R. Montgomery, 11/1/88

Shortest Loops with a Fixed Holonomy.

§1. The Problem and an Introduction.

The physical chemist Alex Pines posed the following constrained variational problem:

Pines' Problem for loops: Among all loops γ based at x_0, and with a fixed holonomy $H[\gamma] = H_0$, find the loops of minimum length.

Here γ is a loop in a Riemannian manifold M, which is the base space of a principal bundle $\pi:P\rightarrow M$ with structure group the connected Lie group G. P is endowed with a fixed connection A. (Pines' interest is in the quantum Berry phase. Here the relevant bundle is $V_{k,n} \rightarrow G_{k,n}$, the bundle of hermitian orthonormal k-frames over the Grassmannian of complex k-planes in complex k+n space, with its canonical connection.)

The curve γ really need not be a loop in order to define Pines' problem. A curve $\gamma \subset M$ joining x_0 to x_1 defines a parallel translation map

$$H[\gamma]: P_0 \rightarrow P_1$$

where P_0, P_1 are the fibers of P over x_0, x_1 respectively. Recall the definition of the parallel translation $H[\gamma]$. Let $\overline{\gamma}$ denote the horizontal lift of γ with initial condition $\overline{\gamma}(0) = p \in P_0$. Then $H[\gamma](p) = \overline{\gamma}(t_1) \in P_1$, where $\gamma(t_1) = x_1$ is the endpoint of the base curve. $H[\gamma]$ satisfies $H[\gamma](pg) = H[\gamma](p)g$. In case γ is a loop we call $H[\gamma]$ the *holonomy of* γ, and by fixing the initial condition p, we can identify it with an element g of G. This is done by writing $\overline{\gamma}(t_1) = pg$. We call g *the holonomy measured from* p. If G is non-Abelian, this identification depends on p: if we change p to pg_1 we find that g becomes $g_1^{-1}gg_1$.

Pines' Problem, general: Fix points $p \in P_0$ and $q \in P_1$. Among all horizontal curves $c \subset P$ joining p to q, find those curves for which the length of $\gamma = \pi c$ is minimized.

Note that $H[\gamma]p = q$ since c is horizontal. Consequently we can restate the general problem:

"minimize the length for a fixed parallel transport $H_0 = H[\gamma]$."

Shapere and Wilczek [1987] pose a similar problem in their beautiful paper on the self-propulsion of microorganisms. Define the *efficiency* of a curve to be

$$E[\gamma] = \chi(H[\gamma]) / {}^1/_2 \int \|\dot{\gamma}\|^2$$

where χ is a given class function on G. Recall that a *class function* is a conjugation invariant function: $\chi(ghg^{-1}) = \chi(h)$. Shapere and Wilczek take

$$\chi(h) = \text{distance from the identity to } h$$

with respect to a bi-invariant metric on G. (Actually they use an infinitesimal version of our E's numerator, namely $\chi'(\delta H/\delta\gamma)$.) In Shapere and Wilczek's situation the numerator of E is the distance travelled by the organism, and the denominator is the power output.

Problem of Shapere and Wilczek. Find the loops of maximum efficiency.

Isoperimetric Problem.

In the case $G = U(1)$ and γ bounds a disc D, we have the formula

$$H[\gamma] = \exp\{i \int_D F\}$$

where F is the curvature of the connection, thought of as a two-form on

M. In case M is two-dimensional and F = const.×area form, Pines' problem then becomes the isoperimetric problem: among all loops enclosing a fixed area, find those of minimum length. The solutions to this problem are curves of constant geodesic curvature. In case $P \to M$ is the Hopf fibration $S^3 \to S^2$ with its standard connection, these loops are "small circles", eg. lines of lattitude. Each such circle is the projection of a geodesic on S^3.

This suggests the following "Kaluza-Klein" approach to Pines' problem. Recall that a Kaluza-Klein metric [K-K for short] on P is constructed by declaring vertical and horizontal subspaces to be orthogonal, putting a fixed bi-invariant metric on the fiber G of P, and putting the base metric $ds^2{}_M$ on the horizontal subspaces. We write $d^2s_P = d^2s_M \oplus d^2s_{\mathfrak{g}}$. Alternatively, a K-K metric on P is one for which G acts by isometries, and such that the fibers are all isometric to G with a fixed bi-invariant metric. Given such a metric on P, the connection is reconstructed by defining the horizontal distribution to be the orthogonal complement to the vertical, and the metric on M is reconstruced by insisting that the projection π is a Riemannian submersion.

Theorem 1. *Assume that G admits a bi-invariant non-degenerate metric, and use this to put a Kaluza-Klein metric on P. Then the projection γ of every Kaluza-Klein geodesic $\tilde{\gamma}$ on P is an extremal for Pines' constrained variational problem. If the connection is fat, or flat, the converse is true: every extremal loop, and in particular every minimal loop $\gamma \subset M$ for Pines' constrained variational problem is the projection onto the base manifold M of a Kaluza-Klein geodesic $\tilde{\gamma}$ on the principal bundle P.*

The bundle with connection is said to be "fat" (Weinstein[1980]) if for every $x \in M$ and $Q \neq 0$ in the co-adjoint bundle at x, $Q \cdot F(\cdot\,,\cdot)$ is a nondegenerate two-form. Here F is the curvature of the connection.

Theorem 2. *Every piecewise* C^1 *extremal* γ *for the problem of Shapere and Wilczek , with* $\chi(H[\gamma]) \neq 0$, *is the projection of a Kaluza-Klein geodesic.*

Remark 1.1. The K-K geodesic $\tilde{\gamma}$ will generally not be horizontal. In fact, a curve $\tilde{\gamma}$ in P is a horizontal geodesic if and only if its projection $\gamma = \pi\tilde{\gamma}$ is a geodesic on M. If $\tilde{\gamma}$ is a geodesic in P, then the horizontal lift, $\overline{\gamma}$,of its projection γ, is

$$\overline{\gamma}(t) = \tilde{\gamma}(t)\exp\{-tQ_0\},$$
$$\text{where } Q_0 = A\cdot d\tilde{\gamma}/dt \in \mathfrak{g} \qquad [1.1].$$

See figure 1. To check this, first note that Q_0 is independent of t. (This is Clairut's theorem, or, conservation of the momentum map for the action of the structure group on TP.) Differentiate [1.1]:

$$d\overline{\gamma}(t)/dt = (d\tilde{\gamma}(t)/dt)g - \tilde{\gamma}gQ_0$$

Here $g = \exp(-tQ_0)$. pQ_0 denotes the infinitesimal generator corresponding to Q_0, evaluated at p. And for $v \in TP$, vg means $TR_g v$. Now apply A:

$$\begin{aligned} A\cdot d\overline{\gamma}(t)/dt &= A\cdot[(d\tilde{\gamma}(t)/dt)g] - Q_0 \\ &= g^{-1}Q_0 g - Q_0 \\ &= Q_0 - Q_0 = 0, \end{aligned}$$

where we have used the fact that g commutes with Q_0.

Formula [1.1] is very helpful in applying the Theorems, as it allows one to calculate the holomomy $H[\gamma]$, given the geodesic $\tilde{\gamma}$. This formula has a Berry phase interpretation: $\exp(tQ_0)$ is the "dynamic phase", and $H[\gamma]$ is the "geometric phase". See Berry [1985].

Remark 1.2. The Pines' minimizer γ is not unique. This can be seen in the case $G = S^1$, $M = R^2$ or S^2 discussed above. There are an infinite number (a circle's worth) of circles through x_0 with fixed area.

Remark 1.3. In the next section we will give a counterexample to the converse in Theorem 1 for the non-fat case.

Outline of Paper. The theorems will be proved by using Wong's equations. These are the classical equations of a particle in a Yang-Mills field A, and can be viewed as the Poisson reduction of the K-K geodesic equations.

We discuss Wong's equations in §2. We also state Theorem 3 there, which says that extremals are solutions to Wong's equations. A precursor of Theorem 3 appears in Shapere and Wilczek. Their equation (30), the Euler-Lagrange equations for their efficiency functional (an infinitesimal version of ours), is Wong's equations! Theorem 1 follows from Theorem 3 and various lemmas proved in §2. It is proved at the end of §2.

In §3 we prove Theorem 3. The proof is by the method of Lagrange multipliers. In §4 we prove Theorem 2. This proof requires calculating the derivative of the holonomy with respect to variations of the loop.

In §5 we calculate some of the extremals for some of the bundles $V_{k,n} \to G_{k,n}$ mentioned at the beginning: the bundle of k-frames over the Grassmannian $G_{k,n}$ of k-planes in C^N, $N = k+n$, with its canonical connection. $G_{k,n}$ can be viewed as a space of projection operators occuring in quantum mechanics, and for this reason these bundles are of basic physical interest. These bundles-with-connection are homogeneous, and for this reason we can give fairly explicit formulas for the extremals: they can all be expressed in terms of exponentials of constant skew-hermitian N×N matrices. See Theorem 5. However this answer is far from complete, as it is rather difficult, if not impossible, to analytically exponentiate an arbitrary N×N matrices. We end up computing all the extremals for the Abelian case, $k = 1$, and certain classes of extremals for the first non-abelian case, $(k,n) = (2,1)$.

In §6 we list some open problems. In the appendix we present a pleasing, but unfortunately heuristic, geometric "proof" of Theorem 1. This involves letting the fiber part of the Kaluza-Klein metric go to infinity, and arguing that the resulting geodesics become horizontal. This appendix is joint work with Alan Weinstein.

§2. Wong's equations.

We will end up proving something more general than Theorem 1, namely that Pines' extremals satisfy Wong's equations. In case G admits a bi-invariant metric, Wong's equations are the (Poisson) reduction of the geodesic equations on P by the action of the structure group. The point is that Wong's equations, and Pines' problem, still make sense when G does not admit a bi-invariant metric, for example when G is the Heisenberg group.

Wong's equations are:

$$\nabla_{\dot{\gamma}}\dot{\gamma} = Q \cdot F(\dot{\gamma}, \cdot\,) \qquad \text{[1st Wong's eqn]}$$

$$DQ/d\tau = 0 \qquad \text{[2nd Wong's eqn].}$$

Explanation: $F = dA + [A, A]$ is the curvature of the connection A. Q is a section of the co-adjoint bundle along γ. Recall that the adjoint bundle is

$$AdP = P \times_{Ad} \mathfrak{g}$$

where $\mathfrak{g}$ is the Lie algebra of G. It is the bundle whose sections are infinitesimal automorphisms of P. The co-adjoint bundle is the dual vector bundle,

$$Ad^*P = P \times_{Ad^*} \mathfrak{g}^*$$

where $\mathfrak{g}^*$ is the dual of $\mathfrak{g}$. The curvature F is a two-form on M with values in AdP. Pairing Q and $F(\dot{\gamma}, \cdot\,)$ yields a one-form along γ. Raise the indices of this one-form (using the metric on M) to form a vector along γ. This vector is denoted $Q \cdot F(\dot{\gamma}, \cdot\,)$, the right hand side of the first Wong's equation.

In the second Wong's equation D denotes the covariant derivative on Ad^*P. In a local trivialization $DQ/dt = dQ/dt + (a \cdot \dot{\gamma})^* Q$, the second term denoting the infinitesimal co-adjoint action of $(a \cdot \dot{\gamma}) \in \mathfrak{g}$ on $Q \in \mathfrak{g}^*$, where a is the local trivialization of the connection one-form A.

Suppose now that $\mathfrak{g}$ admits an adjoint invariant metric. This metric induces a fiber metric on AdP, which we use to identify AdP with Ad^*P. It also induces a Kaluza-Klein metric on P as described above. For $\tilde{\gamma}$ a curve in P, write

$$Q(\tau) = [\tilde{\gamma}(\tau), A\cdot d\tilde{\gamma}/d\tau],$$

a section of the adjoint bundle along $\gamma = \pi\tilde{\gamma}$. Q is essentially the projection of $d\tilde{\gamma}/d\tau$ onto the vertical.

Lemma 1. *Suppose that* $\mathfrak{g}$ *admits an adjoint invariant metric. Then the curve* γ *in* M *is the projection of a K-K geodesic* $\tilde{\gamma}$ *in* P *if and only if there is a section* Q *of the adjoint bundle along* γ *such that* (γ, Q) *satisfy Wong's equations.*

This is proved in Montgomery [1984], and in Kerner [1968]. We have called the equations "Wong's equations", after S.K. Wong who wrote them down in [1970] as the equations of motion for a classical particle in a Yang-Mills field. In the Abelian case ($G = U(1)$) they were first written down by Kaluza [1921] and Klein. There Wong's equations are the Lorentz equations for a particle in an electromagnetic field.

Theorem 1 can now be stated in the more general way:

Theorem 3.
The projection γ *of any solution* (γ, Q) *to Wong's equations which satisfies the boundary conditions* $\gamma(0) = x_0$, $\gamma(t_1) = x_1$, *is an extremal for Pines' problem. Every piecewise* C^1 *extremal for the problem of Pines' , or for the problem of Shapere and Wilczek, is either*

the projection γ *of a solution* (γ, Q) *to Wong's equations* [normal extremal],

or it satisfies the "singular Wong's equations":

$$Q\cdot F(\dot{\gamma}, \cdot\,) = 0, \qquad \nabla_{\dot{\gamma}}\dot{\gamma} \text{ not identically } 0$$
$$DQ/d\tau = 0, \qquad Q \neq 0, \qquad \text{[abnormal extremal].}$$

The terminology "normal" and "abnormal" extremal is due to Bliss [1946]. Consider the holonomy map $H: \Omega \to Aut(P_0, P_1) \cong G$, where Ω denotes the set of piecewise C^1 curves joining x_0 to x_1. A normal extremal γ is one for which $H[\gamma]$ is a regular value of H ($d_\gamma H$ has full rank) and an abnormal extremal is a critical point for H. Thus abnormal extremals are basically extremals at which the set $\Omega_{H_0} = H^{-1}(H_0)$ is not a manifold. More precisely, they are curves for which the tangent space (or cone) to Ω_{H_0} is smaller than $\ker d_\gamma H$.

As mentioned in the outline at the end of the first section, Shapere and Wilczek's equation (30) is Wong's first equation, hence they discovered a precursor of Theorem 3.

An example of abnormality. Take $P = R^2 \times S^1$, $M = R^2$, with a connection whose curvature is $F = \rho dx \wedge dy$, where $\rho \geq 0$, and with the support of ρ a bounded convex open subset U of the plane. Set $\Phi = \int_{R^2} \rho dxdy$, and suppose that $0 < \Phi < 2\pi$. Consider Pines' problem for loops based at $x_0 \in \partial U$, with holonomy $H_0 = e^{i\Phi}$. The minimizer is $\gamma = \partial U$. This does not satisfy Wong's equation, but does satisfy the singular Wong's equations, since $F = 0$ on ∂U. (If U is not convex, the extremal will be the boundary of its convex hull. Note that this example can easily be placed on a compact surface.)

We can eliminate the occurence of abnormal extremals by making assumptions on the curvature.

Lemma 2. *If the curvature is either fat or flat then there are no abnormal extremals.*

Proof. By definition of fat, $Q \cdot F(\dot{\gamma}, \cdot) = 0$ implies that either $\dot{\gamma} = 0$ or $Q = 0$.

To prove the lemma in the flat case, recall that there the holonomy depends only on the homotopy class of γ. Hence we are extremizing the length of γ subject only to a constraint on its homotopy class.

In the case of a normal extremal γ, we would hope to have γ a C^2 curve, and not just piecewise C^2.

Lemma 3. Regularity. *Every normal extremal is* C^2 *(and hence smooth by Lemma 2 and Theorem 3, if we assume the connection and base metric are smooth).*

Proof. This follows from the "Weierstrass-Erdman Corner condition", cor. 74.2 of Bliss [1946]. Let $f(c,\dot{c})$ denote the integrand of the "action" functional S, in eq. [3.4] below. (The curve c is a curve in P.) This condition states that at a corner of piecewise C^1 extremal c for S, both the right and left hand limits of $\partial f/\partial \dot{c}$ exist, and that they are equal. In our case $\partial f/\partial \dot{c}^{\mu} = g_{\mu\upsilon}\dot{\gamma}^{\upsilon}$ where μ and υ denote a horizontal indices. Thus the base curve γ is continuously differentiable. The first Wong's equation now implies that γ is C^2. Q.E.D.

Remark 2.1. Ge Zhong [1988, private communication] has provided a more modern proof of regularity. He uses the projected "energy" $^1/_2\int\|\dot{\gamma}\|^2$ instead of length, and works on the Sobolev space of H_1 curves. His proof requires that H_0 is a regular value of H (eg. the bundle is fat). From this hypothesis it follows that every extremal solves Wong's equations weakly. This, together with the Sobolev embedding theorems, proves regularity.

Regularity for length minimizers is then obtained by parameterizing such a minimizer by arc length, and then noting that the Cauchy-Schwartz inequality implies that a length minimizer is also an energy minimizer.

Remark 2.2. In case γ is a loop it does not follow from Lemma 3 that its derivatives at the endpoints match up, i.e. "boundary regularity" may fail. Examples with such boundary irregularity can be found for the Abelian case in plasma physics texts. One such is obtained by taking the curvature to be $x dx \wedge dy$, on R^2. The solutions to the Lorentz equations (Abelian Wong's equations) with this magnetic field are cycloids. However, if P is a homogeneous bundle with connection (eg. a Stieffel variety) one can prove that the derivatives must match up at the endpoints.

Proof of Theorem 1. Combine Theorem 3 and Lemma 1 to get the first part of the Theorem. To prove the converse, combine Lemma 3 and 2. Q.E.D.

§3. Proof of Theorem 3 (for Pines' Problem).

The isoper. problem is :

minimize the projected length: $\ell(c) = \int \|\pi_* \dot{c}\| dt$ [3.1],
subject to the constraint: c is horizontal.
and the 2 conditions $c(0) = p_0 \in P_0$, $c(1) = p_1 = H_0 p_0 \in P_1$.
We will use the method of Lagrange multipliers. The constraint can be written

$$c^*A = 0 \quad [3.2]$$

where c^*A is the pull-back of the $\mathfrak{g}$-valued connection one-form A. The multipliers will be *functions* $t \mapsto Q(t) \in \mathfrak{g}^*$. We begin by paraphrasing Bliss' [1946] discussion of the method of Lagrange multipliers from his ch. 7 (see Theorem 74.1 and its corollary).

Theorem [Bliss]. *Every piecewise* C^1 *curve which extremizes* ℓ *subject to the constraint* $c^*A = 0$ *must be a critical point of*

$$\lambda_0 \ell(c) - \int Q(t) \cdot c^*A \quad [3.3]$$

for some <u>*non-zero*</u> *value* (λ_0, Q) *of the Lagrange multipliers. The variations allowed in calculating the derivative are piecewise* C^1 *curves* $(c_\varepsilon, Q_\varepsilon)$ *with* c_ε *satisfying the end-point conditions. Those extremals for which* $\lambda_0 \neq 0$ *are called "normal", the extremals with* $\lambda_0 = 0$ *are called "abnormal".*

For alternative, less detailed, descriptions of this result see Courant and Hilbert vol 1. or Arnold et al [1988] p.33. To see why one must allow the possibility of abnormal extremals, consider the usual method of Lagrange multipliers: calculate the critical points of $S = \ell - Q \cdot G$ where G is the constraint function, in our case, $G(c) = c^*A$. In order to conclude that a critical point of S is an extremal for the constrained variational problem, we need to know that 0 is a regular value of G. Otherwise, there may be a vector v in the kernel of d_cG which is not the derivative of any curve lying in the constraint set $\{G = 0\}$, but with $d\ell_c \cdot v \neq 0$. And this can happen even if c is a global minimum to the constrained problem.

In any case, in order to calculate the normal extremals we need only calculate the Euler-Lagrange equations for

$$S(c,Q) = \ell(c) - \int Q(t) \cdot c^*A \qquad [3.4].$$

(This Lagrangian occurs in the physics literature, with Q having the interpretation of a δ-function current.) The variation with respect to Q gives us back the constraint. We will split up the variations of c into vertical and horizontal variations. Vertical variations can be written

$$c_\varepsilon(t) = \eta_\varepsilon(c(t)) = c(t)\exp\varepsilon\xi(c(t))$$

where ξ represents a section of the adjoint bundle, and η_ε its exponential, which is an automorphism of P. ξ must satisfy the boundary conditions $\xi(c(0)) = \xi(c(t_1)) = 0$. The projected length is independent of vertical variations. We compute

$$\begin{aligned}\int \delta S/\delta c^V \cdot \xi &= -d/d\varepsilon \Big|_{\varepsilon=0} \int Q \cdot c^* \eta_\varepsilon{}^* A \\ &= -\int Q \cdot c^* D\xi \\ &= -\int (Q \cdot D\xi/dt)dt \\ &= +\int (DQ/dt \cdot \xi)dt\end{aligned}$$

here we have used the well-known fact that the Lie derivative $L_\xi A$ equals $D\xi$, where D denotes the covariant derivative with respect to A: $D\xi/dt = d\xi/dt + [A(c)\cdot\dot{c}, \xi]$, and $DQ/dt = dQ/dt + ad_{A(c)\cdot\dot{c}}{}^*Q$. Thus the vertical variation is given by

$$\delta S/\delta c^V = DQ/dt \qquad [3.5].$$

The horizontal variation is calculated similarly. Let $\gamma = \pi \circ c$, and let γ_ε be a variation of γ with derivative $\delta\gamma = d\gamma_\varepsilon/d\varepsilon\,|_{\varepsilon=0}$, a tangent vector along γ. Let $\delta\gamma^h$ denote the horizontal lift of an extension of $\delta\gamma$ to a vector field, and let Ψ_ε denote the local flow of $\delta\gamma^h$. Then

$$\int \delta S/\delta c^h \cdot \delta\gamma^h = d/d\varepsilon\,|_{\varepsilon=0}\, S(c_\varepsilon\,,\, Q), \text{ where } c_\varepsilon = \Psi_\varepsilon \circ c$$

The derivative of the length functional is well known:

$$d\ell(c_\varepsilon)/d\varepsilon\,|_{\varepsilon=0} = -\int \langle \|\dot\gamma\|^{-1}\nabla_{\dot\gamma}\dot\gamma, \delta\gamma\rangle dt$$

One calculates

$$\begin{aligned} d/d\varepsilon\,|_{\varepsilon=0}\, c^*\Psi_\varepsilon{}^*A &= c^*L_{(\delta\gamma^h)}A \\ &= F(\delta\gamma^h, \dot c\,)dt \\ &= -F(\dot\gamma, \delta\gamma)dt \qquad [3.6] \end{aligned}$$

where in the final equality we are viewing F as a two-form with values in the adjoint bundle. Consequently, the derivative of the Lagrange multiplier term is

$$-d/d\varepsilon\,|_{\varepsilon=0} \int Q\cdot c^*\Psi_\varepsilon{}^*A = \int Q\cdot F(\dot\gamma, \delta\gamma)dt.$$

Therefore

$$\delta S/\delta c^h = -\|\dot\gamma\|^{-1}\nabla_{\dot\gamma}\dot\gamma + Q\cdot F(\dot\gamma, \cdot\,) \qquad [3.7].$$

Combining the three variational equationals, we see that we have proved that (γ, Q) must satisfy the equations:

$$\nabla_{\dot\gamma}\dot\gamma = \|\dot\gamma\| Q\cdot F(\dot\gamma, \cdot\,)\ ;\ DQ/dt = 0 \qquad [3.8]$$

Now $\|\dot{\gamma}\|$ is constant, as one can see by differentiating it, making use of [3.8] and of the skew-symmetry of F. Thus γ is the projection of a solution to Wong's equations. This proves the first part of Theorem 3.

To prove the rest of Theorem 3, we need only set $\lambda_0 = 0$ in Bliss's Theorem, and calculate the resulting Euler-Lagrange equations. We have already done the work. The equations are the singular Wong's equations.

Q.E.D.

§4. Solution to the Shapere-Wilczek Problem: Proof for Theorem 2 and that part of Theorem 3.

Variation of Holonomy. We will need the formula

$$dH[\gamma_\varepsilon]/d\varepsilon\Big|_{\varepsilon=0} = -H[\gamma]\oint U_\tau^{-1}F(\dot{\gamma},\delta\gamma)U_\tau d\tau \qquad [4.1]$$

for the derivative of the holonomy. Here γ_ε is a smooth deformation of the loop $\gamma = \gamma_0$, and $\delta\gamma = d\gamma_\varepsilon/d\varepsilon\big|_{\varepsilon=0}$. U_τ denotes the operation of parallel translation along γ to the fiber over $\gamma(\tau)$. F is the curvature. We are using matrix notation: $F(\dot{\gamma},\delta\gamma)$ is an element of the fiber of AdP over $\gamma(\tau)$, and conjugating it by U_τ^{-1} parallel translates this element back to the fiber over x_0.

[4.1] can be proved directly from formula [3.6]. We find the derivation suggested in fig. 2 more illuminating. Let $U_1 = U_\tau$, and U_2 parallel translation forward from $\gamma(\tau)$ to x_0. Then

$$H[\gamma] = U_2U_1$$

If $\delta\gamma$ is the deformation epicted in fig. 2 then

$$H[\gamma_\varepsilon] = U_2 P_\varepsilon U_1$$

where P_ε is parallel translation around the little loop C_ε based at $\gamma(\tau)$. It is well known that

P_ε is approximately $I - \varepsilon\Delta\tau F(\dot\gamma, \delta\gamma)$.

where we write $\gamma_\varepsilon = \gamma + \varepsilon\delta\gamma$ and the support of $\delta\gamma$ is the interval $[\tau, \tau+\Delta\tau]$. It follows that for such a deformation,

$H[\gamma_\varepsilon]$ is approximately $H[\gamma] - \varepsilon\Delta\tau U_2 F(\dot\gamma, \delta\gamma)U_1$

Now break an arbitrary deformation up into a sum of deformations with support $\Delta\tau$, take the limit as $\Delta\tau$ goes to 0, and differentiate with respect to ε:

$$dH[\gamma_\varepsilon]/d\varepsilon = -\oint U_2 F(\dot\gamma, \delta\gamma) U_1 \, d\tau \qquad [4.2]$$

To obtain [4.1], write $U_2 = H[\gamma]U_1^{-1}$.

<u>Proof of Theorem 2.</u> Write $E(\gamma) = f(\chi(H[\gamma], e(\gamma))$ where f is a function of two real variables x and y, and where $e(\gamma) = {}^1/_2\int \|\dot\gamma\|^2$. Then

$$dE(\gamma_\varepsilon)/d\varepsilon\big|_{\varepsilon=0} = (\partial f/\partial x)d\chi\cdot dH/d\varepsilon + (\partial f/\partial y)de/d\varepsilon$$

Now $d\chi$ is a one-form on G which is G-invariant, since χ is a class function. We have

$$d\chi\cdot dH/d\varepsilon = \langle d\chi, -H\oint U^{-1}{}_\tau F(\dot\gamma, \delta\gamma)U_\tau d\tau\rangle$$
$$= \langle -H^{-1}d\chi, \oint U^{-1}{}_\tau F(\dot\gamma, \delta\gamma)U_\tau d\tau\rangle$$
$$= -\oint q(\tau)\cdot F(\dot\gamma, \delta\gamma)$$

Here $H^{-1}d\chi = q(0) \in g^* = T^*_I G$ denotes pull-back of the covector $d\chi$ to the identity by left multiplication, and $q(\tau) = U_\tau q(0) U_\tau^{-1}$ is the parallel translate of $q(0)$ along γ to the fiber of the co-adjoint bundle over $\gamma(\tau)$.

Thus

$$dE/d\varepsilon = \{\oint <-aq(\tau)\cdot F(\dot\gamma, \cdot) + b\nabla_{\dot\gamma}\dot\gamma, \delta\gamma> d\tau$$

where $a = \partial f/\partial x$, $b = \partial f/\partial y$ are constants depending on γ. The Euler Lagrange equations are again Wong's equations, provided $b \neq 0$. (They are the singular Wong's equations is $a = 0$, $b \neq 0$.) In the original case of Shapere and Wilczek, i.e. $f = x/y$, $b = -\chi/l^2$, so $b \neq 0$ means $\chi \neq 0$. Q.E.D.

§5. Examples: Extremal Curves on Projective Spaces and Grassmannians.

Hopf Fibrations.

For the Hopf fibration $S^{2n-1} \rightarrow CP^n$ all the extremals are "small" geometric circles sitting in a CP^1 in CP^{n-1}. To see this, fix $x_0 \in CP^n$. Then $\gamma_0 = \pi^{-1}(x_0)$ is a great circle in S^{2n-1}, obtained by intersecting a 2-plane P_0 (which also happens to be a complex line) with the sphere. Let $\tilde\gamma \neq \gamma_0$ be any great circle intersecting γ_0, and let P be the corresponding 2-plane. Then $\mathrm{Span}_C\{P+P_0\}$ is a complex 2-dimensional subspace of C^n. Its intersection with S^{2n-1} is a 3-sphere whose projection onto CP^n is a CP^1. Consequently we are reduced to the case $n = 1$, already covered in the beginning of this paper. The projected extremal curve $\gamma = \pi\tilde\gamma$ is a small circle on this CP^1. The holonomy of γ (and in fact any loop in CP^n) is

$$H[\gamma] = e^{i\Phi}, \text{ where } \Phi = \int_D \omega$$

with D any disc bounded by γ, and ω the curvature two-form, which is also the canonical symplectic form on CP^n, normalized so that $\int_{CP^1}\omega =$

2π.

Homogeneous Bundles. Consider a "tower" of bundles

$$G_1 \to G_1/K = P \to G_1/K\times G = M.$$

where G_1 is a compact group, containing $K\times G$ as a closed subgroup. We have in mind the Stieffel variety $P = V_{k,n}$ of k-frames in C^{k+n} which is a $G=U(k)$ bundle over the Grassmannian $G_{k,n}$ of k-planes in C^{k+n}. Here $G_1 = U(k+n)$ and $K = U(k)$.

Fix a bi-invariant metric on G_1. This induces metrics on P and M such that each projection is a Riemannian submersion, and each structure group acts by isometries. Thus the metrics on G_1 and P are Kaluza-Klein metrics.

Let $p_0 \in P_0$ denote the identity coset. In the $V_{k,n}$ case $p_0 = \{e_1, \ldots, e_k\}$, the frame consisting of the first k vectors of the standard basis. The projection $G_1 \to P$ is simply $g_1 \to g_1 \cdot p_0$. Recall that the geodesics through the identity of G_1 are the one-parameter subgroups, $\exp t\xi$, $\xi \in \mathfrak{g}_1$. Then according to remark 1.1, §1, any geodesic in P through p_0 is of the form

$$\exp t\xi \cdot p_0, \qquad \xi \in \mathfrak{k}^\perp \quad .$$

By our theorem then, the extremal paths on M, are also of this form, pushed down to M. We have thus proved

__Proposition 1__. *For the homogeneous bundle* $P \to M$ *the normal extremal loops through* x_0 *= identity coset of* M *are the paths of the form*

$\gamma(t) = \exp t\xi \cdot x_0$, *where*

(i) $\xi \in \mathfrak{k}^\perp$,

and where

(ii) ξ *is such that there exists a* $t_1 > 0$ *such that* $\exp t_1\xi \in G \times K$.

(This last condition insures that the path γ *is closed.) In particular all extremals are orbits of 1-parameter groups of isometries.*

The problem of finding all the normal extremal loops has been reduced to the Lie theoretic problem of finding all elements $\xi \in \mathcal{k}^{\perp}$ which satisfy property (ii) . Without loss of generality, we can normalize ξ by insisting that its projection onto $(\mathcal{g} \oplus \mathcal{k})^{\perp}$ has length 1. (If this length were 0 then $\gamma(t)$ would be identically x_0.) Then the smallest t_1 such that (ii) holds is the length of γ. The holonomy of γ can be calculated using formula [1.1]:

$$H[\gamma] = \mathbb{P}_G(\exp t_1 \xi \exp -t_1 Q_0),$$

where $\mathbb{P}_G$: $G \times K \rightarrow G$ is the projection, and where $Q_0 = \mathbb{P}_{\mathcal{g}} \xi$ is the orthogonal projection of ξ onto $\mathcal{g}$.

<u>The Complex Grassmanian of Two-planes in Three-space.</u>

This is the first non-trivial non-Abelian example. Our ultimate goal is to find all the solutions to Pines' problem for the Stieffel varieties, and this will hopefully be the subject of a subsequent paper. However, in this paper we will not find all minima, but rather a large class of extrema.

The relevant tower of bundles is

$$\begin{array}{cc} & U(3) \\ U(1) & \downarrow \\ & V_{2,1} \\ U(2) & \downarrow \\ & G_{2,1} \end{array}$$

The corresponding splitting of u(3) is

$$\left(\begin{array}{c|c} Q & -\overline{\mathbf{v}} \\ \hline \mathbf{v} & i\theta \end{array}\right)$$

with Q in u(2), **v** in $\mathbf{C}^2$, and θ real. According to proposition 1, we should take θ = 0, to insure that this matrix generates a geodesic in $V_{2,1}$. The vector **v** represents a tangent vector to $G_{2,1}$. Without loss of generality, we can assume $\|\mathbf{v}\| = 1$, and we can take $\mathbf{v} = (0,1)$ by using the isotropy representation of U(2)×U(1) on $T_{X_0}G_{2,1}$. We are then reduced to studying the exponentials of matrices ξ of the form

$$\left(\begin{array}{cc|c} \multicolumn{2}{c|}{\multirow{2}{*}{Q}} & 0 \\ & & -1 \\ \hline 0 & 1 & 0 \end{array}\right)$$

There are two distinguished classes of ξ's:

$$Q = \begin{matrix} 0 & 0 \\ 0 & i\alpha \end{matrix} \qquad \text{case (I)}$$

and

$$Q = \begin{matrix} 0 & -b \\ b & 0 \end{matrix} \quad , \text{ b real} \qquad\qquad \text{case (II).}$$

(I) generates small geometric circles lying on the $CP^1 \subset CP^2$ which contains x_0 and **v**. (II) generates small geometric circles lying on the $RP^2 \subset CP^2$. The corresponding holonomies, as measured from p_0 = the frame {(1,0,0) , (0,1,0) }, are:

$$H[\gamma] = \begin{matrix} 1 & 0 \\ 0 & e^{i\Phi} \end{matrix} \qquad\qquad \text{for (I)}$$

with $\Phi = \pi\{1 - (\alpha/2)/\sqrt{1+(\alpha/2)^2}\} = \pi(1-\cos\varphi)$, of length $t_1 = \pi/\sqrt{1+(\alpha/2)^2} = \pi\sin\varphi$, with φ as in figure 3.

$$H[\gamma] = \begin{matrix} \cos\Phi & \sin\Phi \\ -\sin\Phi & \cos\Phi \end{matrix} \qquad\qquad \text{for (I)}$$

$\Phi = 2\pi\{1 - b/\sqrt{1+b^2}\} = 2\pi(1-\cos\varphi)$, of length $t_1 = 2\pi/\sqrt{1+b^2} = 2\pi\sin\varphi$ with φ as in figure 3. These lengths and holonomies can be obtained by using the procedure discussed in the paragraph following the statement of the Theorem 5.

A more geometric way of obtaining these results is to note that the corresponding ξ's lie in the Lie algebras for the subgroups

$$U(2) = \begin{matrix} 1 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\multirow{2}{*}{U(2)}} \\ 0 & & \end{matrix} \qquad\qquad \text{for (I)}$$

and

$$SO(3) \qquad\qquad \text{for (II).}$$

$$Q = \begin{matrix} 0 & -\overline{\beta} \\ \beta & 0 \end{matrix} \quad , \ \beta \text{ complex} \qquad \text{(IIa).}$$

Now this Q is obtained from the Q of the form (II) by conjugating the latter by the matrix

$$R_\theta = \begin{matrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{matrix}$$

where $\theta = \arg\sqrt{\beta}$. $R_\theta \oplus Id$ leaves (x_0, V) fixed. The extremal curve corresponding to Q as in (IIa) is obtained by simply conjugating the one for (II) by this $R_\theta \oplus Id$, and its holonomy is obtained by conjugating the holonomy for (II) by R_θ. The extremal curve lies on $RP^2{}_\theta := R_\theta \oplus Id(RP^2)$.

Holonomy around geodesics.

The geodesic through x_0 in the direction v is $\exp t\xi \cdot x_0$ with Q = 0 in the expression for ξ. This lies in both the CP^1 and the RP^2. Now a closed geodesic on a CP^1 of radius r_1 has length $2\pi r_1$, whereas a closed geodesic on an RP^2 of radius r_2 (i.e. its covering sphere has radius r_2 in R^3) has length πr_2. Since our geodesic lies on both CP^1 and RP^2 we must have $2r_1 = r_2$. Cconsequently, as is well known, the sectional curvature of the CP^1 is 4 times that of the RP^2. We have normalized $r_1 = {}^1/_2$. The holonomy of the geodesic is

$$H[\gamma] = \begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix} \ .$$

This is obtained by setting $\alpha = 0$ in the expression for the holonomy in case (I). (Setting b = 0 for case (II) does not work as the consequent curve is twice a closed geodesic on RP^2.)

Some Remarks on Finding all Minimal Loops.

4.1. We conjecture that there are no abnormal extremals in the case where the bundle in question is a Steiffel variety. We also conjecture that Lemma 2 is still valid if the hypothesis of fatness is replaced by the hypothesis that the curvature has full-rank as map $\Lambda^2 T \to AdP$. The first conjecture is valid if the second one is.

<u>4.2</u>. One calculates

$$Q \cdot F(\dot{\gamma}, \cdot) = (\beta, -i\alpha)$$

where

$$Q = \begin{matrix} i\delta & -\bar{\beta} \\ \beta & i\alpha. \end{matrix} \qquad \text{and } \dot{\gamma} = v = (0,1) \in T_{X_0}G_{2,1}.$$

We have discussed two extremes:

[I]: $\beta = 0$, $\delta = 0$. Then $\mathrm{span}\{\dot{\gamma}, \nabla_{\dot{\gamma}}\dot{\gamma}\}$ is the tangent space to CP^1 at $\gamma(t)$.

[II]: $\alpha = 0$, $\delta = 0$. Then $\mathrm{span}\{\dot{\gamma}, \nabla_{\dot{\gamma}}\dot{\gamma}\}$ is the tangent space to RP^2 at $\gamma(t)$.

The next simplest case is

$$Q = \begin{matrix} 0 & -\bar{\beta} \\ \beta & i\alpha. \end{matrix} \qquad \beta \neq 0, \ \alpha \neq 0$$

in which case $\mathrm{span}\{\dot{\gamma}, \nabla_{\dot{\gamma}}\dot{\gamma}\}$ is neither the tangent space to CP^1 or RP^2. Exponentiating the resulting matrices (0 is always an eigenvalue) and thus finding the corresponding extremals and their holonomies is straightforward, but tedious.

<u>4.3.</u> Exponentiating the general $\xi \in U(3)$ requires finding the eigenvalues of ξ, which in turn involves solving a cubic equation. This can of course be done exactly, so in theory we can find all solutions to Pines' problem for the case $(k,n) = (2,1)$.

The first question which comes up is (ii) of Theorem 4: for which ξ does $\exp t\xi$ eventually return to the subgroup $U(2)xU(1)$? Let $i\lambda_1$, $i\lambda_2$, $i\lambda_3$ denote the eigenvalues of ξ. A sufficient condition for return is that the difference vector $(\lambda_3 - \lambda_1, \lambda_2 - \lambda_1) = c(p,q)$, for some integers p,q. In other words, this vector should lie on an integer line. For if this is the case then $\exp t_1\xi = e^{i\mu}Id.$, where $\mu = 2\pi\lambda_1/c$, and $t_1 = 2\pi/c$. (This condition is not neccessary, as one can see by taking Q diagonal.) This shows that there are a dense set of Q's such that the extremal paths $\exp t\xi$ do in fact close.

Problem 4.4. There will be extremal loops which are not minimal. Find an efficient way of eliminating these. Perhaps this can be done by an "isoperimetric inequality" or by a clever calculation of the constrained second variation for Pines' problem.

§5. More Open Problems.

1. Extend Our Results to Surfaces.

The projected Wong solutions $\gamma \subset CP^2$ have constant geodesic curvature in the case of the Hopf fibration. Is the projection of every minimal Legendrian surface $\Sigma \subset P = S^5$ a constant mean curvature Lagrangian surface in CP^2 ? (A Lagrangian surface is one which annihilates the curvature. A Legendrian surface is one which is everywhere horizontal.) This question may be related to recent work of J. Wolfson on minimal Lagrangian surfaces in CP^2.

2. Find a non-Abelian isoperimetric inequality.

Begin by requiring the bundle to be homogeneous, eg. one of the Steiffel varieties. One would like an inequality of the form

$$f(\chi(H[\gamma], l(\gamma)) \leq c.$$

In the Abelian cases, these inequalities are well-known. For example, for R^2 with curvature = area form = $d\Phi$, we have $\Phi/l^2 \leq 2\pi$. In CP^n we get $l^2 \geq 2\pi\Phi - \Phi^2$. This latter is obtained by using the trig identities together with the expressions:

$$\Phi = \pi(1-\cos\alpha)$$
$$l = \pi\sin\alpha$$

for the Pines' minima. (See fig. 3.)

Appendix with Alan Weinstein.
A heuristic geometric proof of Theorem 1.

Wong's equations with $Q \in Ad^*$ are independent of the scale of the bi-invariant metric on the fiber G. (For an interpretation of this fact in terms of Poisson reduction see Montgomery [1984].) Let this scale factor λ go to infinty. This forces the corresponding Kaluza-Klein "geodesics" to be horizontal. The projections of these "geodesics" are still (projections of) solutions to Wong's equations. The "proof" is concluded by noting that the length of a horizontal curve (w.r.t. any K-K metric) is equal to the length of its projection.

One hole in this proof is that it misses the abnormal extremals. To see this hands on, imagine the following construction of a minimizing sequence for Pines' problem. We are taking F compact, and M complete. Then all of the K-K metrics $g_\lambda = d^2s_M \oplus \lambda^2 d^2 s_{\mathfrak{g}}$ on P are complete. Let c_λ be a minimizing geodesic for g_λ which connects $p \in P_0$ to $q \in P_1$ where $H_0 p = q$.

By using the Sobolev embedding theorems and the weak compactness of the unit ball one can show that a subsequence of the c_λ converges in the Sobolev space H_1 to a horizontal curve c. In order to define the Sobolev norms one must fix λ, say $\lambda = 1$, with the resulting norms denoted "$\|\cdot\|_1$". The question now becomes: does the projection $\gamma = \pi \circ c$ satisfy Wong's equations? The vertical projections $Q_\lambda(t) \in \mathfrak{g}$ of the c_l are bounded in the $\lambda^2 d^2 s_{\mathfrak{g}}$ - norm as λ goes to infintiy: $\lambda \|Q_\lambda(t)\|_1 < C$, a fixed constant. Now the $\gamma_\lambda = \pi \circ c_\lambda$ satisfy Wong's equations:

$$\nabla_{\dot\gamma_\lambda} \dot\gamma_\lambda = \lambda^2 Q_\lambda \cdot F(\dot\gamma_\lambda, \cdot)$$

where the inner product of Q_λ with F is with respect to the $\lambda = 1$ metric on $\mathfrak{g}$. But we do not know that $\lambda^2 \|Q_\lambda(t)\|_1$ is bounded. We get γ satisfying Wong's equations, or the singular Wong's equations, depending on whether this quantity is bounded or not.

Acknowledgements. I am pleased to be able to acknowledge Alex Pines for formulating this problem, and for several useful discussions. I would also like to acknowledge useful conversations with Jeeva Anandan, Alan Weinstein, Bruce Kleiner, Eugene Lerman, and Ge Zhong. This work was supported by NSF grant DMS 8702502 and DMS 8807219.

References.

V.I. Arnold, V.V. Kozlov, A.I. Neishtadt, [1988], Dynamical Systems III. v. 3 in the Encyclopaedia of Mathematical Sciences series, Springer.

M. Berry, [1985], "Classical Adiabatic Angles and Quantal Adiabatic Phase", J. Phys. A, 18, 15-27.

G.A. Bliss, [1946], Lectures on Calculus of Variations. Univ. of Chicago Press.

R. Courant and D. Hilbert, [1953], Methods of Mathematical Physics, v.I, Interscience Pub., N.Y.

Th. Kaluza, [1921], "Zum Unitatsproblem der Physik" Berl. Berichte. p. 966.

R. Kerner, [1968],"Generalization of the Kaluza-Klein Theory for an Arbitrary non-Abelian Gauge Group", Ann. Inst. Henri Poincare. v.9, 143-152.

R. Montgomery, [1984], "Canonical Formulations of a Classical Particle in a Yang-Mills Field and Wong's Equations", Lett. Math. Phys., v. 8, 59-67.

A. Shapere and F. Wilczek, [1987], "Self-Propulsion at Low Reynolds Number", Phys. Rev. Lett., v.58, 2051-2054.

A. Weinstein, [1980], "Fat Bundles and Symplectic Manifolds", Adv. in Math. v. 37, 239-250.

S.K. Wong, [1970],"Field and Particle Equations for the Classical Yang-Mills Field and Particles with Isotopic Spin", Nuovo Cimento v.65A, 689-693.

Ge Zhong [1988], private communication.

R. Kerner, [1968],"Generalization of the Kaluza-Klein Theory for an Arbitrary non-Abelian Gauge Group", Ann. Inst. Henri Poincare. v.9, 143-152.

R. Montgomery, [1984], "Canonical Formulations of a Classical Particle in a Yang-Mills Field and Wong's Equations", Lett. Math. Phys., v. 8, 59-67.

A. Shapere and F. Wilczek, [1987], "Self-Propulsion at Low Reynolds Number", Phys. Rev. Lett., v.58, 2051-2054.

A. Weinstein, [1980], "Fat Bundles and Symplectic Manifolds", Adv. in Math. v. 37, 239-250.

S.K. Wong, [1970],"Field and Particle Equations for the Classical Yang-Mills Field and Particles with Isotopic Spin", Nuovo Cimento v.65A, 689-693.

Ge Zhong [1988], private communication.

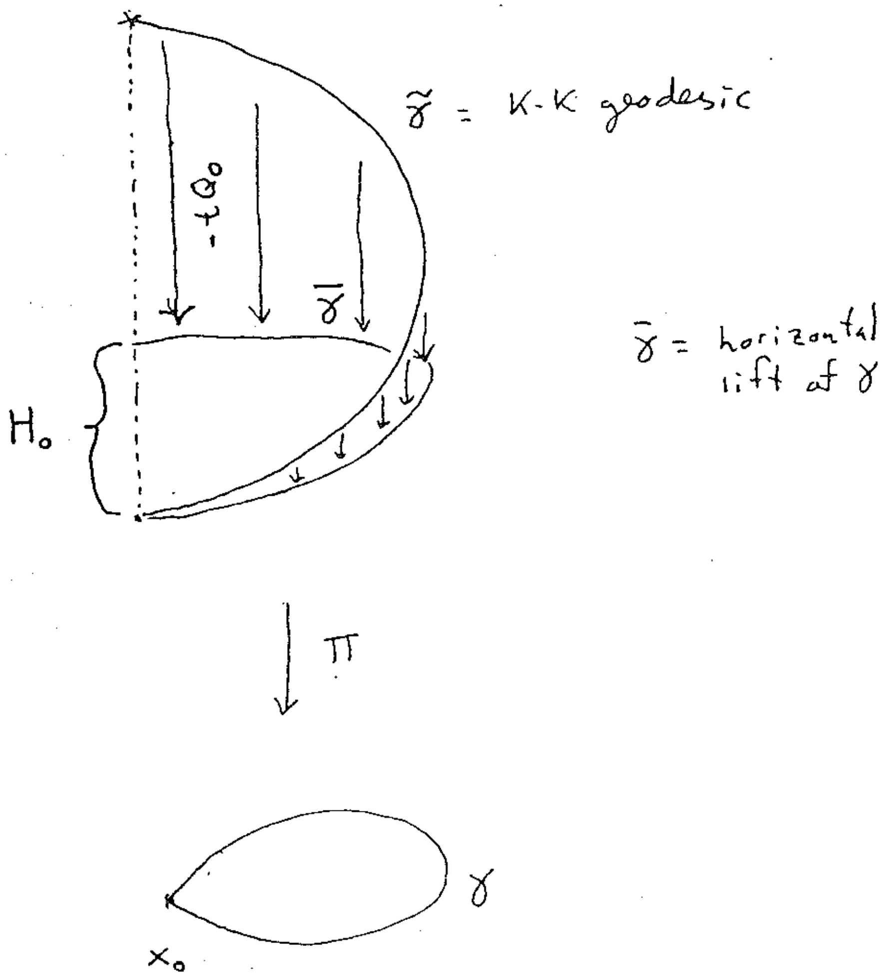
$\tilde{\gamma}$ = K-K geodesic
$-tQ_0$
$\bar{\gamma}$
H_0
$\bar{\gamma}$ = horizontal lift of γ
π
γ
x_0

Figure 1

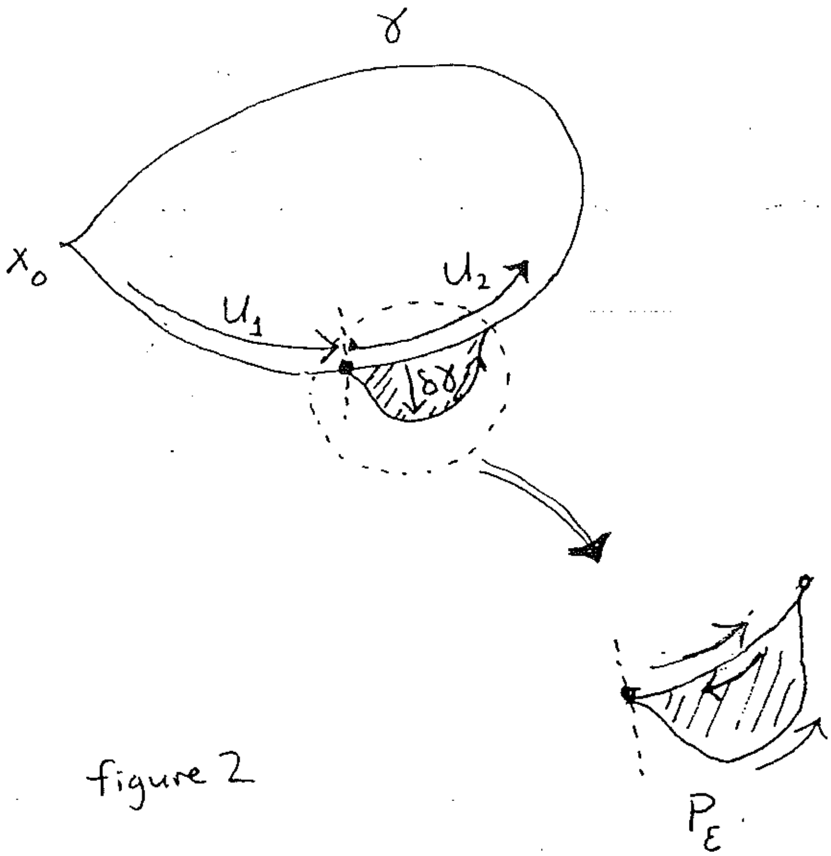

figure 2

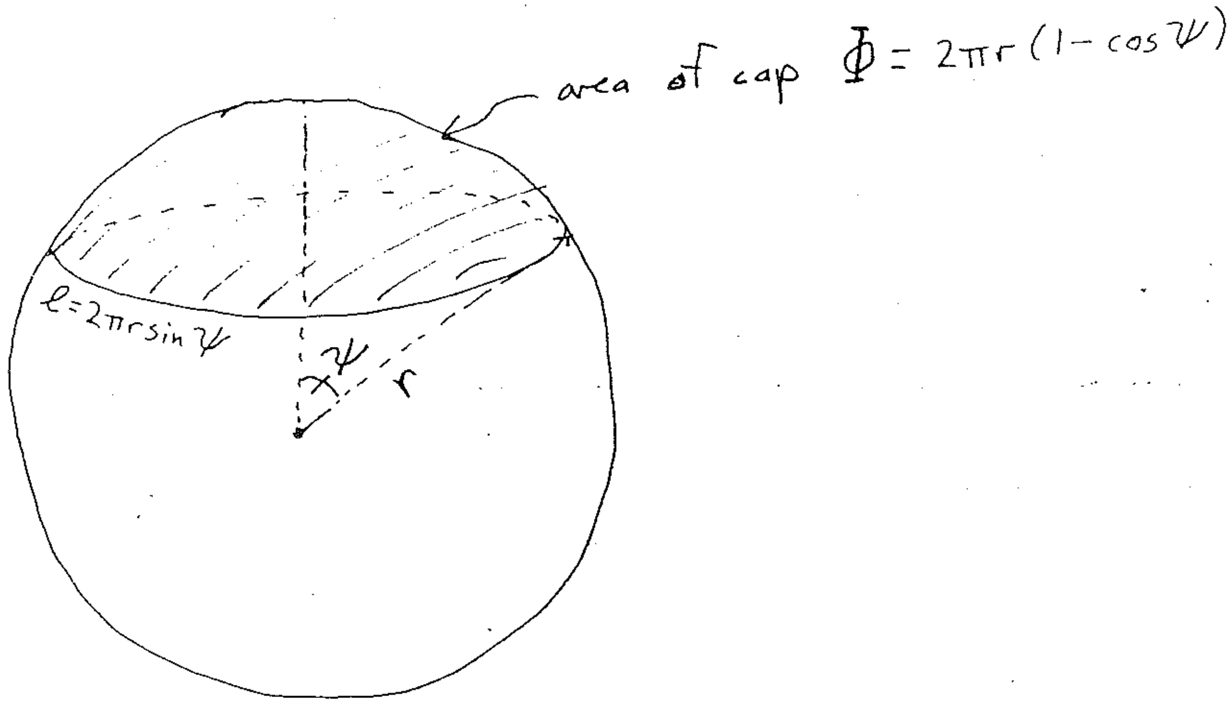

fig. 3

addendum.

The problem:

minimize the length of c (or $\pi(c)$)

subject to the constraint

$$A_{c(t)}\,\dot{c}(t) = \mu \in \mathfrak{g} \quad \text{a const.}$$

Has the same sol'n: Wong trajectories.

PF Use Lag. multipliers for

$$S = \int |\dot{c}| - \int_I Q\cdot\{c^*A - \mu\,dt\}$$

$\frac{\delta S}{\delta Q} \Rightarrow$ constraint

The horiz. & vert. var'ns are the same as already calculated for $\mu = 0$.

A tricky point is: $\frac{\delta\,\text{length}}{\delta\,\text{vert}} = 0$.

Also, do we use the projected, or the total, length (or energy).

Ans: It doesn't matter! From the constraint we get $|\dot{c}|^2 = |\pi_*\dot{c}|^2 + \mu^2$.

(This seems to leave open the q: do we apply the constraint before or after the vertical var'n?) The reason this works is because minimizing f rel the constraint $g = 0$ & " f' " " " " are equivalent, provided $f|_C = f'|_C$, $C = \{g = 0\}$

Note $f' = f + F(g)$ so the Lag mult λ changes by $F'(g)|_{g=0}$.)