Commun. Math. Phys. 128, 565–592 (1990)

Communications in Mathematical Physics

Isoholonomic Problems and Some Applications

R. Montgomery
M.S.R.I., 1000 Centennial Dr., Berkeley, CA 94720, USA

Abstract. We study the problem of finding the shortest loops with a given holonomy. We show that the solutions are the trajectories of particles in Yang–Mills potentials (Theorem 4), or, equivalently, the projections of Kaluza–Klein geodesics (Theorem 2). Applications to quantum mechanics (Berry's phase, Sect. 3) and the optimal control of deformable bodies (Sect. 6) are touched upon.

Contents

1. The Problem and an Introduction

1.1 The Problem which we investigate is *the isoholonomic problem*: **among all loops with a fixed holonomy, find the loop of minimum length.**

The data needed to formulate this problem are a principal bundle

$$\pi: Q \to X \qquad [1.1]$$

with connection A, a Riemannian metric k on X, and a point $x_0 \in X$ at which the loop and its holonomy are based. (The holonomy is called the Wilson loop integral, or the path-ordered exponential of $-A$ in the physics literature.) The structure group of the bundle will be denoted by G. It is a Lie group which acts on Q on the right, and such that $X \cong Q/G$.

1.2. Motivation

1. The physical chemist Alex Pines posed this problem in an effort to better understand and design nuclear magnetic resonance experiments for measuring the non-Abelian Berry's phase. Berry's phase is an element of the unitary group, $G = U(k)$, which is associated to a closed curve of quantum mechanical states. It is the holonomy of the loop of states with respect to a canonical connection. The Abelian ($k = 1$) Berry's phase is the case in which the states are pure states, and has been measured in numerous experiments (Tomita and Chiao [1986], Tycko [1987], Suter, Mueller, and Pines [1988]). The non-Abelian Berry's phase occurs for mixed states, and is related to weak measurements. It has not yet been experimentally measured. See Sect. 3 for details.

There is a booming literature on Berry's phase. Some salient papers are Simon [1983], Berry [1984], Wilczek and Zee [1984], and Aharonov and Anandan [1987]. Our point of view is closest to this last paper.

The length of the loop of states is essentially the energy input required to make the loop. This is shown in Sect. 3. The isoholonomic problem is then the problem of generating a desired phase shift with a minimum amount of energy.

2. Another motivation for studying the isoholonomic problem is that, in certain circumstances, it is equivalent to

The Cat's Problem: **Find the most efficient way to deform a deformable body so as to achieve a desired re-orientation.**

A cat, dropped from upside-down with no angular momentum, changes her shape in such a way as to land on her feet. In doing so, her initial and final shape are essentially the same, but she has re-oriented herself by a rigid rotation of 180 degrees. See Fig. 1. In addition, by conservation, her total angular momentum is zero throughout the motion. For a nice mechanical analysis of this phenomenon, see Kane and Scher [1969]. The cat thus describes a loop in her shape space with the consequence that, in an inertial frame, the beginning and final shapes are related by a rigid motion $g \in G = E(3)$.

Shapere and Wilczek addressed a version of the cat's problem in [1987, 1988]. See also Shapere [1989], and Wilczek [1988]. Their key observation is that certain dynamical constraints, such as "angular momentum equals zero," define a connection on the principal bundle $Q = (\text{inertial configurations}) \to X = (\text{shape space}) = Q/G$. The fiber of this bundle is the group G of rigid motions, an element of which is the cat's desired re-orientation. See Fig. 2.

Iwai [1987a, b, c] also made the observation that angular momentum defines a connection. He noted that the parallel translation for this connection defines the Guichardet frame, which plays an important role in molecular dynamics.

The other key ingredient in Shapere and Wilczek's work is their definition of efficiency in terms of a matric k on shape space. If we *define* the efficiency of a path to be its length (or integrated kinetic energy) then it becomes clear that the cat's problem is the isoholonomic problem.

In Shapere and Wilczek's version of the cat's problem, they do not restrict the

Fig. 1

holonomy, but rather define efficiency as a quotient of (some function of) the holonomy by length. We discuss their problem in Sect. 6.

We discuss the cat's problem in more detail in Sect. 4. Montgomery [1989] is devoted to the problem.

1.3. Perspective. The isoholonomic problem is a generalization of the isoperimetric problem. Take X to be a Riemann surface. Take Q to be a circle bundle over X with a connection whose curvature form is a constant non-zero multiple of the area form on X. Fixing the holonomy of a loop in X is equivalent to fixing the area it encloses and so the isoholonomic problem becomes the classical isoperimetric problem. If X has constant Gaussian curvature (or if we instead took the connection to be the Levi–Civita connection) then the solutions are curves of constant geodesic curvature. For example if X is the sphere or the plane, these curves are geometric circles.

The isoholonomic problem is a special case of the problem of finding sub-Riemannian geodesics. A *sub-Riemannian metric* (Strichartz, [1983]) consists of a distribution Hor on Q, that is a subbundle $\text{Hor} \subset TQ \to Q$, together with a positive definite fiber metric κ_q on Hor. For example, κ could be the restriction

of a Riemannian metric on Q to Hor. Sub-Riemannian metrics are also known as *non-holonomic Riemannian metrics* (Vershik and Gershkovich [1988] *Carnot–Caratheodory metrics* (Hamenstädt [1986, 1988], Bär [1989]), or *singular Riemannian metrics* (Hermann [1973], Brockett [1981]).

Call a curve c *horizontal* if it is piecewise differentiable and its derivative $\dot{c}$, when it exists, lies in Hor. The sub-Riemannian distance between points $p, q \in Q$ is

$$d(p, q) = \inf\{\text{length}(c)\colon c \text{ a horizontal curve joining } p \text{ to } q\}.$$

Here length(c) is the integral of $\sqrt{\kappa(\dot{c}, \dot{c})}dt$ over the curve. (If there are no horizontal paths joining p to q, set $d(p, q) = \infty$.) By taking

$$\text{Hor} = \ker(A), \tag{1.2a}$$

the horizontal distribution for our connection A, and

$$\kappa_q(v, w) = k_{\pi(q)}(d_q\pi \cdot v, d_q\pi \cdot w), \tag{1.2b}$$

we see that the isoholonomic problem becomes a special case of the *Sub-Riemannian geodesic problem.* Find the horizontal curve joining p to q whose length is $d(p, q)$.

The o.d.e. (see Theorem 5 below) which *ought* to characterize sub-Riemannian geodesics has been known for decades. Bär [1988, 1989], following a partial proof of Strichartz [1983], proved that this o.d.e. does in fact characterize them. We restate Bär's theorem here as Theorem 5 in Sect. 5. The 'hard" half of our main result, Theorem 1, is an immediate consequence of Bär's theorem.

1.4. Results and Outline. Our key result, Theorem 1 below, states that solving the isoholonomic problem is equivalent to solving the Hamiltonian differential equations (with the correct endpoint conditions) generated by a certain Hamiltonian H_0. More precisely, first relax the condition that the curve in X be a loop. (This eliminates worry over the endpoint conditions.) consider the

The Isoparallel Problem. **Among all piecewise $\mathbf{C}^1$ curves** c **in** X **joining** x_0 **to** x_1 **with a fixed** *parallel translation operator*

$$\text{Hol}[c]\colon Q_0 \to Q_1,$$

find the loop of minimum length. Here Q_i **is the fiber** $\pi^{-1}(x_i)$ **over** $\mathbf{x}_i, i = 1, 2$.

(Recall that $\text{Hol}[c](q_0) = q_1$, where $q(t)$ is the unique horizontal path covering (that is, $\pi \circ q =$) x and satisfying $q(0) = q_0$. We assume here that x is parametrized by $0 \leqq t \leqq 1$.) In case $x_0 = x_1$ this is the isoholonomic problem. Theorem 1 states that the *extremals* for the isoparallel problem are exactly the projections of the solutions to the Hamiltonian equations. The rest of our results follow directly from Theorem 1 and our earlier results [1984] concerning the equations of a particle in a Yang–Mills fields.

Theorem 2 states that the isoparallel extremals are the projections to X of geodesics for a Kaluza–Klein metric on Q. To define this metric we must have an adjoint invariant inner product on the Lie algebra of G.

Theorem 3.1 of Sect. 3 characterizes the isoparallel extremals for the data (bundle, connection) of Pine's interest. Theorem 3.1 follows immediately from Theorem 3.2 which describes the isoparallel extremals when the metric and connection are homogeneous.

Theorem 4 states that these extremals are the trajectories of a "particle" travelling in the Yang–Mills potential A. The differential equations of such a particle are called Wong's equations (Eqs. [4.2a–c] below), after Wong [1970].

Following the statements of Theorems 2, 3.1 and 4 we present some examples.

Theorem 5 is Bär's theorem, which we restate in order to prove half of Theorem 1.

Theorem 6 is a rephrasing of Theorem 1 in the context of the cat's problem.

Section 7 concerns Wilczek and Shapere's problem of maximizing the efficiency of a loop. Theorem 7 states that the solutions to this problem are isoparallel extremals, and hence projections of solutions to Wong's equations.

1.5. Solvability and Controllability. There may be no loops whose holonomy is $h_0 \in \mathrm{Aut}(Q_0)$. In this case the isoholonomic problem (for this particular holonomy constraint) has no solution.

The Ambrose–Singer theorem (Ambrose and Singer [1953], see also Kobayashi–Nomizu [1963], pp. 83–89) gives a sufficient condition for every holonomy to be realized. This theorem is a restatement of a theorem of Chow [1939], now familiar to people in control theory. In control theory a distribution with the property that any two points can be joined by a horizontal path is said to be locally *controllable*, or to provide local *accessibility*.

The horizontal distribution is said to satisfy "Hormander's condition" at $q \in Q$ if the horizontal vector fields, together with all of their iterated Lie brackets span the tangent space at q. The Chow–Ambrose–Singer theorem implies that if Hor satisfies Hormander's condition at some $q \in Q$, then any two nearby points in Q can be joined by a horizontal path, and hence any holonomy near the identity is realized. (The distribution must come from a connection for this implication to hold. One uses its G-invariance in the proof.) The Hormander condition can be expressed purely in terms of the curvature $F = dA + [A, A]$ and its covariant derivatives. This gives the following consequence of the Ambrose–Singer theorem.

Proposition 1. *Suppose X and G are connected, and that the Riemannian structure k on X is complete. Let $\mathfrak{g}$ denote the Lie algebra of G, and $\Delta(q)$ the Lie subalgebra of $\mathfrak{g}$ which is generated by the values of the curvature $F_q(X, Y)$ together with all of its covariant derivatives $D_Z F(X, Y), D_W D_Z F(X, Y)$, etc., at q. If there is a point $q \in Q$ such that $\Delta(q) = \mathfrak{g}$, then the isoparallel problem is solvable for every choice of the parallel transport constraint.*

Proof. According to the Ambrose–Singer theorem there exists a sequence c_i of loops with the given holonomy, whose lengths approach the infimum of the lengths of all loops with this holonomy. Apply the Arzela–Ascoli theorem to get a convergent subsequence.

1.6. Earlier Papers. This paper is an extension of two earlier preprints, Montgomery [1988], and Montgomery [1989].

2. Two Theorems and Kaluza–Klein Metrics

2.1. We begin with some definitions. Call a family $c_\varepsilon, 0 < \varepsilon < 1$, of piecewise C^1 curves on X an *isoparallel deformation of* c if as $\varepsilon \to 0$ it converges uniformly (i.e. the C^0 topology) to c, and if every member of the family has the same end points and the same parallel translation operator as c. We say that the piecewise C^1 curve c is an *extremal* for the isoparallel problem if for every isoparallel deformation c_ε of c, we have

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \text{length}(c_\varepsilon) = 0 \qquad [2.1]$$

whenever the derivative exists.

Define the *horizontal kinetic energy*

$$H_0 : T^*Q \to \mathbf{R} \qquad [2.2a]$$

by

$$H_0(q,p) = \tfrac{1}{2}\| h_q^* p \|^2, \quad p \in T_q^*Q. \qquad [2.2b]$$

Here $\|\cdot\|^2$ represents the squared length of a covector with respect to the metric on the base space X, and

$$h_q^* : T_q^*Q \to T_{\pi(q)}^*X \qquad [2.3a]$$

is the dual of horizontal lift operator h. The horizontal lift operator

$$h : \pi^*TX \to TQ; \quad h_q : T_{\pi(q)}X \to T_qQ \text{ (linear)} \qquad [2.3b]$$

is a vector bundle map which can be defined by requiring that

$$\text{image}(h_q) = \text{Hor}_q; \quad h_q \circ d_q\pi = \text{identity on } T_{\pi(q)}X. \qquad [2.3c]$$

Here $\text{Hor}_q = \ker(A_q)$ is the horizontal space defined by the connection, and π^*TX denotes the pullback by π of the vector bundle $TX \to X$. See Sect. 4.2 for a coordinate expression for H_0.

Theorem 1. *The loop c in X is an extremal for the isoparallel problem if and only if there is a curve $z = (q,p)$ in T^*Q which satisfies $\pi \circ q = c$, and which is a solution curve to Hamilton's differential equation for the Hamiltonian H_0.*

The curve q in Q is the cotangent projection of z.

2.2 Theorem 2 will be a reformulation of Theorem 1 which is applicable whenever the Lie algebra $\mathfrak{g}$ admits an adjoint invariant inner product β. We will use β to define a Kaluza–Klein type metric $d^2s = \beta \oplus k$ on Q. Let $\text{Vert} = \ker(d\pi)$ denote the vertical subbundle of TQ. The connection defines a splitting

$$TQ = \text{Vert} \oplus \text{Hor}. \qquad [2.4a]$$

Also,

$$\text{Vert} \oplus \text{Hor} \cong \mathfrak{g} \oplus \pi^* TX. \qquad [2.4b]$$

Here $\mathfrak{g}$ stands for the trivial bundle $Q \times \mathfrak{g} \to Q$. The isomorphism $\mathfrak{g} \to \text{Vert}$ is the infinitesimal G action. The isomorphism $\text{Hor} \to \pi^* TX$ is $d\pi$ restricted to Hor. Define the inner product $\beta \oplus k$ on $T_q Q$ by declaring that

$$\text{Vert}_q \perp \text{Hor}_q \text{ with respect to } \beta \oplus k,$$

and

$$\beta \oplus k = \beta \quad \text{on} \quad \mathfrak{g} \cong \text{Vert}_q,$$
$$\beta \oplus k = k \quad \text{on} \quad T_{\pi(q)} X \cong \text{Hor}_q.$$

Theorem 2. *Suppose* $\mathfrak{g}$ *admits an adjoint invariant inner product* β, *and use this to put the metric* $\beta \oplus k$ *on* Q. *Then the following conditions for a curve* c *in* X *are equivalent.*

A. *The curve* c *is an extremal for the isoparallel problem.*
B. *There is a geodesic* $\tilde{q} \subset Q$ *which satisfies* $\pi \circ \tilde{q} = c$.

2.3. Riemannian Submersions and Examples. The construction of $\beta \oplus k$ is often turned around. A G-invariant metric on Q defines a connection on $Q \to X$ and metric k on X, by *declaring* that $\text{Hor} = \text{Vert}^{\perp}$, and that π is a *Riemannian submersion.* The connection and base metric for the Hopf fibrations $S^{2n-1} \to \mathbf{CP}^n$ are induced by the standard metric on S^{2n-1} in this manner. We recall that a *Riemannian submersion* $\pi: Q \to X$ of Riemannian manifolds is a submersion with the property that $d_q \pi$ is an isometry, when restricted to $\text{Hor} = (\ker d_q \pi)^{\perp}$. The metric on Q also induces a family of right invariant metrics on G. If these are all isometric to a fixed bi-invariant metric on G, then the original metric on Q is of the form $\beta \oplus k$.

Example A. Consider the Hopf fibration $S^3 \to S^2 = \mathbf{CP}^1$, with its canonical connection and the standard metric on the base. These structures are induced by the standard metric on S^3, as just described. The geodesics on S^3 are great circles. One easily checks, for example by using coordinate formulas, that great circles in S^3 project to small circles (lines of latitude, or curves of constant geodesic curvature) on S^2. Hence these small circles are the isoparallel extremals.

Note that such c's are exactly the solutions to the isoperimetric problem on S^2, as they should be according to Sect. 1.3, "perspectives."

Example B. The same reasoning applies to the Hopf fibrations $S^{2n+1} \to \mathbf{CP}^n$. Each isoparallel extremal is a small circle which lies on some $\mathbf{CP}^1$ in $\mathbf{CP}^n$.

This reasoning also applies to the quaternionic Hopf fibration $S^7 \to \mathbf{HP}^1 = S^4$, a bundle with structure group $G = SU(2)$. The connection is induced by the standard metric on S^7 and is the standard Yang–Mills potential of a symmetric instanton. The base metric is the round one. The extremals are projected great circles, which are again small circles on S^4.

Experiments. Avron *et al* [1989] showed that this instanton connection occurs for families of time reversal invariant spin 3/2 systems. Koenig, Mueller, and

Zwanziger [1989] of Pines' group are currently designing an NMR type experiment based on such a family, namely axially symmetric crystals of potassium or sodium chlorate. The purpose of the experiment is to detect non-Abelian effects of this $SU(2)$ holonomy.

2.4. *Relations between Theorems 1 and 2*

1. The geodesics $\tilde{q}(t)$ of Theorem 2 will generally *not* be horizontal, whereas the curves $q(t)$ of Theorem 1 are always horizontal. In order to obtain $q(t)$ from $\tilde{q}(t)$, project $\tilde{q}(t)$ to X, and then horizontally lift this projection to form the horizontal curve $q(t)$ through $\tilde{q}(0)$. There is a formula for this operation:

$$q(t) = \tilde{q}(t) \exp\{-t\xi\}, \quad \text{where } \xi = A \cdot \frac{d\tilde{q}}{dt} \in \mathfrak{g}. \tag{2.5}$$

See Fig. 2. To check this formula, note that ξ is independent of t. This is the content of Clairut's theorem, or equivalently, of the conservation of the momentum

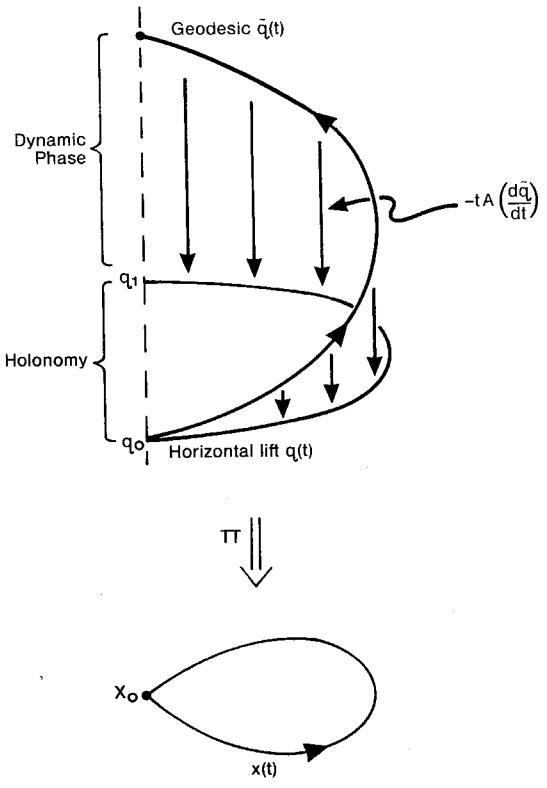

Fig. 2

map for the action of the structure group on TQ. Next, differentiate [2.5]:

$$\frac{dq(t)}{dt} = \frac{d\tilde{q}}{dt}g - \tilde{q}g\xi.$$

Here $g = \exp(-t\xi) \in G$, $q\xi$ denotes the infinitesimal generator corresponding to ξ, evaluated at $q \in Q$, and for $v \in T_qQ$, vg means $d_qR_g \cdot v$, where R_g is the action of g on Q. Now apply A:

$$A \cdot \frac{dq}{dt} = g^{-1}\left\{A \cdot \frac{d\tilde{q}}{dt} - g\xi g^{-1}\right\}g = g^{-1}\xi g - \xi = \xi - \xi = 0,$$

where we have used the fact that g commutes with ξ.

Formula [2.3] is very helpful in applying the Theorems, as it allows one to calculate the holonomy of the extremal $\pi \circ \tilde{q}$, given the geodesic $\tilde{q}$. This formula has a "Berry phase" interpretation: $\exp(t\xi)$ is the "dynamic phase," and the holonomy is the "geometric, or Berry, phase." See Berry [1985] and Fig. 2.

2. A horizontal geodesic on Q projects to a geodesic on X. Conversely, the horizontal lift of a geodesic on X is a geodesic on Q.

3. According to Theorem 2, the class $\{\pi \circ \tilde{q}\}$ of projected geodesics is independent of the choice of adjoint invariant form β. How is this possible?

Let $H_\beta : T^*Q \to \mathbf{R}$ denote the kinetic energy for the metric $\beta \oplus k$. Let

$$C_\beta(q, p) = \beta^{-1}(J(q, p), J(q, p)), \qquad [2.6]$$

where β^{-1} is the induced co-adjoint invariant inner product on $\mathfrak{g}^*$, the dual of the Lie algebra, and where

$$J : T^*Q \to \mathfrak{g}^*$$

is the momentum map for the G action on T^*Q. ($J(q, p) = \sigma_q^* p$, where $\sigma_q : \mathfrak{g} \to T_qQ$ is the infinitesimal generator of the G action.) We have the formula

$$H_\beta = H_0 + C_\beta \qquad [2.7]$$

which is simply the splitting of the total $(\beta \oplus k)$ kinetic energy into horizontal and vertical kinetic energies.

Let X_β denote the Hamiltonian vector field corresponding to H_β. This is the vector field whose flow is the $(\beta \oplus k)$-geodesic flow. Since G acts by isometries on Q, X_β is a G-invariant vector field on T^*Q. Let Y_β denote the pushforward of X_β to $(T^*Q)/G$:

$$Y_\beta = pr * X_\beta, \qquad [2.8]$$

where $\mathrm{pr} : T^*Q \to (T^*Q)/G$ is the projection. Y_β is called the *reduction* of X_β.

There is a natural sequence of projections $T^*Q \to (T^*Q)/G \to X = Q/G$. Any two geodesics (viewed as curves in the cotangent bundle), which are related by an isometry $g \in G$ have the same projections to $(T^*Q)/G$ and to X. It follows that these projected geodesics $\pi \circ \tilde{q} \subset X$ are the projections of the integral curves of Y_β.

Now C_β is a *Casimir*, that is, it Poisson commutes with all G-invariant functions

on T^*Q. This is implies that the push-forward of C_β's Hamiltonian vector field to $(T^*Q)/G$ is zero, so that

$$Y_\beta = Y_0,$$

where Y_0 is the push-forward to $(T^*Q)/G$ of the Hamiltonian vector field of H_0. Consequently, the class of projected geodesics is independent of β, as claimed.

In addition, this proves that Theorem 1 and Theorem 2 are equivalent, provided the Lie algebra admits an adjoint invariant inner product. The preceding discussion is a synopsis of one of the main results of Montgomery [1984].

4. Replace β by $\lambda\beta$, $\lambda \in \mathbf{R}$ and let $\lambda \to \infty$. This makes the vertical kinetic energy go to infinity (when written in terms of tangent vectors) and so "forces" curves to become horizontal. Now $C_{\lambda\beta} = \lambda^{-1} C_\beta \to 0$, and so $H_{\lambda\beta} \to H_0$. This gives a heuristic proof of Theorem 1.

5. In order to solve the isoholonomic problem, c must be a loop. Consequently, q must reintersect the fiber it starts from. It can be a difficult problem to describe such "re-intersecting" geodesics. See the next section.

3. Pines' Motivation, Homogeneous Bundles, and some Open Problems

Simon [1983] and Berry [1984] pointed out that the concept of holonomy appears naturally in quantum mechanics. This holonomy takes its values in the group $G = U(k)$. In the Abelian case ($k = 1$) the holonomy is popularly known as Berry's phase, and has been measured in numerous experiments (see Sect. 1.2). Alex Pines and co-workers Joe Zwanzinger, Marianne Koening, and Carl Mueller, in attempting to understand and measure the non-Abelian ($k > 1$) holonomy were led to the question: what are the best loops which give rise to a desired holonomy.

To make sense of this question, we will begin by reviewing the Abelian case. The space of states in the standard quantum mechanics is the projective Hilbert space $X = \mathbf{P}\mathscr{H}$. Over it we have the natural $U(1)$-bundle, $Q = S(\mathscr{H}) =$ unit sphere in Hilbert space, together with its canonical connection

$$A(\psi) = \langle \psi, d\psi \rangle.$$

(*A is a one-form on Q with values in $i\mathbf{R}$* = Lie algebra of $U(1)$, since $\langle \psi, \psi \rangle = 1$.)

A Schrödinger evolution

$$\frac{d\psi}{dt} = iH(t)\psi(t)$$

on $\mathscr{H}$ induces a flow on $\mathbf{P}\mathscr{H}$. Here $H(t)$, the Hamiltonian, is a t-dependent self-adjoint operator. Let $c(t) = [\psi(t)]$ be an closed orbit for this flow which has period T. Thus

$$\psi(T) = \exp(i\beta)\psi(0).$$

Writing

$$\frac{d\psi}{dt} = \left\{ \frac{d\psi}{dt} - \left\langle \psi, \frac{d\psi}{dt} \right\rangle \psi \right\} + \left\langle \psi, \frac{d\psi}{dt} \right\rangle \psi$$

and noting that this is the horizontal-vertical split of the vector $(d\psi/dt) \in TQ$, we find that

$$\beta = \mathrm{Hol}\,[c] + \int_0^T \omega(t)dt,$$

where $\mathrm{Hol}\,[c]$ is the holonomy of the loop c, and where $\omega(t) = \langle \psi(t), H(t)\psi(t) \rangle$ is the usual frequency, or energy, of oscillation. (Planck's constant is set equal to 1.) This formula for β is Berry's result, as reformulated by Aharonov and Anandan.

To measure the holonomy, take two identical systems, each prepared so that ψ is initially an eigenstate of the background Hamiltonian $H(0)$. Alter the first system by imposing fields which have the effect of changing the Hamiltonian from $H(0)$ to $H(t)$ in time t. $H(t)$ is to be chosen so that $\omega(t)$ is constant. Interfere the two systems after time t and measure the resulting phase shift. The integrals in the formulae for β cancel, and this phase shift is the holonomy of c.

What is the physical meaning of the length of c? The metric on $\mathbf{P}\mathscr{H}$ is defined by declaring that $\pi: S(\mathscr{H}) \to \mathbf{P}\mathscr{H}$ is a Riemannian submersion. This means that

$$\begin{aligned}\left\|\frac{dc}{dt}\right\| &= \left\langle \left\{\frac{d\psi}{dt} - \left\langle \psi, \frac{d\psi}{dt}\right\rangle \psi\right\}, \left\{\frac{d\psi}{dt} - \left\langle \psi, \frac{d\psi}{dt}\right\rangle \psi\right\} \right\rangle^{1/2} \\ &= \{\langle \psi, H^2\psi \rangle - \langle \psi, H\psi \rangle^2\}^{1/2} \\ &= \Delta E, \text{ the root mean square deviation in energy.}\end{aligned}$$

In matrix terms

$$\left\|\frac{dc}{dt}\right\| = \left\{ \sum_{i \neq 1} |H_{1i}|^2 \right\}^{1/2},$$

where we have picked a moving unitary frame $\{e_i\}$ in which $e_1 = \psi$. This represents the average energy needed to leave the state $[\psi]$. In other words $\|dc/dt\|dt$ is a measure of the energy output required to move from the state $c(t)$ to the state $c(t + dt)$ Thus the isoholonomic problem is essentially the following.

Find those time-dependent field configurations which generate a given phase shift with a minimum energy expenditure.

3.2. In order to investigate the non-Abelian Berry's phase, we find it helpful to view the base space X as a manifold of quantum mechanical states. Many authors prefer to view the base space as a Grassmannian of k-planes in $\mathscr{H}$. For the relation between these points of view see example C below.

Recall that a *state* is a linear functional defined on the set of observables (the self-adjoint operators on $\mathscr{H}$) which is nonnegative on the non-negative observables. A state is normalized if it has the value 1 on the unit operator. In finite dimensions, the normalized states are identified with density matrices ρ. These are non-negative hermitian operators of trace 1, the corresponding linear functional being $H \to \mathrm{trace}\,(\rho H)$. The set of density matrices can in turn be identified with a certain cone in $su(N)$, the Lie algebra of skew-symmetric trace-free matrices; the identification being $\rho \to i(\rho - 1/N)$. Here N is the dimension of $\mathscr{H}$. (Pines' main interest is in spin systems, for which $\mathscr{H}$ is finite-dimensional.)

Density matrices evolve according to the Heisenberg equation (also called the Liousville equation in this context)

$$i\frac{d\rho}{dt} = [H(t), \rho], \qquad [3.1]$$

where $H(t)$, the Hamiltonian, is a time-dependent Hermitian operator. The set $X = X(\rho_0)$ of all density matrices reachable from an initial matrix ρ_0 by all such evolutions is our base space. In other words, X is the set of all density matrices unitarily equivalent to ρ_0. Under our last identification X is an adjoint orbit in $su(N)$.

We define the bundle $Q \to X$ by focusing on a particular eigenvalue λ_1 for ρ_0. Let k be the multiplicity of λ. All operators $\rho \in X$, being conjugate to ρ_0, have λ_1 as an eigenvalue with the same multiplicity. Attach the corresponding eigenspace E_ρ to each $\rho \in X$, thus obtaining a rank k vector bundle E over X. The principal bundle $Q \to X$ is the associated frame bundle. Its fibers Q_ρ consist of all unitary frames for the vector space E_ρ.

The Abelian case is regained by taking ρ_0 to be a pure state. This means that it is a density matrix of rank 1. Any two normalized pure states are unitarily equivalent, and the set X of such states is a projective Hilbert space $\mathbf{P}\mathscr{H}$. Using Dirac's notation, the Hopf projection $Q = S(\mathscr{H}) \to X = \mathbf{P}\mathscr{H}$ is given by $\psi \to |\psi\rangle\langle\psi|$.

The Hilbert space structure of $\mathscr{H}$ induces a natural $U(N)$ invariant connection on the vector bundle E, and hence on Q. E is a sub-bundle of the trivial bundle $X \times \mathscr{H}$. A local section of E is just a function $s: U \subset X \to \mathscr{H}$ satisfying $s(\rho) \in E_\rho$. For $\rho \in X$, let $\mathbf{P}_\rho: \mathscr{H} \to E_\rho$ denote orthogonal projection. The covariant derivative D on E is defined by

$$(Ds)(\rho) = \mathbf{P}_\rho ds(\rho). \qquad [3.2]$$

D defines the connection on Q in the usual way: a moving unitary frame $\{s_i\}_{i=1,\ldots,k}$ is declared to be horizontal at $\{s_i(\rho)\} \in Q$ if each $Ds_i = 0$ at ρ.

Problem. Find the isoparallel extremals in this situation.

For this problem to be well-defined, we need a metric on X. There are two natural choices. Both are $U(N)$ invariant. The first choice is the induced metric obtained by thinking of X as a submanifold of $su(N)$ with its Euclidean (i.e. Killing form) metric. In this case, with ρ satisfying [3.1], we have

$$\left\| \frac{d\rho}{dt} \right\|^2 = C \operatorname{tr}([\rho, H]^2).$$

A convenient choice for the constant C is $1/2$.

We call the second choice of metric the *bi-invariant metric* since it is induced from the bi-invariant metric on $S = U(N)$ by declaring that the projection

$$U(N) \to X = U(N)/\{U(k_1) \times \cdots \times U(k_i)\} \qquad [3.3]$$

be a Riemannian submersion. Here the k_j are the multiplicities of the eigenvalues of ρ_0, so that $U(k_1) \times \cdots \times U(k_i)$ is the isotropy subgroup for the action of $U(N)$ on X. (See Sect. 2.3 for the definition of a Riemannian submersion.)

These two choices do not agree in general. However, they always agree (up to scale) in the very important case in which ρ_0 has exactly two distinct eigenvalues. Then X forms a Grassmannian, as can be seen by mapping $\rho \in X$ to the eigenspace Λ_ρ for (say) its top eigenvalue. To better understand the meaning of length in this case, let us further assume that ρ_0 is $1/k$ times a projection operator onto the k-dimensional subspace Λ. Then, using the evolution Eq. [3.1] one easily calculates that

$$\left\| \frac{d\rho}{dt} \right\|^2 = (\text{const}) \sum_{i \leq k < \mu} |H_{i\mu}|^2,$$

where the $H_{i\mu}$ are the matrix coefficients of H relative to a unitary frame $\{e_i\}$ whose first k elements span Λ. As in the Abelian case, this is a kind of a measure of the energy required to move the state $\rho(t)$ to the state $\rho(t + dt)$. This non-Abelian isoholonomic problem thus has the same physical meaning as the previously discussed Abelian case (Sect. 3.1).

If we choose the bi-invariant metric, then we can solve the isoholonomic problem.

Theorem 3.1. *Put the bi-invariant metric on X, the set of all density matrices unitarily equivalent to ρ_0. Let Q be the frame bundle associated to the 1st eigenbundle over X, endowed with its canonical connection [3.2]. Then the isoparallel extremals through $\rho_0 \in X$ are precisely the curves*

$$c(t) = \exp(itH_0)\rho_0 \exp(-itH_0), \qquad [3.4a]$$

where H_0 is any Hermitian matrix which staisfies $\sum_{i \neq 1} \mathbf{P}_i H_0 \mathbf{P}_i = 0$. Here identity $= \mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_i$ is the spectral decomposition of ρ_0. The parallel transport operator along c from $c(0)$ to $c(t)$ is

$$\text{Hol}(t) = \exp(itH_0)\exp(-it\mathbf{P}_1 H_0 \mathbf{P}_1) \in U(N). \qquad [3.4b]$$

Theorem 3.1 is an immediate consequence of Theorem 3.2 below.

Note that $[H_0, \mathbf{P}_1 H_0 \mathbf{P}_1] \neq 0$ in general, so that [3.4b] does not equal $\exp(it(H_0 - \mathbf{P}_1 H_0 \mathbf{P}_1))$. Also note that if $U(t)\mathbf{P}_1 = U(t)''\mathbf{P}_1$, then $U(t)$ and $U(t)''$ define the same parallel translation operator on E or Q, so this equality is to be taken modulo this relation. This formula can be found in Avron et al as Eq. [7.4].

Example A, 2nd time. Take $\mathscr{H} = \mathbf{C}^2$, and ρ_0 a density matrix with non-equal eigenvalues. Then the orbit X is isomorphic to $S^2 = \mathbf{CP}^1$, and the two eigenbundles $E_\pm \to X$ are the canonical line bundle and its negative. The frame bundle Q_+ is S^3. The projection $\pi: Q_+ \to S^2$ is the Hopf fibration. ($Q_- \to S^2$ is the anti-Hopf fibration $\pi \circ$ antipodal map.) $U(2)$ acts on $X = S^2$ by isometries. Consequently, an extremal curve $c(t)$ is the orbit of the point ρ_0 under rotation about a fixed axis. These are again the small circles.

Example C. $\mathscr{H} = \mathbf{C}^N$, and ρ_0 is a density matrix with exactly two non-equal eigenvalues, one of multiplicity k, the other of multiplicity n, where $k + n = N$. Then $X \cong G_k(\mathscr{H}) = G_{k,n}$, the Grassmannian of complex k-planes in $k + n$-space. If we focus on the eigenspaces of multiplicity k, then E is the canonical k-plane

bundle over X, and $Q \cong V_{k,n}$, the Stiefel variety consisting of all unitary k-frames in our N-dimensional space.

Choose a basis of $\mathbf{C}^N$ so that ρ_0 is diagonal. And suppose that the first eigenvalue (with multiplicity k) is the one we are focusing on. Then, the Hamiltonian H_0 of [3.4a] has the block diagonal form

$$H_0 = i\begin{bmatrix} 0 & b \\ b^t & d \end{bmatrix}$$

and its horizontal projection is

$$\mathbf{P}_1 H_0 \mathbf{P}_1 = i\begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix}.$$

We have not been able to characterize the extremals in any more detail than that given by just plugging these expressions in to [3.4a, b] of Theorem 3.1. See Sect. 3.4, "Open Problems."

3.3. Homogeneous Bundles. Consider the tower of bundles

$$S \to Q = S/K_0 \to X = S/K, \qquad [3.5]$$

S is a "big" compact Lie group containing K and K_0 as Lie subgroups. K_0 is a normal subgroup of K and $G \cong K/K_0$. Keep in mind the case of Example C immediately above. There $Q = V_{k,n}$, the Steifel variety of k-frames in C^N, with $k+n=N$, $S = U(N) \supset K = U(k) \times U(n) \supset K_0 = [I] \times U(n)$, and $G = U(k)$.

Fix an adjoint invariant metric on the Lie algebra $\mathfrak{s}$ of S. This defines a bi-invariant metric on S, which in turn induces metrics and connections on every space and projection in [3.5]. See Sect. 2.2. The metrics on S and Q are of the bi-invariant type occurring in Theorem 2. The geodesics through the identity in S are the one-parameter subgroups, $\exp t\xi$, $\xi \in \mathfrak{s}$. Such a geodesic is horizontal relative to the connection on $S \to Q$ if and only if $\xi \in \mathfrak{k}_0^\perp$. According to Remark 2.4.2, every geodesic $q(t)$ in Q through the identity coset q_0 is the projection of such a horizontal geodesic:

$$(\exp(t\xi))q_0, \quad \xi \in \mathfrak{k}_0^\perp. \qquad [3.6]$$

According to Theorem 2, the extremal paths on X, are exactly these curves, pushed down to X. We have proved

Theorem 3.2. *The isoparallel extremal loops through the identity coset $x_0 \in X$ are the paths of the form*

$$x(t) = \exp t\xi \cdot x_0, \quad \textit{where} \quad \xi \in \mathfrak{k}_0^\perp. \qquad [3.7a]$$

If the exact sequence $1 \to K_0 \to K \to G = K/K_0 \to 1$ splits, so that G is embedded as a subgroup of K (and $K \cong G \times K_0$) then the parallel transport operator, $\mathrm{Hol}(t)$, *along x from $x(0)$ to $x(t)$ is given by*

$$\mathrm{Hol}(t)(q) = \exp(t\xi)\exp(-t\mathbf{P}_{\mathfrak{g}}(\xi))q, \qquad [3.7b]$$

where $\mathbf{P}_{\mathfrak{g}}(\xi)$ is the orthogonal projection of ξ onto $\mathfrak{g}$, identified as a subalgebra of $\mathfrak{s}$.

The last fact follows from Eq. [2.5], the fact that $A(q_0)(\xi q_0) = \mathbf{P}_g(\xi)$, where q_0 is the identity coset, and the fact that $q_0 \exp(-t\mathbf{P}_g(\xi)) = \exp(-t\mathbf{P}_g(\xi))q_0$.

3.4. Open Problems. In these problems we focus on the case where $Q \to X$ is as in Example C above, the Steifel variety over $X =$ the Grassmannian.

1. Finish solving the isoholonomic problem. The extremals $x(t)$ will not close in general. In order to understand which of the isoparallel extremals are isoholonomic, that is, form closed loops, we must answer the following question.

For which $\xi \in \mathfrak{k}_0^\perp$ *does there exists a* $t > 0$ *such that* $\exp(t\xi) \in K$?

This seems to be a hard problem. We do not know the solution even for the simple case in which Q is the Steifel variety $V_{2,1}$, so that $S = U(3)$, $K = U(2) \times U(1)$ and $K_0 = \{I\} \times U(1)$.

It would also be of interest ot characterize those extremal loops which give rise to the trivial holonomy. Write the Lie algebra of $S = \mathfrak{k}_0 \oplus \mathfrak{g} \oplus \mathfrak{m}$, so that $\mathfrak{k} = \mathfrak{k}_0 \oplus \mathfrak{g}$ and $\mathfrak{k}_0^\perp = \mathfrak{g} \oplus \mathfrak{m}$. This is the problem of characterizing those pairs $(\xi_1, \xi_2) \in \mathfrak{g} \oplus \mathfrak{m}$ such that

$$\exp(t(\xi_1 + \xi_2)) \exp(-t\xi_1) \in K_0.$$

2. What is the cut locus and conjugate locus for the isoparallel extremals on the Grassmannians? Ge Zhong [1989] has made some progress on this problem. This problem, and to a lesser extent problem 1, are related to the Morse theory of horizontal paths in Q, which is one of the main investigations of Ge Zhong.

3. Find an isoholonomic inequality relating the lengths of closed loops in Grassmannians, and their holonomy. This would be a "non-Abelian" isoperimetric inequality.

In particular, according to our calculations, the length of a path in X which is parametrized by arclength is $\Delta E \Delta t$, where Δt is the length of time required to traverse the path, and ΔE is the energy. *Is there a relation between the uncertainty principle*

$$\Delta E \Delta t \geqq h/4\pi$$

and this alleged non-Abelian isoperimetric inequality? (Here h is Planck's constant.)

For example, in the Abelian case of example A, where X is the round two-sphere (or more generally for $X = \mathbf{CP}^n$) we have the isoperimetric inequality

$$\text{length} \geqq \sqrt{2\pi\Phi - \Phi^2}, \quad \text{when} \quad 0 \leqq \Phi \leqq \pi. \tag{3.8}$$

Here $\exp(i\Phi)$ is the holonomy, and $\Phi =$ solid angle/2. This isoperimetric inequality follows immediately from the more standard isoperimetric inequality

$$(\text{length})^2 \geqq 4\pi(\text{Area}) - \frac{(\text{Area})^2}{r^2}$$

for the smaller of the two areas enclosed by a Jordan curve on the sphere of radius r, together with the equalities $\Phi = \text{Area}/2r$, and $r = 1/2$. That $r = 1/2$ follows because we take the sphere in Hilbert space to have radius 1, and because the Hopf fibration is a Riemannian submersion. This "standard" isoperimetric inequality follows from

trigonometry identities together with the expressions

$$\text{Area} = 2\pi r^2(1-\cos\alpha), \quad \text{length} = 2\pi r \sin\alpha$$

for the small circles which are the isoperimetric minima. Here α is the angle between a point on the sphere and the z-axis. If we are measuring in units of Planck's constant, and if we make the change of variables $\theta = \Phi/2\pi$, then the isoperimetric inequality reads

$$\Delta E \Delta t \geqq 2\pi h \sqrt{\theta - \theta^2},$$

which is to be compared with the Heisenberg uncertainty relation.

4. Electromagnetic Analogies and Half of the Proof of Theorem 1

4.1. Particle in a Yang–Mills Field. After reduction by G, the Hamiltonian equations for H_0 become the differential equations for the trajectory of a particle in the Yang–Mills potential A. This fact can be found in Montgomery [1984]. Also compare Eq. [4.4] below with Balachandran, Borchardt and Stern [1978]. Here we review this fact, and rephrase Theorem 1 in this language.

We call the reduced differential equations "Wong's equations" after Wong [1970], and write them as Eqs. [4.2a–c]. They are equations for a curve $e(t)$ in the co-adjoint bundle

$$\mathfrak{g}^*(Q) = Q \times_{\mathrm{Ad}^*} \mathfrak{g}^* \cong V^*/G, \qquad [4.1]$$

which is a vector bundle over X with fiber $\mathfrak{g}^*$. Here V^* denotes the dual of the vertical bundle $V = \ker d\pi$. The points $e \in \mathfrak{g}^*(Q)$ are called *charges*.

Write $x(t) = \pi(e(t)) \in X$, and $\dot{x} = dx/dt \in TX$. (We abuse notation by letting π also denote the projection $\mathfrak{g}^*(Q) \to X$.) Let D denote the connection induced on the co-adjoint bundle by the connection A on Q. In coordinates,

$$De = de - (\mathrm{ad}_A)^* e.$$

Let ∇ be the *Riemannian* (Levi–Civita) connection on X induced by the metric k. Let

$$F = dA + [A, A]$$

denote the curvature of A, viewed as a two-form on X with values in the adjoint bundle $\mathfrak{g}(Q) = Q \times_{\mathrm{Ad}} \mathfrak{g} \cong V/G$. Then $e \cdot F(\dot{x}, \cdot)$ is a one-form along x, (a "force"), and $e \cdot F(\dot{x}, \cdot)^{\#}$ is a vector field along x, where "#" denotes the operation of raising indices with respect to the metric k on X. Wong's equations are

$$\nabla_{\dot{x}} \dot{x} = e \cdot F(\dot{x}, \cdot)^{\#}, \qquad [4.2a]$$

$$De/dt = 0. \qquad [4.2b]$$

They are second order in x and first order in the fiber coordinate e. We can write them as a "single" first order differential equation on the vector bundle $\mathfrak{g}^*(Q) \oplus T^*X$ by adding the equation

$$\dot{x} = y^{\#}, \qquad [4.2c]$$

($y \in T_{x(t)}{}^*X$) and then rewriting [4.2a, b] in terms of y and dy/dt. In the beginning of this section we stated that Eqs. [4.2a–c] are equivalent to the equations for an integral curve of the vector field Y_0 of Sect. 2.4.3. Recall that Y_0 is the push-down to $(T^*Q)/G$ of the Hamiltonian vector field for H_0. Equations [4.2a–c] define a vector field on the manifold $\mathfrak{g}^*(Q) \oplus T^*X$. The connection defines a G equivariant isomorphism:

$$T^*Q = V^* \oplus \mathrm{HOR}^* \cong \mathfrak{g}^* \times \pi^* T^*X,$$

which is the dual of the usual vertical-horizontal splitting. Dividing by G, we obtain the isomorphism

$$(T^*Q)/G \cong \mathfrak{g}^*(Q) \oplus T^*X. \qquad [4.3]$$

Under this identification, the equations defined by Y_0 become Eqs. [4.2a–c].

Conversely, given Y_0, the vector field X_0 on T^*Q is uniquely determined by the conditions that it project to Y_0, and that the projection of its trajectories onto Q are horizontal. It follows that Theorem 1 is equivalent to

Theorem 4. *The following conditions for a curve $x(t)$ in X are equivalent.*
A. *The curve x is an extremal for the isoparallel problem.*
B. *There is a solution $e(t) \in \mathfrak{g}^*(Q)$ to Wong's equations such that*

$$\pi \circ e = x.$$

Example A, 3rd time. The curvature of the Hopf fibration is a multiple of the area form. This the magnetic field of a monopole at the sphere's center. Wong's equations are the Lorentz equations for the motion of a charged particle constrained to the sphere. Again, these are small circles.

Example D. Let Q be the bundle of orthonormal frames of the Riemannian manifold X, and A the Levi–Civita connection. Then $\mathfrak{g}^*(Q) \cong \mathfrak{g}(Q) =$ skew symmetric endomorphisms of $TX \cong \Lambda^2(T^*X)$. (The isomorphisms are defined by the metric k.) Thus the charge $e(x(t))$ is a skew symmetric endomorphism of the tangent space at $x(t)$. We will write $F = R$ for the Riemannian curvature.

If X has constant sectional curvature K, we will now show that the corresponding isoparallel extremals are the curves whose curvatures $\{k_1, \ldots, k_{n-1}\}$ are all constant ($\dim X = n$). This is inspired by Arnold's [1961].

The constant curvature condition implies that $e \cdot R(\dot{x}, \cdot)^{\#} = Ke \cdot \dot{x}$, so that Wong's equations read

$$\nabla_{\dot{x}} \dot{x} = Ke \cdot \dot{x}, \quad \nabla_{\dot{x}} e = 0.$$

It follows that $x^{(j+1)} = Ke \cdot x^{(j)}$, where $x^{(j+1)}$ denotes the j^{th} covariant derivative of $\dot{x}$ along x, and $x^{(1)} = \dot{x}$. Using this, and the fact that e is skew, one can show that the functions $\langle x^{(j)}, x^{(k)} \rangle$ are constant along x. (They are identically zero if $k - j$ is odd.)

We now recall the definition of the curvatures k_i. For simplicity, suppose that the $\{x^{(1)}, \ldots, x^{(n)}\}$ are linearly independent. Apply the Graam–Schmidt procedure to this frame in order to obtain an orthonormal frame $\{e_1, \ldots, e_n\}$ along x, the Frenet–Serret frame. This frame satisfies the differential equations of the form

$De_1/dt = k_1 e_2$, $De_j/dt = -k_{j-1}e_{j-1} + k_j e_{j+1}$, $2 \leqq j \leqq n-1$, $De_n/dt = -k_{n-1}e_{n-1}$. The coefficients k_i are the curvatures. Set $\kappa_i^2 = \langle x^{(i)}, x^{(i)} \rangle$, and recall that these are constant. We find $e_1 = x^{(1)}/\kappa_1$, $e_2 = x^{(2)}/\kappa_2$. This shows that the first curvature is $k_1 = \kappa_2/\kappa_1$. (It is called the geodesic curvature for space curves.) An inductive argument, based on the fact that the $\langle x^{(j)}, x^{(k)} \rangle$ are constant, shows that the k_i are all constant. Compare with Arnol'd [1961].

This Gram–Schmidt procedure stops if $x^{(j+1)}$ depends linearly on the lower derivatives, but in this case $k_j = 0$ (constant) and the higher curvatures are all identically zero.

4.2. Coordinates. (See Montogomery [1984]). Trivialize Q over $U \subset X$: $\pi^{-1}(U) \cong U \times G$. Put coordinates $\{x^\mu\}$ on U, choose a basis $\{e_a\}$ for $\mathfrak{g}$, and trivialize $T^*G \cong G \times \mathfrak{g}^*$ by right translating covectors to the identity. Then, over U, $T^*Q \cong U \times G \times \mathbf{R}^n \times \mathfrak{g}^*$ and we have coordinates (x_μ, g, P_μ, e_a) on T^*Q. The e_a are linear coordinates on $\mathfrak{g}^*$, and (x^μ, P_μ) are canonical coordinates on T^*U. In these coordinates the horizontal kinetic energy is

$$H_0 = \tfrac{1}{2} k^{\mu\nu}(x)(P_\mu - e_a a_\mu^a)(P_\nu - e_a a_\nu^a). \qquad [4.4]$$

Here $k_{\mu\nu}$ is the expression for the metric k on X, and $k^{\mu\nu}$ is its inverse. The connection form A is

$$A = g^{-1}(ag + dg) \quad g \in G, \quad \text{where} \quad a = \sum a_\mu^a dx^\mu \otimes e_a,$$

the pull-back of A to X by the local section $g = \text{identity}$. Note that (x^μ, P_μ, e_a) coordinatize $(T^*Q)/G$ over U.

In case $G = U(1)$, $\mathfrak{g}^*$ is one-dimensional, and the corresponding linear coordinate e is a Casimir: $\{e, x^\mu\} = \{e, P_\mu\} = 0$, and so e is an automatic constant of the motion. If we interpret e as the electric charge then it is well-known that the Hamiltonian H_0 governs the motion of a particle travelling on X in the magnetic field da.

The substitution

$$v_\mu = P_\mu - e_a a_\mu^a$$

expresses physical momenta v_μ in terms of canonical moment P_μ and color charges e_a. (In other words, one of the equations of motion ([4.2c]) is $dx^\mu/dt = k^{\mu\nu}v_\nu$.) Together with $e_a \to e_a$, $x^\mu \to x_\mu$ this substitution defines the isomorphism [4.3].

4.3. Proof of Half of Theorem 4, and Hence Theorem 1: Wong Implies Extremal. We rephrase the isoparallel problem in terms of curves q: $[0,1] \to Q$.

Minimize the projected length:	length $(\pi \circ q)$	
subject to the constraint:	q is horizontal	(4.5)
and the fixed endpoint conditions:	$q(0) = q_0, q(1) = q_1$.	

We use the method of Lagrange multipliers. The constraint can be written

$$q^*A = 0, \qquad [4.6]$$

(q^*A is a $\mathfrak{g}$-valued connection one-form on the interval.) The Lagrange multiplier will be a *function*

$$t \mapsto e(t) \in \mathfrak{g}^*.$$

(See, for example, Courant and Hilbert [1953], vol. 1, pp. 221–222.) The functional to be extremized is

$$S(q, e) = \text{length}(\pi \circ q) - \int e(t) \cdot q^* A. \qquad [4.7a]$$

This Lagrangian is precisely the Lagrangian used to derive Wong's equations. See Balachandran, Borchardt and Stern [1978, Case 2]. In order to see this we write it in a local trivialization $U \times G \cong \pi^{-1}(U) \subset Q$. Then $q(t) = (x(t), g(t)) \in U \times G$, and $A = g^{-1}(ag + dg)$, where a is a $\mathfrak{g}$-valued connection one-form on $U \subset X$. And

$$S = \int_0^1 \left\{ \left\| \frac{dx}{dt} \right\| - e(t) \cdot g^{-1} \frac{Dg}{dt} \right\} dt, \qquad [4.7b]$$

where

$$\left\| \frac{dx}{dt} \right\| = \sqrt{k_{\mu\nu}(x) \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}}$$

and

$$\frac{Dg}{dt} = a_\mu(x) \frac{dx^\mu}{dt} g + \frac{dg}{dt}.$$

Equation [4.7b] is exactly formula (2.6) of Balachandran et al. At this point we could just quote their result to complete the proof. *The only real difference between our calculation and theirs is a matter of interpretation.* For us e is a Lagrange multiplier. For them $e(t)\delta(x - x(t))$ is the (color) current of a point particle. For completeness and clarity we will complete the proof.

When $G = U(1)$, so that $g = e^{i\theta}$, we have

$$S = \int_0^1 \left\| \frac{dx}{dt} \right\| - e(t) \cdot \left(a + \frac{d\theta}{dt} \right) dt.$$

If e were constant, then the term $e(\mathrm{d}\theta/dt)dt$ could be ignored as it represents a closed one-form. The integrand would then be the Lagrangian for a particle of charge e, travelling in X under the influence of the (electro)magnetic field $F = da$. It is well-known that the resulting Euler–Lagrange equations are the Lorentz equations, which are Wong's equations for $G = U(1)$. Now e is a constant, since

$$\frac{\delta S}{\delta \theta} = \frac{de}{dt}.$$

This proves the half of Theorem 4 (and hence Theorem 1) for $G = U(1)$.

For general G, essentially the same calculation yields Wong's equations. Varying S with respect to e yields the constraint [3.2] which says that the curve q is horizontal.

Split the variations of q into vertical and horizontal variations. Vertical

variations can be written

$$(q_\varepsilon(t)) = q(t)\exp(\varepsilon\xi(t)),$$

where $\xi(t)$ is any differentiable curve in $\mathfrak{g}$ satisfying the boundary conditions $\xi(0) = \xi(1) = 0$. Now

$$q_\varepsilon^* A = \mathrm{Ad}_{\exp}\{\varepsilon\xi(t)\}q^*A + \frac{d\xi}{dt}.$$

Imposing the horizontal constraint $q^*A = 0$, we obtain

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S(q_\varepsilon, e) = \int e\cdot\frac{d\xi}{dt} = -\int\frac{de}{dt}\cdot\xi$$

from which it follows that e is constant. (Note that the projected length is independent of vertical variations, so does not enter into the variation.) Since e is constant and q is horizontal, it follows that the projection

$$e = [q, e]\in\mathfrak{g}^*(Q) \text{ is convariantly constant.}$$

(Excuse the double use of e, please.) This is Eq. [4.2b].

Let x_ε be a variation of $x = \pi\circ q$, with derivative δx at $\varepsilon = 0$, a tangent vector along x, satisfying $\delta x(0) = \delta x(1) = 0$. Let δq denote the horizontal lift of δx, which we can extend to define a horizontal projectable vector field in a neighborhood of q. Let η_ε denote the local flow of δq. Then

$$q_\varepsilon = \eta_\varepsilon\circ q$$

is a horizontal variation. The derivative of the length functional with respect to such a variation is well known from Riemannian geometry:

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \text{length}\,(x_\varepsilon) = -\int\|\dot{x}\|^{-1}\langle\nabla_{\dot{x}}\dot{x}, \delta x\rangle dt.$$

To determine the variation of the Lagrange multiplier term, one calculates

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} q_\varepsilon^* A = q^*\mathscr{L}(\delta q)A = F(\delta q, \dot{q})dt = -F(\delta x, \dot{x})dt. \qquad [4.8]$$

Here $\mathscr{L}$ denotes the Lie derivative, and in the final equality we view F as a two-form with values in the adjoint bundle. Consequently, the derivative of the Lagrange multiplier term is

$$\int e\cdot F(\delta x, \dot{x})dt.$$

Therefore the horizontal variation is

$$\frac{\delta S}{\delta x} = -\|\dot{x}\|^{-1}k^{-1}\nabla_{\dot{x}}\dot{x} + e\cdot F(\dot{x}, \cdot). \qquad [4.9]$$

Setting this equal to zero is almost the first Wong's equation [4.2a]. Setting it equal to zero and using the skew symmetry of F and the covariant constancy of e implies that $\|\dot{x}\|$ is constant along x. Then redefining e to be $\|\dot{x}\|e$, we obtain Wong's Eqs. [4.2a–c]. Q.E.D.

5. Sub-Riemannian Metrics and Proof of the Hard Half

5.1. Bär's Result. The hard part of proving Theorem 1 is to show that every extremal satisfies the Hamiltonian differential equation. We will do this by simply quoting a recent result of Bär's [1988, 1989] concerning sub-Riemannian metrics.

Recall from the introduction that a sub-Riemannian structure on Q is a "horizontal" distribution Hor, together with a metric κ on Hor. As noted in the introduction (see Eqs. [1.2a, b]), the isoparallel problem is a special case of the sub-Riemannian geodesic problem.

Definition. A sub-Riemannian geodesic $q(t)$ on Q is a horizontal curve which extremizes the integrated energy functional

$$E[\gamma] = \tfrac{1}{2}\int \kappa(\dot\gamma(t), \dot\gamma(t))dt$$

among all piecewise C^1 *horizontal* paths γ which join $q(0)$ to $q(1)$.

As in Riemannian geometry, the extremals of E and of the length functional are the same, when viewed as unparametrized curves. The energy functional is more convenient from the point of view of analysis.

A sub-Riemannian metric defines (and is defined by) a constant rank co-metric C. C is a symmetric non-negative vector bundle endomorphism $C: T^*Q \to TQ$, which is defined by the requirements that

(1) $\mathrm{im}C = \mathrm{Hor}_q \subset T_qQ$,
(2) if $v = C(q)\cdot p$, then $k_q(v, v) = \mathrm{p{\cdot}v}$.

Alternatively, C is a smooth, constant rank, contravariant, symmetric, non-negative two tensor:

$$C(q)(p_1, p_2) = p_1 \cdot [C(q)\cdot p_2].$$

The fiber-wise quadratic form

$$H_0(q, p) = \tfrac{1}{2}C(q)(p, p) \qquad [5.2]$$

will be called the *horizontal kinetic energy*, or *sub-Riemannian kinetic energy*. It is a smooth function on T^*Q. A straightforward calculation shows that this H_0 equals the earlier H_0 in the case of the isoparallel problem.

Theorem 5. [Bär, [1988, 1989]]. *Every sub-Riemannian geodesic is the cotangent projection to Q of a solution on T^*Q to the Hamiltonian differential equations for the Hamiltonian H_0.*

The other half of Theorem 1 is a special case of this theorem.

In canonical coordinates (q^i, p_i) on T^*Q the differential equations of the theorem are

$$dq^i/dt = \sum C^{ij}p_j;$$
$$dp_i/dt = -\tfrac{1}{2}\sum [\partial(C^{kj})/\partial q^i]p_k p_j.$$

Here $H_0 = \frac{1}{2}\sum C^{kj}(q)p_k p_j$. Note that the first equation implies that $q(t)$ is horizontal.

5.2. Remarks and History. There is a large literature on the sub-Riemannian geodesic problem. Vershik and Gershkovich [1988] is a review with a summary of facts, some intriguing pictures, and an extensive bibliography. Beyond the works mentioned in Sect. 1.3, the following works have come to our attention: Hermannn [1962, 1973], Brockett [1981], Baillieul [1975], Gunther [1982], Faibusovich [1988], and Taylor [1989]. The sub-Riemannian geodesic problem is a special case of the problem of Lagrange in the Calculus of Variations. This is treated by Carathéodory [1967, final chapter], and by Bliss [1930].

The converse to Bär's theorem (our previous section) was proved for H^1 paths by Hamenstädt [1986, 1988]. She also showed that any solution to the differential equations is locally length minimizing. Bär has an interesting counterexample which shows that locally length minimizing curves need not satisfy the Hamiltonian differential equations globally, a situation impossible in Riemannian geometry.

The proof of Bär's theorem is easy in the extreme cases where the connection A is flat or *fat*. (See Ge Zhong [1989], or Strichartz [1983].) This is because in these cases the set of horizontal paths joining q_0 to q_1 forms a smooth manifold, and so standard calculus techniques, such as the Lagrange multiplier technique of Sect. 4 apply. In the flat case one can work on a single integrable leaf of the horizontal distribution, and the problem is identical to the Riemannian geodesic problem. The condition that a connection be fat is equivalent to the condition that its horizontal distribution satisfy what is sometimes called the "strong bracket generating condition." "Fatness" means that for every non-zero $v \in \mathrm{Hor}_q$, the map $w \to F_q(v, w)$ is onto $\mathfrak{g}$. (F is the curvature of A.) Fatness implies that every $q_1 \neq q_0$ is a regular value of the end-point map

$$e: \{\text{horizontal } H_1\text{-paths starting at } q_0\} \to Q; e(q) = q(1).$$

In general the rank of e varies (Hamenstädt).

Bär's proof does not make any assumptions regarding the horizontal distribution. In fact, the co-metric C may even have variable rank, in which case the "distribution," $\mathrm{im}(C)$ is a singular one. His proof is based on a partial proof of Strichartz [1983]. (Strichartz's proof contains an error. He ignored the possibility that his H, defined by a minimization procedure could have the value zero.) Strichartz's idea is to apply the Pontrjagin maximum principle, as found in Cesari [1983], Chap. 7.

The essence of the difficulty in the proof is that extremals may be *abnormal.* The method of Lagrange multipliers, *in full*, is to find $(e_0, e(t))$, not identically zero, with $e_0 \in \mathbf{R}$, such that

$$e_0 \,\mathrm{length}(\pi \circ q) - \int e(t) \cdot q^* A$$

has a critical point as a function of q. (Compare with [4.7a].) In Sect. 4 we set $e_0 = 1$. Abnormal extremals (Bliss's terminology) are ones for which $e_0 = 0$. *Truly abonrmal extremals* (our terminology) are ones for which *every* nonzero multiplier satisfies $e_0 = 0$. If one can eliminate these, then the standard Euler–Lagrange equations of Sect. 4 apply. The crux of Bär's argument is then to eliminate these.

6. The Cat's Problem

The configuration space Q for a deformable body is a submanifold of the space of embeddings of the body B into Euclidean 3-space. A point q of Q is then a map

$$\mathbf{q}: B \to \mathbf{R}^3; \quad x = \mathbf{q}(X) \in \mathbf{R}^3, \quad X \in B. \tag{6.1}$$

The body B is assumed to have a mass density, $dm(X)$, which together with the inner product $\langle \cdot, \cdot \rangle$ on $\mathbf{R}^3$ defines a Riemannian metric d^2s on Q:

$$d^2s_q(\delta q, \delta q) = \int_B \langle \delta \mathbf{q}(X), \delta \mathbf{q}(X) \rangle dm(X). \tag{6.2}$$

The group G of rigid motions (isometries of $\mathbf{R}^3$) acts isometrically on Q. The action is left composition: $g\mathbf{q} = g \circ \mathbf{q}$, and corresponds to rigidly rotating and translating B. If the body is never colinear ($q(B)$ is never contained in a single line) then the action is free.

We thus have the following situation. The Lie group G acts freely, properly, and by isometries on the Riemannian manifold Q. From this data we can recover the data (Sect. 1.1) needed to state the isoholonomic problem. Set $X = Q/G$. It is the *shape space* of our deformable body, and $\pi: Q \to X$ forms a (left) principal G-bundle. X inherits a Riemannian metric by declaring π to be a Riemannian submersion (Sect. 2.3). Q inherits a connection by declaring that horizontal is orthogonal to vertical ($= \ker d\pi$).

We will now show that in this setting the isoholonomic problem is

The Cat's Problem: Given a deformable body in free-fall with initial angular momentum zero, find the most efficient way to deform it so as to achieve a desired re-orientation.

We will ignore the translational degrees of freedom in G, because changing the shape of a freely falling body cannot affect the motion is its center of mass. Consequently, we will fix the center of mass $\int \mathbf{q}(X) dm(X)$ by setting it equal to zero. This defines a new fixed center of mass configuration space which we again call Q. The group G becomes the group of rotations about the center of mass. (Shapere and Wilczek are interested in translations of their paramecium. Affecting the translation is possible here because strong friction is present, so that linear momentum is not conserved.)

The basic observation which translates one problem into the other is the following:

Observation. A tangent vector $(\mathbf{v}, \mathbf{q}) \in T_{\mathbf{q}}Q$ is horizontal if and only if its angular momentum is zero.

Check. A vertical tangent vector at $\mathbf{q} \in Q$ is an infinitesimal rigid rotation:

$$\delta \mathbf{q}(X) = \omega \times \mathbf{q}(X). \tag{6.3}$$

A vector $\mathbf{v} \in T_{\mathbf{q}}Q$ is horizontal, by definition, if and only if it is orthogonal to all such variations, that is, if and only if

$$\int \mathbf{v}(X) \cdot \{\omega \times \mathbf{q}(X)\} dm(X) \quad \text{for all} \quad \omega \in \mathbf{R}^3. \tag{6.4}$$

After a simple rearrangement, this becomes the statement

$$\int \mathbf{q}(X) \times \mathbf{v}(X) dm(X) = 0, \qquad [6.5]$$

which is the statement that the angular momentum vanishes. This shows that the cat's problem is equivalent to the isoholonomic problem, *provided we define efficiency in the cat's problem to be the integrated kinetic energy.* Note that the re-orientation of the body after a shape change is the holonomy of the loop in shape space.

One calculates the H_0 of [2.2] to be

$$H_0(q,p) = \tfrac{1}{2}\{\|p\|^2 - \mathbf{I}_q^{-1}(J(q,p), J(q,p))\}. \qquad [6.6]$$

The terms in H_0 are as follows: $\frac{1}{2}\|\mathbf{p}\|^2$ is the standard kinetic energy for the metric on Q. $\frac{1}{2}\mathbf{I}_q^{-1}(J(q,p), J(q,p))$ is the vertical kinetic energy, so when subtracted off it yields the horizontal kinetic energy. The factors within this vertical kinetic energy are as follows. $\mathbf{I}_q$ is the *locked interia tensor.* This is the interia tensor of our cat if we froze it in the shape q.

$$\mathbf{I}_q = (\operatorname{tr} \Psi_q)1 - \Psi_q, \qquad [6.7a]$$

where

$$(\Psi_q)^{ij} = \int q(X)^i q(X)^j dm(X) \quad \text{for} \quad i,j = 1,2,3. \qquad [6.7b]$$

Geometrically, $\mathbf{I}$ is the pull-back of the metric k to $\mathfrak{g}$:

$$\mathbf{I}_q(\omega_1 . \omega_2) = d^2 s_q(\sigma_q \omega_1, \sigma_q \omega_2) \quad \text{for } \omega_i \in \mathfrak{g}. \qquad [6.7c]$$

Here $\sigma_q \omega = q \times \omega$ is the infinitesimal generator, [6.3]. $\mathbf{I}_q$ is invertible, since the G-action is free. Thus $\mathbf{I}_q^{-1}$ is well-defined as a positive definite quadratic form on $\mathfrak{g}^*$. J is the total angular momentum, *written as a function of the canonical momenta p, and not of velocity.* In mathematical terms $J: T^*Q \to \mathfrak{g}^*$ is the momentum map for the action of G.

Warning. Be careful of the difference between this angular momentum and the corresponding angular momentum

$$M: TQ \to \mathfrak{g}^*$$

written in terms of velocity (the left-hand side of [6.5]). The two are related by $J(q,p) = M(q,v)$ *provided*

$$v = p^{\#} \qquad [6.8]$$

($^{\#}$ is the index raising operation relative to $d^2 s_q$.) However [6.8] does not hold along general integral curves $(q(t), p(t))$ of H_0. In fact every such curve satisfies

$$M(q, \dot{q}) = 0,$$

since q is horizontal, but J is a constant of the H_0-motion whose value $J(q(t), p(t))$ is an arbitrary constant (depending on initial conditions).

Theorem 6. *A curve q in Q is an extremal for the cat problem if and only if there exists a smooth covector $p(t)$ along q such that $(q(t), p(t))$ satisfies Hamilton's differential equation for the Hamiltonian H_0.*

Theorem 6 follows immediately from Theorem 1 and the above discussion.

Remarks 1. We could have stated the cat problem for a spinning cat. Then the constraint would have been $M(q, v) = \mu$, a fixed constant vector. Theorem 6 still holds provided we replace H_0 by $H_0 + \mathbf{I}_q^{-1}(J(q, p), \mu)$.
2. Theorem 6 still holds if the G action is only locally free (all isotropy groups are discrete).
3. If d^2s is the bi-invariant metric $\beta \oplus k$ on Q which was described in Sect 2.2, then the inertia tensor $\mathbf{I}_q$ is identically equal to β. The vertical kinetic energy, the second term of [6.6] is the Casimir C_β of 2.2.3.
4. Shapere and Wilczek [1987, 1989] give a formula for the connection one-form A which defines the horizontal subspace here (i.e. the "zero angular momentum connection"). Their formula is

$$A_q = \mathbf{I}_q^{-1} M(q, \cdot): T_q Q \to \mathfrak{g}. \qquad [6.9]$$

5. J is a constant of the H_0-dynamics. If we fix the value of this constant, then we can view the motion as that of a particle in a potential field defined by the second term of [6.6]. This potential is exactly the negative of what is usually called the effective potential, $V_{\text{eff}} = \frac{1}{2}\|\alpha\|^2$, the square of the covector α which is the J-component of the connection form A.

7. A problem of Shapere and Wilczek

Shapere and Wilczek [1987] posed a problem closely related to the isoholonomic and cat's problem in their beautiful paper on the self-propulsion of microorganisms. For them the group G is $E(3)$, the group of Euclidean motions, and the metric k measures power output for a given path x in the space X. Let $\chi: G \to \mathbf{R}^+$ be the length of the translational factor: $\chi(g, v) = \|v\|^2$, $v \in \mathbf{R}^3$. Set

$$E[x] = \tfrac{1}{2}\int \|\dot{x}\|^2 dt.$$

They define the *effeciency* of a curve x into X to be

$$\text{Eff}[x] = \frac{\chi(\text{Hol}[x])}{E[x]}. \qquad [7.1a]$$

More generally, let

$$\chi: G \to \mathbf{R}^+$$

be a *class function* on G. (A class function is a conjugation invariant function, $\chi(ghg^{-1}) = \chi(h)$, for example, the trace on the unitary group.) And fix

$$f: \mathbf{R} \times \mathbf{R}^+ \to \mathbf{R},$$

a smooth function. Set

$$\text{Eff}[x] = f(\chi(\text{Hol}[x]), E[x]) \qquad [7.1b]$$

and call this the *efficiency* of the path $x: [0, 1] \to X$. The problem of Shapere and

Wilczek is to

find the loops of maximum efficiency.

Shapere and Wilczek actually state an infinitesimal version of this problem. They look for infinitesimal loops. Their definition of efficiency is the infinitesimal version of ours: replace the holonomy by the curvature, and the integral by the integand $\frac{1}{2}\|\dot{x}\|^2$.

Theorem 7. *Assume that $x(t)$ is a loop in X which maximizes the efficiency* [4.1a], *is piecewise smooth, and satisfies $\chi(\mathrm{Hol}[x]) \neq 0$. Then x is the projection of a solution to Wong's equation,* [4.3a–c].

Proof. Theorem 4 states that isoholonomic extremals solve Wong's equations. The isoholonomic extremals solve the following constrained variational problem: extremize E subject to the constraint Holonomy = constant.

In general, suppose one is trying to extremize a function E subject to a constraint $h = \text{const}$. The resulting Euler–Lagrange equations are $\lambda_0 dE + \lambda dh = 0$ for some choice of non-zero multipliers λ_0, λ. Compare this with extremizing $\mathrm{eff}(x) = f(E(x), \chi(h(x)))$:

$$d(\mathrm{eff}) = \frac{\partial f}{\partial E} dE + \frac{\partial f}{\partial \chi}\frac{\partial \chi}{\partial h} dh.$$

This demonstrates that if p is a critical point of (eff) for which at least one of these two coefficients, $\lambda_0 = \partial f/\partial E$, and $\lambda = (\partial f/\partial \chi)(\partial \chi/\partial h)$, are non-zero, then p is an extremal for the constrained variational problem. For Shapere and Wilczek, $f = \chi(h)/E$, so that $\partial f/\partial E = -\chi(h)/E^2$ in non-zero, provided $\chi(\mathrm{Hol}[x]) \neq 0$. (That this be non-zero is actually the condition that the extremal be *normal*. See the end of Sect. 5.) Q.E.D.

Remark. This theorem can also be proved by direct calculation using the formula

$$d\,\mathrm{Hol}[x] \cdot \delta x = \int \{U_2(t) F(\dot{x}(t), \delta x(t)) U_1(t)\} dt$$

for the variation of the holonomy. Here $U_1(t)$ denotes the operation of parallel translation along x from $x(0)$ to $x(t)$, and $U_2(t)$ is parallel translation along x from $x(t)$ to $x(1)$. Using the fact that $\mathrm{Hol}[x] = U_2(t)U_1(t)$, this can be rewritten in terms of just $U_1(t)$ and $\mathrm{Hol}[x]$.

Acknowledgements. I am pleased to acknowledge Alex Pines for formulating this problem, and for useful discussions. I would also like to thank Malcolm Adams, Jeeva Anandan, Juan Simo, Alan Weinstein, Bruce Kleiner, Ge Zhong, Ralf Spatzier, Tadeusz Januszkiewicz and Ursula Hamenstädt for helpful conversations and directions to the literature. Eugene Lerman translated some of the encyclopaedia article of Vershik and Gershkovich. Richard Cushman provided some editorial criticism. Gorky, Claudine Swickard's cat, provided inspiration and experimental know-how for Sect. 6.

This work was done while at M.S.R.I., funded by NSF Postdoctoral grant #DMS-8807219.

References

Aharonov, Y., Anandan, J.: Phase change during cyclic quantum evolution. Phys. Rev. Lett. **58**, 1593–1596 (1987)

Ambrose, W., Singer, I. M.: A theorem on holonomy. Trans. AMS **75**, 428–453 (1953)
Arnol'd, V. I.: Some remarks on flows of frames. Sov. Math, translations of Doklady. USSR, **2**, 562–564 (1961)
Arnol'd, V. I., Kozlov, V. V., Neishtadt, A. I. (1988): Dynamical systems III. vol. 3. In: The Encyclopaedia of Mathematical Sciences series. Berlin, Heidelberg, New York: Springer 1988
Avron, J. E., Sadun, L., Segert, J., Simon, B.: Chern numbers and Berry's phases in fermi systems. Commun. Math. Phys. **124**, 595–627 (1989)
Baillieul, J. B.: Geometric methods for nonlinear optimal control problems. J. Optimization Th. Applications **25**, 519–548 (1975)
Balachandran, A P., Borchardt, S., Stern, A.: Lagrangian and Hamiltonian descriptions of Yang–Mills particles. Phys. Rev. **D17**, 3247–3256 (1978)
Bär, C.: Carnot–Caratheodory–Metriken. Diplomarbeit, Bonn 1988
Bär, C.: Geodesics for Carnot–Caratheodory Metrics. Preprint 1989
Berry, M. V.: Quantal phase factors accompanying adiabatic changes. J. Phys. A. **18**, 15–27 (1984)
Bliss, G. A.: Lectures on calculus of variations. Chicago, IL: Univ. of Chicago Press 1946
Bliss, G. A.: The problem of Lagrange in the calculus of variations. Am. J. Math. **52**, 674–713 (1930)
Brockett, R. W.: Control theory and singular Riemannian geometry. In: New directions in applied mathematics. Hilton, P. J., Young, G. S. (eds). Berlin, Heidelberg, New York: Springer 1981
Carathéodory, C.: Calculus of variations and partial differential equations of the first order, vol. 2. Holden-Day, S.F., CA 1967
Cesari, L.: Optimization—Theory and applications. Berlin, Heidelberg, New York: Springer 1983
Chow, W. L.: Uber Systeme van Linearen partiellen Differentialgleichungen erster Ordnung. Math. Ann **117**, 98–105 (1939)
Courant, R., Hilbert, D.: Methods of mathematical physics vol. I, New York: Interscience 1953
Faibusovich, L. E.: Explicitly solvable nonlinear optimal controls. Int'l J. Control **48**, 2507–2526 (1988)
Gunther, N. L.: Hamoltonian mechanics and optimal control. Harvard thesis 1982
Ge Zhong: On a constrained variation problem and the space of horizontal paths. M.S.R.I. preprint #04224-89 (1989)
Hamenstädt, U.: Über Theorie von Carnot–Caratheodory–Metriken und ihren Anwendungen. Doktorarbeit, Bonn 1986
Hamenstädt, U.: Some regularity theorems for Carnot–Caratheodory metrics. Preprint, Cal. Tech. 1988
Hermann, R.: Some differential geometric aspects of the lagrante variational problem. Indiana Math. J. 634–673 (1962)
Hermann, R.: Geodesics of singular Riemannian metrics. Bull. AMS **79**, 780–782 (1973)
Iwai, T.: A gauge theory for the quantum planar three-body system. J. Math. Phys. **28**, 1315–1326 (1987a)
Iwai, T.: A geometric setting for internal motions of the quantum three-body system. J. Math. Phys. **28**, 1315–1326 (1987b)
Iwai, T.: A geometric setting for classical molecular dynamics. Ann. Inst. Henri Poincairé, Phys. Th., **47**, 199–219 (1987c)
Kane, T. R., Scher, M. P.: A dynamical explanation of the falling cat phenomenon. Intl. J. Solids Structures, **5**, 663–670 (1969)
Koenig, M., Mueller, C., Zwanziger, J.: private conversations (1989)
Montgomery, R.: Canonical formulations of a classical particle in a Yang–Mills field and Wong's equations. Lett. Math. Phys. **8**, 59–67 (1984)
Montgomery, R.: Shortest loops with a fixed holonomy. MSRI preprint series #01224-89 (1988)
Montgomery, R.: Optimal control of deformable bodies, isoholonomic problems, and sub-Riemannian geometry. MSRI preprint series #05324-89 (1989)
Shapere, A.: Gauge mechanics of deformable bodies. PhD. thesis, Physics, Princeton (1989)
Shapere, A., Wilczek, F.: Self-propulsion at low Reynolds number. Phys. Rev. Lett. **58**, 2051–2054 (1987)
Simon, B.: Holonomy, the quantum adiabatic theorem, and Berry's phase. Phys. Rev. Lett. **51**, 2167–2170 (1983)
Strichartz, R.: Sub-Riemannian geometry. J. Diff. Geom. **24**, 221–263 (1983)
Suter, D., Mueller, K. T., Pines, A.: Study of the Aharonov–Anandan quantum phase by NMR interferometry. Phys. Rev. Lett. **60**, 1218–1220 (1988)

Taylor, T. J. S.: Some aspects of differential geometry associated with hypoelliptic second order operators. Pac. J. Math. **136**, 355–378 (1989)
Tomita, A., Chiao, R. Y.: Observation of Berry's topological phase by use of an optical fiber. Phys. Rev. Lett. **57**, 937–940 (1986)
Tycko, R.: Adiabatic rotational splittings and Berry's phase in nuclear quadraplole resonance. Phys. Rev. Lett. **58**, 2281–2284 (1987)
Vershik, A. M., Ya Gershkovich, V.: Non-holonomic Riemannian manifolds. In: Dynamical systems vol. 7, part of the new Mathematical Encyclopaedia series vol. 16. In Russian, MIR pub. Berlin, Heidelberg, New York: Springer 1988
Weinstein, A.: Fat bundles and symplectic manifolds. Adv. Math. **37**, 239–250 (1980)
Wilczek, F.: Gauge theory of deformable bodies. Inst. Adv. Studies preprint #-88/41 (1988)
Wilczek, F., Zee, A.: Appearence of gauge structure in simple dynamical systems. Phys. Rev. Lett. **52**, 2111–2114 (1984)
Wong, S. K.: Field and particle equations for the classical Yang–Mills field and particles with isotopic spin. Nuovo Cimento **65A**, 689–693 (1970)

Communicated by B. Simon

Received July 17, 1989; in revised form August 29, 1989

Note added in proof. It was brought to our attention that Guichardet defined and used the connection "angular momentum equals zero" in his 1984 paper "On Rotation and vibration motions of molecules", Ann. Inst. Henri Poincare, *40*, 329–342. This paper contains Shapere and Wilczek's "master formula" for the connection, our Eq. [6.9], and also a nice descriptions of its curvature. Guichardet proves that when the deformable body consists of four or more point particles, that the distribution satisfies Hormander's condition, and hence is controllable (see our Sect. 1.5).

Zwanziger, Koenig, and Pines have completed their experiment to measure the non-Abelian holonomy (Berry's phase) and have submitted the work to Phys. Rev. Lett.. Their experiment concerns the nuclear quadrapole resonance spectrum of a crystal of sodium chlorate which is rotating simultaneously about two axes (curves of the form $\exp(ta)\exp(tb)$ in $SO(3)$).