Z_* = Z_*(x, \epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!} \right) Z_j^0(x) \tag{2.11}$$

where each Z_j^0 is a tensor field of fixed type. Specifically assume that

$$Z_j^0 \in \mathcal{P}_j \subset \mathcal{T}_{pq}, \text{for } j = 0, 1, 2, \ldots \tag{2.12}$$

where $\mathcal{P}_j$ is a linear subspace of $\mathcal{T}_{pq}$. In order to simplify the problem the method of normal forms seeks a near identity change of variables of the form $x = \xi + O(\epsilon)$ such that the tensor field Z_* in the new coordinates is simpler. The traditional approach is simple: assume a general series for the change of variables, substitute it in the series for Z_*, collect terms, and try to choose the coefficients in the change of variables series so that the tensor Z_* in the new coordinates is as simple as possible. For simple problems that will suffice, however there are several disadvantages to this approach. The bookkeeping of the terms of the series can become a major problem especially if the problem has some special structure or symmetry. For example if Z_* is a Hamiltonian vector field one would want the vector field in the new coordinates to be Hamiltonian also. Or if Z_* is invariant under some symmetry group one would want this to be true in the new coordinates also. Figuring out what the form of the n^{th} term in new series can be quite difficult using the straight plug and chug method. Also, this procedure is not easily coded in a symbolic computer language.

Hori(1966) was interested in perturbation theory for Hamiltonian vector fields and suggested that the near identity transformation be given as the solution of an autonomous ordinary differential equation. Unfortunately, not all near identity transformations are solutions of autonomous equations and so Hori was not able to develop a general theory. Deprit(1969) took Hori's idea one step further by using non-autonomous equations. He was able to give a simple set of recursive formulas that overcomes the objections given above. Hori and Deprit worked with Hamiltonian systems, but soon afterwards Kamel(1970) and Henrard(1970) considered the general case.

Thus to simplify the perturbation problem given by Z_* in (2.11) we seek a near identity change of coordinates of the form

$$x = x(\xi, \epsilon) = \xi + \cdots \tag{2.13}$$

where $x(\xi, \epsilon)$ is constructed as a formal solution of the system of equations and initial conditions

$$\frac{dx}{d\epsilon} = W(x, \epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!} \right) W_{j+1}(x), \qquad x(0) = \xi. \tag{2.14}$$

It can easily be shown that for any change of coordinates of the form (2.13) there is a unique differential equation of the form (2.14) for which it is the solution function. The W above is

a smooth vector field on D for each ϵ, so we take

$$W_j \in \mathcal{R}_i \subset \mathcal{V}, \text{for all } i = 0, 1, 2, \dots \tag{2.15}$$

where $\mathcal{R}_i$ is a linear subspace of $\mathcal{V}$, the space of smooth vector fields on D. The problem defined by Z_* may have some special symmetry, like a reflective symmetry, or a special structure, like being Hamiltonian, and this is reflected in the assumption that we have identified the subspace $\mathcal{P}_i$ to which the Z_i belong. To preserve this symmetry or structure it may be necessary to restrict the change of variables by requiring the W_i to lie in the subspaces $\mathcal{R}_i$.

In the new coordinates ξ the tensor $Z_*(x, \epsilon)$ becomes

$$Z^* = Z^*(\xi, \epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!} \right) Z_0^j(\xi). \tag{2.16}$$

We say (2.13) or (2.14) transforms (2.11) into (2.16). Also we shall say the tensor Z^* in (16) is in normal form and hence simplified by definition if we have identified subspaces $\mathcal{Q}_i, i = 1, 2, \dots$ such that

$$Z_0^i \in \mathcal{Q}_i \subset \mathcal{P}_i, \text{for } i = 1, 2, 3, \dots . \tag{2.17}$$

The fundamental theorem of the theory is:

Theorem 2.1. *Assume (i) $[\mathcal{P}_i, \mathcal{R}_j] \subset \mathcal{P}_{i+j}$ for $i, j = 1, 2, 3, \dots$ and (ii) for any $i = 1, 2, 3, \dots$ and for any $A \in \mathcal{P}_i$ there exists $B \in \mathcal{Q}_i$ and $C \in \mathcal{R}_i$ such that*

$$B = A + [Z_0^0, C]. \tag{2.18}$$

Then one can compute a formal expansion for W as given in (2.14) with $W_i \in \mathcal{R}_i$ for all i which transforms (2.11) to (2.16) where $Z_0^i \in \mathcal{Q}_i$ for all i.

The proof of this theorem in almost this level of generality can be found in Meyer and Schmidt(1977) and is given in sightly more generality in Section 8, see Theorem 8.2.. The proof is completely constructive in the sense that an effective algorithm is given to find the expansion of W and Z^* term by term. In practice Z_0 is given and so one takes the subspaces $\mathcal{P}_i$ as small as possible. The spaces $\mathcal{Q}_i$ and $\mathcal{R}_i$ come from an analysis of the equation in (2.18).

3. Function Applications.

In this section we will show some applications of the method of Lie transforms when the problem involves simply functions as opposed to vector fields.

The implicit function theorem. One of the fundamental theorems of analysis is the implicit function theorem. We will show how to compute the implicitly defined function using Lie transforms.

Consider a function (or formal power series) $f(u, x)$ defined in neighborhood of the origin in $\mathbb{R}^m \times \mathbb{R}^n$ into $\mathbb{R}^n$ such that $f(0, 0) = 0$ and $\frac{\partial f}{\partial x}(0, 0) = D$ is nonsingular. Then the implicit function theorem asserts that there is an analytic function (or formal power series) $\psi(u)$ defined in a neighborhood of the origin in $\mathbb{R}^m$ into $\mathbb{R}^n$such that $\psi(0) = 0$ and $f(u, \psi(u)) \equiv 0$.

Introduce a small parameter ϵ by scaling $u \to \epsilon^2 u, x \to \epsilon x$ and $f \to \epsilon^{-1} f$, that is define F_* by

$$F_*(u,x,\epsilon) = \epsilon^{-1} f(\epsilon^2 u, \epsilon x) = \sum_{i=0}^{\infty} \left(\frac{\epsilon^i}{i!}\right) F_i^0(u,x) \tag{3.1}$$

and $F_0^0(u,x) = Dx$. Let x be the variable and treat u simply as a parameter in the problem. The functions $F_i^0(u,x)$ are vector of polynomials in u and x and so let $\mathcal{P}_i$ be the vector space of such vectors of polynomials in u and x.

By Theorem 2.1 we must be able to solve (2.18) where A is any polynomial. In this case the Lie bracket is $[F_0^0, C] = \mathrm{DC}$. Clearly we can solve $[F_0^0, C] + A = B$ by taking $B = 0$ and $C = D^{-1}A$. Thus if we define $\mathcal{Q}_i = \{0\}$ and $\mathcal{R}_i = \mathcal{P}_i$, then for any $A \in \mathcal{P}_i$, we can solve (2.18) for $B \in \mathcal{Q}_i = \{0\}$ and $C \in \mathcal{R}_i = \mathcal{P}_i$. Thus one can compute a transformation such that $F^*(u,\xi,\epsilon) = D\xi$. But $F^*(u,\xi,\epsilon) = F_*(u,\phi(u,\xi,\epsilon),\epsilon) = \epsilon^{-1} f(\epsilon^2 u, \epsilon\phi(u,\xi,\epsilon))$. So $\phi(u,o,1) = \psi(u)$ satisfies $f(u,\psi(u)) \equiv 0$. This shows that the implicit function can be computed by Lie transforms. In general the method of Lie transforms only produces a formal series, but in this case the implicit function theorem assures that formal series converges when the series for f does. Note that the parameter ϵ was only used to order the terms in the series since it was set to 1 in the end.

The splitting lemma. The splitting lemma is an important tool in the analysis of critical points of a function and catastrophe theory (see Poston and Stewart(1978)). Let $V(x)$ be a real valued analytic function defined in a neighborhood of the origin in $\mathbb{R}^n$ and $x \in \mathbb{R}^n$. Assume that the origin is a critical point for V and for simplicity assume that $V(0) = 0$. Assume that the rank of the Hessian, $\dfrac{\partial^2 V}{\partial x^2}(0)$, is s, $0 \le s \le n$. Then the splitting lemma says that there is a change of coordinates $x = \phi(y)$ such that in the new coordinates

$$V(y) = \frac{1}{2}(\pm y_1^2 \pm \cdots \pm y_s^2) + v(y_{s+1}, ..., y_n). \tag{3.2}$$

Scale by $x \to \epsilon x$, and $V \to \epsilon^{-2} V$ or define

$$U_*(x,\epsilon) = \epsilon^{-2} V(\epsilon x) = \sum_{i=0}^{\infty} \left(\frac{\epsilon^i}{i!}\right) U_i^0(x). \tag{3.3}$$

Here the $U_i^0(x,\mu)$ are polynomials in x of degree $i+2$, so let $\mathcal{P}_i$ be the vector space of such polynomials. $U_0^0(x)$ is a quadratic form in x and so by making a linear change of variable if necessary we may assume that

$$U_0^0(x) = \frac{1}{2}(\pm x_1^2 \pm x_2^2 \pm \cdots \pm x_s^2). \tag{3.4}$$

To solve (2.18) let

$$C = c x_1^{k_1} \cdots x_n^{k_n}. \tag{3.5}$$

be a monomials of degree $i+2$ and where $c = (c_1, \ldots, c_n)^T$ is an n-vector. Then

$$[U_0^0, C] = \pm c_1 x_1^{k_1+1} x_2^{k_2} \cdots x_n^{k_n} \pm \cdots \pm c_s x_1^{k_1} x_2^{k_2} \cdots x_s^{k_s+1} \cdots x_n^{k_n} \tag{3.6}$$

so the kernel of $[U_0^0, C]$ consists of all homogeneous polynomials of degree $i+2$ in $x_s, \ldots, x_n$ and the range of $[U_0^0, C]$ consists of the span of all monomials which contain one of $x_1, \ldots, x_s$

to a positive power or equivalently those polynomials which are zero when $x_1 = \cdots = x_s = 0$. Thus we can solve (2.18) by taking $\mathcal{P}_i$ as the space of all scalar homogeneous polynomials of degree $i+2$, $\mathcal{Q}_i$ the subspace of $\mathcal{P}_i$ consisting of all scalar homogeneous polynomials of degree $i+2$ in $x_s, \ldots, x_n$ alone, and $\mathcal{R}_i$ the space of all n-vectors of homogeneous polynomials of degree $i+1$ in $x_1, \ldots, x_n$.

Thus the method of Lie transforms will construct a change of coordinates so that in the new coordinate

$$U^*(y,\epsilon) = \sum_{i=0}^{\infty} \left(\frac{\epsilon^i}{i!}\right) U_0^i(y) \tag{3.7}$$

where for $i \geq 1$ the $U_0^i(y)$ depend only on $y_s, \ldots, y_n$. Setting $\epsilon = 1$ gives the form given by the splitting lemma in (3.2).

In Meyer and Schmidt(1988) the problem for finding bifurcations of relative equilibria in the N-body problem was reduced to finding the bifurcation of critical points of the potential constrained to a constant moment of inertia manifold. The constraint equation was solved by the method of Lie transforms to compute the implicitly defined function. Then by applying the splitting lemma algorithm we obtained the bifurcation equations in a form that could be analyzed by hand.

4. Autonomous Differential Equations

In this section we will show how Theorem 2.1 can be used to study autonomous differential equations. There are many more applications than the ones given here.

The classical normal form. Consider the equation

$$\dot{x} = Lx + f(x) \tag{4.1}$$

where $x \in \mathbb{R}^n$, L is an $n \times n$ constant matrix, f is an analytic function defined in a neighborhood of the origin in $\mathbb{R}^n$ whose series expansion starts with second degree terms. Scale the equations by $x \to \epsilon x$ and divide the equation by ϵ so that equation (4.1) becomes

$$\dot{x} = \sum_{i=0}^{\infty} \left(\frac{\epsilon^i}{i!}\right) F_i^0(x), \tag{4.2}$$

where $F_0^0(x) = Lx$ and F_i^0 is an n-vector of homogeneous polynomials of degree $i+1$ so let $\mathcal{P}_i$ be the space of all such polynomials.

Assume that L is diagonal so $L = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$. In order to solve (2.18) let

$$\begin{aligned} &A = ax^k, \quad B = bx^k, \quad C = cx^k \\ &k = (k_1, \ldots, k_n), \quad x = (x_1, \ldots, x_n), \quad x^k = x_1^{k_1} \cdots x_n^{k_n} \end{aligned} \tag{4.3}$$

and substitute into (2.18) to get

$$bx^k = ax^k + (L - (\Sigma k_s \lambda_s) I) c x^k. \tag{4.4}$$

The coefficient matrix, $L-(\Sigma k_s\lambda_s)I$, of cx^k is diagonal with entries $\lambda_j-\Sigma k_s\lambda_s$. So to solve (2.18) take

$$
\begin{aligned}
c_j &= \frac{-a_j}{\lambda_j-\sum k_s\lambda_s}, \quad b_j=0 \quad \text{when} \quad \lambda_j-\sum k_s\lambda_s\neq 0 \\
c_j &= 0, \quad b_j=a_j \quad \text{when} \quad \lambda_j-\sum k_s\lambda_s=0
\end{aligned} \tag{4.5}
$$

Let $e_j=(0,\dots,0,1,0,\dots,0)^T$ be the standard basis for $\mathbb{R}^n$. From the above we define

$$
\begin{aligned}
\mathcal{Q}_i &= \text{span}\,\{e_jx^k : \lambda_j-\sum k_s\lambda_s=0,\ \sum k_k=i+1\} \\
\mathcal{R}_i &= \text{span}\,\{e_jx^k : \lambda_j-\sum k_s\lambda_s\neq 0,\ \sum k_k=i+1\}
\end{aligned} \tag{4.6}
$$

so the condition in (ii) of the Theorem 2.1 is satisfied. So (4.2) can be formally transformed to

$$
\dot{y}=\sum_{i=0}^{\infty}\left(\frac{\epsilon^i}{i!}\right)F_0^i(y), \tag{4.7}
$$

where $F_0^i\in\mathcal{Q}_i$ for all $i\geq 1$. Setting $\epsilon=1$ brings the equations to the form

$$
\dot{y}=Ly+g(y) \tag{4.8}
$$

where the terms in g lie in some $\mathcal{Q}_i$. It is easy to check that a term $h(y)$ is in some $\mathcal{Q}_i$ if and only if $h(e^{Lt}y)=e^{Lt}h(y)$ for all y and t. Thus g in (4.8) satisfies

$$
g(e^{Lt}y)=e^{Lt}g(y). \tag{4.9}
$$

This formulation for the normal form does not require that L be in diagonal form (L must be diagonalizable!) . This is the classical normal form as found in Diliberto(1961) et al. For example if $n=3$ and

$$
L=\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{4.10}
$$

so L has eigenvalues -1, and $\pm i$ then the normal form is

$$
\begin{aligned}
\dot{u} &= -v-va(u^2+v^2)+ub(u^2+v^2) \\
\dot{v} &= u+ua(u^2+v^2)+vb(u^2+v^2) \\
\dot{w} &= -w+wc(u^2+v^2),
\end{aligned} \tag{4.11}
$$

where the a,b and c are arbitrary series. This normal form yields the so called center manifold since the plane $w=0$ is invariant and the equations on this center manifold are in Poincare's normal form for a center.

Invariant tori. Consider a system of coupled van der Pol equations written in polar coordinates. Or more generally a system of the form

$$\begin{aligned} \dot{r} &= R_*(r,\theta,\epsilon) = \sum_{i=0}^{\infty} R_i^0(r,\theta) \\ \dot{\theta} &= \Theta_*(r,\theta,\epsilon) = \sum_{i=0}^{\infty} \Theta_i^0(r,\theta) \end{aligned} \tag{4.12}$$

where r is a m-vector, θ is an n-vector of angles, R_i^0 and Θ_i^0 have finite Fourier series in the θ's with coefficients which are polynomials in the r variables. Let $\mathcal{P}_i$ be the space of all such functions.

Assume that $\Theta_0^0 = \omega$ is a constant vector, $R_0^0 = P(r)$ and that there exist a constant vector r_0 such that $P(r_0) = 0$ and $\dfrac{\partial P}{\partial r}(r_0)$ has no eigenvalue with zero real part. Then there is a formal change of variables $(r,\theta) \to (\rho,\phi)$ such that the equations (4.12) are of the form

$$\begin{aligned} \dot{\rho} &= R^*(\rho,\phi,\epsilon) = \sum_{i=0}^{\infty} R_0^i(\rho,\phi) \\ \dot{\phi} &= \Phi^*(\rho,\phi,\epsilon) = \sum_{i=0}^{\infty} \Phi_0^i(\rho,\phi) \end{aligned} \tag{4.13}$$

where R^* and Φ^* are like R_* and Θ_* but have the additional property that

$$R^*(r_0,\phi,\epsilon) \equiv 0 \quad \text{and} \quad \Phi^*(r_0,\phi+\omega t,\epsilon) \equiv 0. \tag{4.14}$$

The first condition in (4.14) says that $r = r_0$ is an invariant torus for the equations (4.13) and the second condition says that the equations on the invariant torus are in normal form. If there are no resonances among the frequencies ω then $\Phi^*(r_0,\phi,\epsilon) \equiv 0$.

Here

$$\begin{aligned} &Z_0^0 = \begin{pmatrix} P(r) \\ \omega \end{pmatrix}, \quad C = \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} u(r)e^{ik\theta} \\ v(r)e^{ik\theta} \end{pmatrix} \\ &A = \begin{pmatrix} a(r)e^{ik\theta} \\ \alpha(r)e^{ik\theta} \end{pmatrix}, \qquad B = \begin{pmatrix} b(r)e^{ik\theta} \\ \beta(r)e^{ik\theta} \end{pmatrix} \\ &k\theta = k_1\theta_1 + \cdots + k_n\theta_n \end{aligned} \tag{4.15}$$

then

$$
\begin{aligned}
[Z_0^0, C] &= \begin{pmatrix} \frac{\partial P}{\partial r} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} - \begin{pmatrix} \frac{\partial U}{\partial r} & \frac{\partial U}{\partial \theta} \\ \frac{\partial V}{\partial r} & \frac{\partial V}{\partial \theta} \end{pmatrix} \begin{pmatrix} P \\ \omega \end{pmatrix} = \\
&= \begin{pmatrix} \frac{\partial P}{\partial r} u e^{ik\theta} - \frac{\partial u}{\partial r} - iuk\omega e^{ik\theta} \\ -\frac{\partial v}{\partial r} e^{ik\theta} P - ivk\omega e^{ik\theta} \end{pmatrix} = \begin{pmatrix} (b-a)e^{ik\omega} \\ (\beta - \alpha)e^{ik\theta} \end{pmatrix}.
\end{aligned} \tag{4.16}
$$

To solve the second set of equations take

$$
\begin{aligned}
v &= \frac{\alpha}{k\omega} \qquad \beta = \frac{d\alpha}{dr}\frac{P}{k\omega} \quad \text{when} \quad k\omega \neq 0 \\
v &= 0 \qquad \beta = \alpha \quad \text{when} \quad k\omega = 0.
\end{aligned} \tag{4.17}
$$

For the first equation in (4.16) first let $D = \frac{\partial P}{\partial r}(r_0)$ and note that $u = (D - ik\omega I)^{-1}$ solves $Du - ik\omega u = -a$ for all k since D has no eigenvalue with zero real part by assumption. So we take

$$
\begin{aligned}
u &= (D - ik\omega)^{-1} a \\
b &= \left(\frac{\partial P}{\partial r}(r) - \frac{\partial P}{\partial r}(r_0) \right) u - \frac{\partial u}{\partial r} P(r)
\end{aligned} \tag{4.18}
$$

This formulas satisfy the equations and clearly $b(r_0) = 0$. The space $\mathcal{Q}_i$ is the span of all the solutions given for B and the space $\mathcal{R}_i$is the span of all the solutions given for C above. Thus we have verified the conditions of the Theorem 2.1. This was the procedure used in Meyer and Schmidt(1977) to calculate the regions in parameter space where two coupled van der Pol oscillators had frequencies that were locked in. The so called entrainment domains.

5. Non-Autonomous Differential Equations

In many applications the differential equations involve time explicitly so one must consider equations of the form $\dot{x} = f(t, x)$. In this case one would allow the transformation generated by W to depend on t also. But this case can be reduced to the previous case by replacing the original system with the equivalent autonomous system $\dot{x} = f(\tau, x)$, $\dot{\tau} = 1$ where τ is a new variable.

Consider the system

$$
\dot{x} = Z_*(t, x, \epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!} \right) Z_j^0(t, x), \tag{5.1}
$$

and the near identify transformation

$$
x = x(t, \xi, \epsilon) = \xi + \cdots \tag{5.2}
$$

generated as a solution of the equation

$$\frac{dx}{d\epsilon} = W(t,x,\epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) W_{j+1}(t,x), \qquad x(0) = \xi \tag{5.3}$$

which transforms (5.1) to

$$\dot{\xi} = Z^*(t,\xi,\epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) Z_0^j(t,\xi). \tag{5.4}$$

The translation of the Theorem 2.1 to the non-autonomous case goes as follows.

Theorem 5.1. *Let $\mathcal{P}_j$ ($\mathcal{R}_j$ respectively) be linear spaces of smooth time dependent tensor (respectively vector) fields defined for $j = 1, 2, ...$, $x \in D \subset \mathbb{R}^n$ and $t \in \mathbb{R}$ and let $\mathcal{Q}_j$ be a subspace of $\mathcal{P}_j$. If (i) $Z_j^0 \in \mathcal{P}_j$ for $j = 0, 1, 2, ...$, (ii) $[\mathcal{P}_i, \mathcal{R}_j] \subset \mathcal{P}_{i+j}$, $i, j = 0, 1, 2, ...$ (iii) for any $i = 1, 2, 3, ...$ and any $A \in \mathcal{P}_i$ there exist $B \in \mathcal{Q}_i$ and $C \in \mathcal{R}_i$ such that*

$$B = A + [Z_0^0, C] - \dot{C} \tag{5.5}$$

then one can construct W as in (5.3) with $W_i \in \mathcal{R}_i$ which generates a transformation (5.2) which takes (5.1) to (5.4) with $Z_0^i \in \mathcal{Q}_i$.

The method of averaging. The method of averaging is a special case of the normal form theorem given above. The method of averaging deals with a periodic system of the form (5.1) where $Z_0^0 = 0$, i.e. $\dot{x} = \epsilon Z_1^0(t,x) + \cdots$. One seeks a periodic change of variables, so the function W must be periodic in t also. Equation (5.5) reduces to $B = A - \dot{C}$. Given a periodic A in order to have a periodic C it is necessary and sufficient that we take B as the average over a period of A, so B is independent of t, and C as any indefinite integral of $A - B$. This shows that the normalized equation (5.4) are autonomous, i.e. Z_0^i is independent of t. The name comes from the fact that Z_0^1 is the time average of Z_1^0.

The Floquet exponents and the Liapunov transformation. A classical problem is to compute the characteristic exponents of Mathieu's equation $\ddot{x} + (a + b\cos 2\pi t)x = 0$ or other similar linear periodic systems. Assume that $Z_0^0(t,x) = Lx$ where L is diagonal matrix $L = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $Z_i^0(t,x) = A_i(t)x$ where $A_i(t)$ in an $n \times n$ T-periodic matrix, so let $\mathcal{P}_i$ be the space of all linear T-periodic systems. Seek a linear T-periodic change of variables, so seek $W_i(t,x) = C_i(t)x$ where $C_i(t)$ is to be T-periodic also and take $\mathcal{R}_i$ be the space $\mathcal{P}_i$. Equation (5.5) becomes

$$B(t) = A(t) + C(t)L - LC(t) - \dot{C}(t) \tag{5.6}$$

where A, B and C are matrices. The equation for the ij^{th} component is

$$b_{ij} = a_{ij} + (\lambda_i - \lambda_j)c_{ij} - \dot{c}_{ij}. \tag{5.7}$$

This is a linear first order differential equation in c_{ij}. If $(\lambda_i - \lambda_j)T \neq n2\pi i$ then take b_{ij} to be the average of a_{ij} and c_{ij} the unique T-periodic solution of (5.7). Thus the space $\mathcal{Q}_i$ is all linear systems with constant diagonal coefficient matrices. Thus we can compute a linear periodic change of coordinates which reduces the linear periodic system (5.1) to the linear diagonal constant system (5.4), this transformation is known as the Liapunov transformation. The entries on the diagonal are the Floquet exponents. The equation (5.6) has been studied

in the more general case when L is not necessarily diagonal. The presentation given here is merely a simple example.

A very similar problem is to calculate the series expansion of a solution of a linear differential equation at a regular singular point.

6. The Computational Darboux Theorem

To our knowledge the method of Lie transforms has not been used on tensor fields more complicated than vector fields. Here we will give a somewhat frivolous example to illustrate the generality of the method. In order to avoid the notational overload found in modern treatises like Kobayashi and Nomizu(1963) or Abraham and Marsden(1978), we shall use classical tensor notation. Thus repeated indices are summed over. Since the problem is a computational one we must use coordinates in the end anyway. Flanders(1963) is a highly recommended introduction to differential forms. The fundamental geometry of Hamiltonian mechanics is embodied in a *symplectic structure*, Ω, i.e. a closed, non-degenerate 2-form. In a neighborhood of the origin in $\mathbb{R}^{2n}$

$$\Omega = \Omega_{ij}(x)dx^i \wedge dx^j \tag{6.1}$$

where we have used the summation convention, $\Omega_{ij} = -\Omega_{ji}$, and the $\Omega_{ij}(x)$ are real analytic in x. $\{\Omega_{ij}\}$ is a 2-covariant tensor, so if you change coordinates by $x = x(y)$ then the tensor in the y coordinates is

$$\Omega(y) = \Omega_{ij}(x(y))\frac{\partial x^i}{\partial y^m}\frac{\partial x^j}{\partial y^n}dy^m \wedge dy^n. \tag{6.2}$$

Sometimes we will think of $\Omega(x)$ as the skew-symmetric matrix $(\Omega_{ij}(x))$, the coefficient matrix of the form (6.1). Ω is non-degenerate means that the matrix $\Omega(x)$ is nonsingular for all x. (6.1) means that the matrix Ω transforms by

$$\Omega \to \frac{\partial x}{\partial y}^T \Omega \frac{\partial x}{\partial y}. \tag{6.3}$$

Ω is closed means that

$$d\Omega = \frac{\partial \Omega_{ij}}{\partial x_k}dx^i \wedge dx^j \wedge dx^k = 0. \tag{6.4}$$

Since we are working locally, a closed form is exact by Poincare's lemma so there is a one form $\alpha(x) = \alpha_i(x)dx^i$ such that $\Omega = d\alpha$.

This matrix $\Omega(0)$ is nonsingular and skew symmetric so there is a nonsingular matrix P such that

$$P^T\Omega(0)P = J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \tag{6.5}$$

which means that after a linear change of coordinates the coefficient matrix of $\Omega(0)$ is J. Darboux's theorem says there is a nonlinear change of coordinates defined in a neighborhood of the origin in $\mathbb{R}^{2n}$ so that in the new coordinates the coefficient matrix of Ω is identically J in the whole neighborhood. Our computational procedure follows the proof given by Weinstein(1971).

Assume that the linear change of variables has been made so that $\Omega(0) = J$ and scale by $x \to \epsilon x, \Omega \to \epsilon^{-1}\Omega$ so that

$$\Omega = \sum_{s=0}^{\infty} \left(\frac{\epsilon^s}{s!} \right) \omega_s^0, \tag{6.6}$$

where ω_s^0 is a closed 2-form with coefficients that are homogeneous polynomials in x of degree s. Let $\mathcal{P}_s$ be the vector space of such forms and $\mathcal{Q}_s = \{0\}$. Let $A \in \mathcal{P}_s$, $B = 0 \in \mathcal{Q}_s$, and $C \in \mathcal{R}_s$, where $\mathcal{R}_s$ is the vector space of vector fields which are homogeneous polynomials of degree $s+1$ condition. In coordinates equation (2.18) for this problem is

$$0 = A_{sm} + J_{im}\frac{\partial C^i}{\partial x^s} + J_{sj}\frac{\partial C^j}{\partial x^m}. \tag{6.7}$$

(In general there would be a term $+\frac{\partial J_{sm}}{\partial x^i}C^i$ in (6.7) but this term is zero since J is constant.)

Since A is a closed two form there is a one form α such that $A = d\alpha$ so (6.7) becomes

$$0 = \frac{\partial \alpha_s}{\partial x^m} - \frac{\partial \alpha_m}{\partial x^s} + J_{im}\frac{\partial C^i}{\partial x^s} + J_{sj}\frac{\partial C^j}{\partial x^m}. \tag{6.8}$$

This equation has a solution $C^i = \alpha_{i+n}$ for $1 \le i \le n, C^i = -\alpha_{i-n}$ for $n \le i \le 2n$, or $C = J\alpha$. Thus there is a solution of (2.18) and so the coordinate change given by Darboux's theorem can be computed by Lie transforms.

7. Hamiltonian Systems

For Hamiltonian systems the Lie bracket is replaced by the Poisson bracket. Let F, G and H be smooth real valued functions defined in an open set in $\mathbb{R}^{2n}$, the Poisson bracket of F and G is the smooth function $\{F, G\}$ defined by

$$\{F, G\} = \frac{\partial F}{\partial x}^T J \frac{\partial G}{\partial x} \tag{7.1}$$

where J is as in (6.5) the usual $2n \times 2n$ skew symmetric matrix of Hamiltonian mechanics. A Hamiltonian differential equation (generated by the Hamiltonian H) is

$$\dot{x} = J\frac{\partial H}{\partial x}. \tag{7.2}$$

The Poisson bracket and the Lie bracket are related by

$$J\frac{\partial}{\partial x}\{F, G\} = \left[J\frac{\partial F}{\partial x}, J\frac{\partial G}{\partial x} \right] \tag{7.3}$$

so the Hamiltonian vector field generated by $\{F, G\}$ is the Lie bracket of the Hamiltonian vector fields generated by G and F, see Abraham and Marsden(1978).

Consider a Hamiltonian perturbation problem given by the Hamiltonian

$$H_*(x, \epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!} \right) H_j^0(x). \tag{7.4}$$

A near identity symplectic change of coordinates $x = \phi(\xi, \epsilon) = \xi + \cdots$ can be generated as the solution of the Hamiltonian differential equations

$$\frac{dx}{d\epsilon} = J\frac{\partial W}{\partial x}(x,\epsilon), \quad x(0) = \xi, \quad W(x,\epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) W_{j+1}(x). \tag{7.5}$$

It transforms (7.4) to

$$H^*(x,\epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) H_0^j(x). \tag{7.6}$$

Theorem 7.1. *Let $\mathcal{P}_j$, $\mathcal{Q}_j$, and $\mathcal{R}_j$ be vector spaces of smooth Hamiltonians on D with $\mathcal{Q}_j \subset \mathcal{P}_j$. Assume that (i) $H_j^0 \in \mathcal{P}_j$ for $j = 1, 2, 3...$ (ii) $\{\mathcal{P}_i, \mathcal{R}_j\} \subset \mathcal{P}_{i+j}$ for $i, j = 1, 2, 3, ...$ (iii) for any j and any $A \in \mathcal{P}_j$ there exist $B \in \mathcal{Q}_j$ and $C \in \mathcal{R}_j$ such that*

$$B = A + \{H_0^0, C\}. \tag{7.7}$$

Then one can compute a formal expansion for W in (7.5) with $W_j \in \mathcal{R}_j$ for all j which transforms (7.4) to (7.6) where $H_0^j \in \mathcal{Q}_j$ for all j.

The classical Birkhoff normal form for a Hamiltonian system near an equilibrium point is as follows. Assume that the Hamiltonian (7.4) came from scaling a system about an equilibrium point at the origin. That is, $H_0^0(x)$ is a quadratic form and H_j^0 is a homogeneous polynomial of degree $j + 2$. Assume that the linear Hamiltonian system

$$\dot{x} = J\frac{\partial H_0^0}{\partial x} = Ax \tag{7.8}$$

is such that A is diagonalizable. Then one can compute a symplectic change of variables generated by (7.5) which transforms (7.4) to (7.6) with

$$H^*(e^{At}x, \epsilon) = H^*(x, \epsilon). \tag{7.9}$$

For a Lie transform proof see Meyer(1974).

Kummer(1976) has shown that Lie algebra theory is useful in studying normal forms in some special cases in celestial mechanics. Taking this lead Cushman, Deprit and Mosak(1983) have used results from representation theory to give a complete description of the normal forms for Hamiltonian systems without the diagonalizable assumption.

8. The General Lie Transform Algorithm

In this section we will give a proof of the main algorithm of Deprit, Theorem 8.1, and the main perturbation algorithm, Theorem 8.2, for general tensor fields. Theorem 8.2 is a slight extension of Theorem 2.1. A general reference for the tensor analysis and notation used here is Abraham and Marsden(1978).

Let $\mathbb{E}, \mathbb{F}, \mathbb{G}$ and $\mathbb{E}_1, \ldots, \mathbb{E}_k$ be vector spaces over $\mathbb{K}$ where $\mathbb{K}$ is the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$, $L(\mathbb{E}; \mathbb{F})$ be the space of bounded linear functions from $\mathbb{E}$ to $\mathbb{F}$, $\mathbb{E}^* = L(\mathbb{E}, \mathbb{K})$ be the dual space of $\mathbb{E}$, and $L^k(\mathbb{E}_1, \ldots, \mathbb{E}_k; \mathbb{K})$ be the space of bounded multilinear maps from $\mathbb{E}_1 \times \cdots \times \mathbb{E}_k$ into $\mathbb{K}$. Define $T_s^r(\mathbb{E}) = L^{r+s}(\mathbb{E}^*, \ldots, \mathbb{E}^*, \mathbb{E}, \ldots, \mathbb{E}; \mathbb{K}) - -r$ copies of $\mathbb{E}^*$ and s copies of $\mathbb{E}$, so if $Z \in T_s^r(\mathbb{E})$ then $Z : \mathbb{E}^* \times \ldots \times \mathbb{E}^* \times \mathbb{E} \times \ldots \times \mathbb{E} \to \mathbb{K}$ is linear in each argument. The elements, $Z \in T_s^r(\mathbb{E})$, are called r-contravariant, s-covariant tensors or simply

(r,s)-tensors. In the case $r=s=0$ we define $T_0^0(\mathbb{E})=\mathbb{K}$. If $A:\mathbb{E}\to\mathbb{E}$ is an invertible linear map and $A^*:\mathbb{E}^*\to\mathbb{E}^*$ is the dual map, then $A_s^r:T_s^r(\mathbb{E})\to T_s^r(\mathbb{E})$ is the invertible linear map defined by $(A_s^rZ)(\alpha^1,\dots,\alpha^r,\beta_1,\dots,\beta_s)=Z(A^*\alpha^1,\dots,A^*\alpha^r,A^{-1}\beta_1,\dots,A^{-1}\beta_s)$.

Let M be a smooth manifold modeled on a vector space $\mathbb{E}$ and $p\in M$ any point. In the classical and still most important case M is simply an open set D in $\mathbb{R}^m$ and $\mathbb{E}$ is $\mathbb{R}^m$ itself. The tangent space to M at p, denoted by T_pM is isomorphic to $\mathbb{E}$ itself; the cotangent space to M at p, denoted by T_p^*M, is the dual of T_pM; and the space of r-contravariant, s-covariant tensors at p is $T_s^r(T_pM)$. The vector bundles built on T_pM, T_p^*M, and $T_s^r(T_pM)$ are respectively: TM, the tangent bundle; T^*M, the cotangent bundle; and T_s^rM, the (r,s)-tensor bundle. Smooth sections in these bundles are called respectively: vector fields (or contravariant vector fields or ordinary differential equations); covector fields (or one forms); and (r,s)-tensor fields. Let $\mathcal{T}(M)$ be the space of smooth vector fields, $\mathcal{T}^*(M)$ the space of smooth one-forms, and $\mathcal{T}_s^r(M)$ the space of smooth (r,s)-tensors. Let $V:M\to M$ be a diffeomorphism, $p\in M, q=V(p)$ and $DV(p):T_pM\to T_qM$ be the derivative of V at p then $DV_s^r(p):T_s^r(T_pM)\to T_s^r(T_qM)$. The results of this section are quite general so M could be a Banach manifold modeled on a reflexive Banach space $\mathbb{E}$, but the author has no examples which require this level of generality.

Consider the case where M is an open set in $\mathbb{R}^m$ with coordinates $(x^1,\dots,x^m)$. A $(0,0)$-tensor field is simply a smooth function $Z:M\to\mathbb{K}$. A vector field, Z, is given by

$$Z=Z^1(x)\frac{\partial}{\partial x^1}+\cdots+Z^m(x)\frac{\partial}{\partial x^m}, \tag{8.1}$$

where $Z^1,\dots,Z^m$ are smooth real valued functions on M. The vector field Z is the same as the differential equation

$$\dot{x}=Z(x)\quad(\text{or}\quad \dot{x}^i=Z^i(x),\quad i=1,\dots,m). \tag{8.2}$$

A covector field, Z, is given by

$$Z=Z_1(x)dx^1+\cdots+Z_m(x)dx^m, \tag{8.3}$$

where again $Z_1,\dots,Z_m$ are smooth functions.

Let U be a smooth vector field (autonomous differential equation) on M and let $X(\tau,y)$be the general solution of the differential equation

$$x'=\frac{dx}{d\tau}=U(x) \tag{8.4}$$

which satisfies $X(0,y)=y$. That is, $X'(\tau,y)=U(X(\tau,y))$. Assume that there is an $\tau_0>0$ such that $X:(-\tau_0,\tau_0)\times M\to M$ is defined and smooth. X is a function of two arguments and let $'$ denote the partial derivative with respect to the first argument, $'=\partial/\partial\tau$, and let D denote the partial derivative with respect to the second argument, $D=\partial/\partial y$, thus $DX(\tau,p):T_pM\to T_qM, q=X(\tau,p)$ and $DX_s^r(\tau,p):T_s^r(T_pM)\to T_s^r(T_qM)$. Let $Z:M\to\mathcal{T}_s^r(M)$ be a smooth (r,s)-tensor field on $M, p\in M, q=X(\tau,p)$. Then $Z(p)\in T_s^r(T_pM), Z(X(\tau,p))\in T_s^r(T_qM)$, and $A(\tau)=DX_s^r(\tau,p)^{-1}Z(X(\tau,p))\in T_s^r(T_pM)$, so $A(\tau)$ is a smooth curve of (r,s)-tensors in the fixed tensor space $T_s^r(T_pM)$. The Lie derivative of Z in the direction of U (or along U) is denoted by $[Z,U]$ and is defined as

$$[Z,U](p)=\frac{\partial}{\partial\tau}A(\tau)\bigg|_{\tau=0}=\frac{\partial}{\partial\tau}DX_s^r(\tau,p)^{-1}Z(X(\tau,p))\bigg|_{\tau=0} \tag{8.5}$$

Since $A(\tau) \in T_s^r(T_pM)$ for all τ its derivative is in $T_s^r(T_pM)$ so $[Z,U](p) \in T_s^r(T_pM)$ and $[Z,U]$ is a smooth (r,s)-tensor field also and $[\cdot,\cdot] : \mathcal{T}_s^s(M) \times \mathcal{T}(M) \to \mathcal{T}_s^r(M)$ is bilinear. $[\cdot,\cdot]$ is called the Lie bracket.

If M is an open set in $\mathbb{R}^m$ and $Z : M \to \mathbb{R}$ is a smooth function ((0,0)-tensor field) then in classical notation

$$[Z,U](x) = \nabla Z(x) \cdot U(x) \tag{8.6}$$

so $[Z,U]$ is the directional derivative of Z in the direction U. If Z is a smooth vector field (ordinary differential equation) as in (8.2) then

$$[Z,U](x) = \frac{\partial Z}{\partial x}(x)U(x) - \frac{\partial U}{\partial x}(x)Z(x) \tag{8.7}$$

where Z and U are column vectors. If Z is a one form thought of as a column vector then

$$[Z,U](x) = \frac{\partial Z}{\partial x}(x)^T U(x) + \frac{\partial U}{\partial x}(x)^T Z(x). \tag{8.8}$$

Suppose that the perturbation problem is given as an (r,s)-tensor field $Z = Z_*$ on M which has a formal expansion in a small parameter ϵ. Consider

$$Z(\epsilon,x) = Z_*(\epsilon,x) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) Z_j^0(x) \tag{8.9}$$

where each $Z_j^0 : M \to T_s^r M$ is an (r,s)-tensor field.

To simplify the perturbation problem given by Z_* in (8.9) we seek a near identity change of coordinates of the form

$$x = X(\epsilon,y) = y + \cdots \tag{8.10}$$

where $X(\epsilon,y)$ is constructed as a formal solution of the nonautonomous system of differential equations

$$\frac{dx}{d\epsilon} = W(x,\epsilon) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) W_{j+1}(x), \tag{8.11}$$

satisfying the initial condition

$$x(0) = y \tag{8.12}$$

where each $W_j : M \to TM$ is a smooth vector field.

The Lie transform of $Z(= Z_*)$ by W, denoted by $\mathcal{L}(W)Z$ or Z^* for short, is the tensor field Z_* expressed in the new coordinates and so is an (r,s)-tensor field depending on the parameter ϵ also. Specifically,

$$Z^*(\epsilon,y) = \mathcal{L}(W)Z(\epsilon,y) = DX_s^r(\epsilon,y)^{-1} Z_*(\epsilon,X(\epsilon,y)) \tag{8.13}$$

In the new coordinates y the tensor $Z_*(x,\epsilon)$ becomes

$$Z^*(y,\epsilon) = \mathcal{L}(W)Z(\epsilon,y) = \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) Z_0^j(y). \tag{8.14}$$

We say (8.10) or (8.11) transforms (8.9) into (8.14). The method of Lie transforms introduces a double indexed array of tensor fields $\{Z_j^i\}, i,j = 0,1,\ldots$ which agree with the definitions

given in (8.9) and (8.14) when either i or j is zero. The other terms are intermediary terms introduced to facilitate the computation. The main theorem on Lie transforms by Deprit(1969) in this general context is the following.

Theorem 8.1. *Using the notation given above, the tensor fields* $\{Z_j^i\}, i = 1, 2, \ldots, j = 0, 1, \ldots$ *satisfy the recursive identities*

$$Z_j^i = Z_{j+1}^{i-1} + \sum_{k=0}^{j} \binom{j}{k} [Z_{j-k}^{i-1}, W_{k+1}]. \tag{8.15}$$

Remarks. The above formulas contain the binomial coefficient $\binom{j}{k} = \dfrac{j!}{k!(j-k)!}$. Note that since the transformation generated by W is a near identity transformation the first term in Z_* and Z^* are the same, namely Z_0^0. Also note that the first term in the expansion for W starts with W_1. This convention imparts some nice properties to the formulas in (8.15). Namely, each term in (8.15) has indices summing to $i + j$ and each term on the right hand side has upper index i-1.

The interdependence of the $\{Z_j^i\}$ can easily be understood by considering the Lie triangle

$$\begin{array}{ccccccc} Z_0^0 & & & & & & \\ \downarrow & & & & & & \\ Z_1^0 & \to & Z_0^1 & & & & \\ \downarrow & & \downarrow & & & & \\ Z_2^0 & \to & Z_1^1 & \to & Z_0^2 & & \\ \downarrow & & \downarrow & & \downarrow & & \\ Z_3^0 & \to & Z_2^1 & \to & Z_1^2 & \to & Z_0^3 \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \end{array} \tag{8.16}$$

The coefficients of the expansion of the old tensor field Z_* are in the left column and those of the new tensor field Z^* are on the diagonal. The formula (8.15) says that to calculate any element in the Lie triangle you need the entries in the column one step to the left and up.

Proof. Let $Y(\epsilon, x)$ be the inverse of $X(\epsilon, y)$ so $Y(\epsilon, X(\epsilon, y)) \equiv y$, $X(\epsilon, Y(\epsilon, x)) \equiv x$, $DX(\epsilon, y)^{-1} = DY(\epsilon, X(\epsilon, y))$, and $DX_s^r(\epsilon, y)^{-1} = DY_s^r(\epsilon, X(\epsilon, y))$. Thus (8.13) becomes $\mathcal{L}(W)Z(\epsilon, y) = DY_s^r(\epsilon, X(\epsilon, y))Z_*(\epsilon, X(\epsilon, y))$.

Define the differential operator $\mathcal{D} = \mathcal{D}_W$ acting on (r, s)-tensor fields depending on a parameter ϵ by

$$\mathcal{D}K(\epsilon, x) = \frac{\partial K}{\partial \epsilon}(\epsilon, x) + [K, W](\epsilon, x). \tag{8.17}$$

In computing the Lie bracket in (8.17) the ϵ is simply a parameter and so held fixed during any differentiation. With this notation we have

$$\frac{d}{d\epsilon} \left\{ DY_s^r(\epsilon, x)K(\epsilon, x)\Big|_{x=X(\epsilon,y)} \right\} = DY_s^r(\epsilon, x)\mathcal{D}K(\epsilon, x)\Big|_{x=X(\epsilon,y)}. \tag{8.18}$$

Define new functions by $Z^0 = Z$, $Z^i = \mathcal{D}Z^{i-1}, i \geq 1$. Let these functions have series expansions

$$Z^i(\epsilon, x) = \sum_{k=0}^{\infty} \left(\frac{\epsilon^k}{k!}\right) Z_k^i(x) \tag{8.19}$$

so

$$\begin{aligned} Z^i(\epsilon, x) &= \mathcal{D}\sum_{k=0}^{\infty} \left(\frac{\epsilon^k}{k!}\right) Z_k^{i-1}(x) \\ &= \sum_{k=0}^{\infty} \left(\frac{\epsilon^{k-1}}{(k-1)!}\right) Z_k^{i-1} + \sum_{k=0}^{\infty} \left[\left(\frac{\epsilon^k}{k!}\right) Z_k^{i-1}(x), \sum_{s=0}^{\infty} \left(\frac{\epsilon^s}{s!}\right) W_{s+1}(x)\right] \\ &= \sum_{j=0}^{\infty} \left(\frac{\epsilon^j}{j!}\right) \left(Z_{j+1}^{i-1} + \sum_{k=0}^{i} \binom{j}{k} \left[Z_{j-k}^{i-1}, W_{k+1}\right]\right). \end{aligned} \tag{8.20}$$

So the functions Z_j^i are related by (8.15). It remains to show that $Z_* = G$ has the expansion (8.14). By Taylor's theorem and (8.18)

$$\begin{aligned} G(\epsilon, y) &= \sum_{n=0}^{\infty} \left(\frac{\epsilon^n}{n!}\right) \frac{d^n}{d\epsilon^n} G(\epsilon, y)\Big|_{\epsilon=0} \\ &= \sum_{n=0}^{\infty} \left(\frac{\epsilon^n}{n!}\right) \frac{d^n}{d\epsilon^n} \left(DY_s^r(\epsilon, x) Z(\epsilon, x)\Big|_{x=X(\epsilon,y)}\right)_{\epsilon=0} \\ &= \sum_{n=0}^{\infty} \left(\frac{\epsilon^n}{n!}\right) \left(DY_s^r(\epsilon, x) \mathcal{D}^n Z(\epsilon, x)\Big|_{x=X(\epsilon,y)}\right)_{\epsilon=0} \\ &= \sum_{n=0}^{\infty} \left(\frac{\epsilon^n}{n!}\right) Z_0^n(x) \end{aligned} \tag{8.21}$$

□

In the cases of interest the tensor field is given and the change of variables is sought to simplify it. When the field is sufficiently simple it is said to be in 'normal form'. The main Lie transform algorithm starts with a given field which depends on a small parameter, ϵ, and constructs a change of variables so that the field in the new variables is simple. The algorithm is built around the following observation.

Consider the series (8.9) as given so all the Z_i^0 are known. Assume that all the entries in the Lie triangle are known down to the N row, so the Z_j^i are known for $i+j \leq N$ and assume the W_i are known for $i \leq N$. Let $\tilde{Z}_j^i$ be computed from the same differential equation, so $\tilde{Z}_i^0 = Z_i^0$ for all i, and with $\tilde{W}_1, \ldots, \tilde{W}_N$ where $\tilde{W}_i = W_i$ for $i = 1, 2, \ldots, N-1$ but $\tilde{W}_N = 0$.

Then

$$Z_j^i = \tilde{Z}_j^i \quad \text{for} \quad i+j < N$$

$$Z_j^i = \tilde{Z}_j^i + [Z_0^0, W_N] \quad \text{for} \quad i+j = n \tag{8.22}$$

This is easily seen from the recursive formulas in Theorem 8.1. Recall the remark that the sum of all the indices must add to the row number, so W_N does not effect the terms in the first $N-1$ rows. The second equation in (8.22) follows from a simple induction across the N^{th} row. The algorithm can be used to prove a general theorem which includes almost all applications, see Meyer and Schmidt(1977).

Theorem 8.2. *Let $\{P_i\}_{i=0}^{\infty}$, $\{Q_i\}_{i=1}^{\infty}$ and $\{R_i\}_{i=1}^{\infty}$ be sequences of linear spaces of smooth fields defined on a manifold M where $\{P_i\}_{i=0}^{\infty}$ and $\{Q_i\}_{i=1}^{\infty}$ are (r,s)-tensor fields and $\{R_i\}_{i=1}^{\infty}$ are a vector fields. Assume:*

(i) $Q_i \subset P_i$, $i = 1, 2, \ldots$
(ii) $Z_i^0 \in P_i$, $i = 0, 1, 2, \ldots$
(iii) $[P_i, R_j] \subset P_{i+j}$, $i, j = 0, 1, 2, \ldots$
(iv) *for any $A \in P_i$, $i = 1, 2, \ldots$ there exists $B \in Q_i$ and $C \in R_i$ such that*

$$B = A + [Z_0^0, C]. \tag{8.23}$$

Then there exists a W with a formal expansion of the form (8.11) with $W_i \in R_i$, $i = 1, 2, \ldots$, which transforms the tensor field Z_* with the formal series expansion given in (8.9) to the field Z^* with the formal series expansion given by (8.14) with $Z_0^i \in Q_i, i = 1, 2, \ldots$.

Proof. Use induction on the rows of the Lie triangle. Induction Hypothesis I_n: Let $Z_j^i \in P_{i+j}$ for $0 \le i+j \le n$ and $W_i \in R_i$, $Z_0^i \in Q_i$ for $1 \le i \le n$.

I_0 is true by assumption and so assume I_{n-1}. By (8.15)

$$Z_{n-1}^1 = Z_n^0 + \sum_{k=0}^{n-2} \binom{n-1}{k} [W_{k+1}, Z_{n-1-k}^0] + [W_n, Z_0^0]. \tag{8.24}$$

The last term is singled out because it is the only term that contains the element, W_n, which is not covered either by the induction hypothesis or the hypothesis of the theorem. All the other terms are in P_n by I_{n-1} and (iii). Thus

$$Z_{n-1}^1 = K^1 + [W_n, Z_0^0] \tag{8.25}$$

where $K^1 \in P_n$ is known. A simple induction on the columns of the Lie triangle using (8.15) shows that

$$Z_{n-s}^s = K^s + [W_n, Z_0^0] \tag{8.26}$$

where $K^s \in P_n$ for $s = 1, 2, \ldots, n$ and so

$$Z_n^0 = K^n + [W_n, Z_0^0]. \tag{8.27}$$

By (iv) solve (8.27) for $W_n \in R_n$ and $Z_0^n \in Q_i$. Thus I_n is true. □

The theorem given above is formal in the sense that the convergence of the various series is not discussed. In interesting cases the series diverge, but useful information can be obtained in the first few terms of the normal form. One can stop the process at any order, N, to

obtain a W which is a polynomial in ϵ and so converges. From the proof given above it is clear that the terms in series for Z^* up to order N are unaffected by the termination.

References

Abraham, R. and Marsden J. E. 1978: **Foundations of Mechanics,** Benjamin/Cummings Publ. Co., Reading, Mass.

Birkhoff, G. D. 1927: **Dynamical Systems**, Am. Math. Soc., Providence, R.I.

Cushman, R., Deprit, A. and Mosak, R. 1983: *Normal forms and representation theory*, **J. Math. Phy.** 24(8), 2102-2116.

Cushman, R., and Sanders, J. 1986: *Nilpotent normal forms and representation theory of* $sl(2,\mathbb{R})$, **Contempory Mathematics 56,** Amer. Math. Soc., Providence, R.I., 31-51.

Deprit, A. 1969: *Canonical transformation depending on a small parameter*, **Celestial Mechanics 72**, 173-79.

Diliberto, S. P. 1961: *Perturbation theorems for periodic systems*, **Circ. Mat. Palermo 9(2)**, 265-299, **10(2)**, 111-112.

Diliberto, S. P. 1967: *New results on periodic surfaces and the averaging principle*, **Differential and Integral Equations**, Benjamin, New York, 49-87.

Flanders, H. 1963: **Differential Forms**, Academic Press, New York.

Gantmacher, F. R. 1960: **The Theory of Matrices**, Chelsea Publ., New York.

Henrard, J. 1970: *On a perturbation theory using Lie transforms*, **Celestial Mech. 3**, 107-120.

Hori, G. 1966: *Theory of general perturbations with unspecified canonical variables,* **Publ. Astron. Soc. Japan 18(4)**, 287-296.

Kamel, A. 1970: *Perturbation method in the theory of nonlinear oscillations,* **Celestial Mechanics 3**, 90-99.

Kobayashi, S. and Nomizu, K. 1969: **Foundations of Differential Geometry**, Interscience, New York.

Kummer, M. 1976: *On resonant non-linear coupled oscillators with two equal frequencies*, **Comm. Math. Phy. 48**, 137-139.

Meyer, K. R. 1974: *Normal forms for Hamiltonian systems*, **Celestial Mech. 9**, 517-522.

Meyer, K. R. 1984: *Normal forms for the general equilibrium*, **Funkcialaj Ekv. 27(2)**, 261-271.

Meyer, K. R. 1989: *A Lie transform tutorial*, it never appeared in the Proceedings of the Conference on Symbolic Computations at IBM, Yorktown Heights.

Meyer, K. R. and Schmidt, D. S. 1977: *Entrainment domains*, **Funkcialaj Ekvacioj 20(2)**, 171-92.

Meyer, K. R. and Schmidt, D. S. 1988: *Bifurcations of relative equilibria in the N-body and Kirchhoff problems*, **SIAM Journal on Mathematical Analysis**, 19, 6, 1295-1313.

Poston, T. and Stewart, I. 1978: **Catastrophe Theory and its Applications**, Pitman, Boston.

Weinstein, A. 1971: *Symplectic manifolds and their Lagrangian submanifolds*, **Adv. Math. 6**, 329-346.

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

E-mail address: ken.meyer@uc.edu