

# A variational problem on Stiefel manifolds

Anthony M Bloch<sup>1</sup>, Peter E Crouch<sup>2</sup> and Amit K Sanyal<sup>3</sup>

<sup>1</sup> Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>2</sup> Department of Electrical Engineering, Arizona State University, Tempe, AZ 85281, USA

<sup>3</sup> Department of Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ 85281, USA

E-mail: [abloch@math.lsa.umich.edu](mailto:abloch@math.lsa.umich.edu), [peter.crouch@asu.edu](mailto:peter.crouch@asu.edu) and [sanyal@asu.edu](mailto:sanyal@asu.edu)

Received 3 November 2005, in final form 25 July 2006

Published 25 August 2006

Online at [stacks.iop.org/Non/19/2247](http://stacks.iop.org/Non/19/2247)

Recommended by E S Titi

## Abstract

In their paper on discrete analogues of some classical systems such as the rigid body and the geodesic flow on an ellipsoid, Moser and Veselov introduced their analysis in the general context of flows on Stiefel manifolds. We consider here a general class of continuous time, quadratic cost, optimal control problems on Stiefel manifolds, which in the extreme dimensions again yield these classical physical geodesic flows. We have already shown that this optimal control setting gives a new symmetric representation of the rigid body flow and in this paper we extend this representation to the geodesic flow on the ellipsoid and the more general Stiefel manifold case. The metric we choose on the Stiefel manifolds is the same as that used in the symmetric representation of the rigid body flow and that used by Moser and Veselov. In the extreme cases of the ellipsoid and the rigid body, the geodesic flows are known to be integrable. We obtain the extremal flows using both variational and optimal control approaches and elucidate the structure of the flows on general Stiefel manifolds.

Mathematics Subject Classification: 49Q99, 34K35, 65K10

## 1. Introduction

This paper presents a variational problem on the Stiefel manifold of orthogonal  $n$  frames in  $N$ -dimensional real Euclidean space and its corresponding optimal control counterpart. Solutions to the variational and optimal control problems are obtained and some of their geometric and analytic properties studied. This is an extension of earlier work by Bloch *et al* (2002) on the symmetric representation of the rigid body equations (which correspond to the extreme case of  $n = N$ ) on the Cartesian product  $SO(N) \times SO(N)$ . We characterize the space of solutions of the optimal control problem and the nature of the geodesic flows on the Stiefel

manifold. The discrete version of this problem has been analysed in the seminal work by Moser and Veselov (1991). Theorem 4 of Moser and Veselov (1991) gives a set of isospectral deformations for the discrete geodesic flow, which can be viewed as a discrete analogue of the parameter-dependent Lax representation.

Recent results by Bolsinov and Jovanovic (2004) demonstrate that bi-invariant geodesic flows on Stiefel manifolds are integrable for a  $SO(N)$ -invariant metric. However, integrability has not yet been demonstrated for geodesic flows on general Stiefel manifolds with left-invariant metrics.

We present a generalization of the Lax pair form for the equations of motion and show how this reduces to the classical Lax pair form in the case of the rigid body equations and the geodesic flow on the ellipsoid. The integrability of the rigid body equations by a Lax pair formulation with parameter had been shown by Manakov (1976); Mischenko and Fomenko (1978) showed that a similar formalism exists for any semisimple Lie group. Further references on parameter-dependent Lax pair formulations for integrable systems are given in Fedorov (1995). The geodesic flow on an ellipsoid, which corresponds to the  $n = 1$  case, and its integrability have been treated by Moser and Veselov (1991), Knorrer (1980) and others. The paper by Moser and Veselov (1991) on the discrete variational version of this problem also gave a discrete Lax pair formulation with parameter, thereby demonstrating that the discrete geodesic equations on the Stiefel manifold are indeed integrable.

We describe here how to obtain geodesic flows using the maximum principle of optimal control theory (see Bloch *et al* (2003), Gelfand and Fomin (2000) and Kirk (2004)). We show how to relate this optimal control formulation to the form naturally derived from variational calculus. We also relate these extremal flows to the Hamiltonian flow using the natural symplectic structure on the cotangent bundle of the Stiefel manifold. The extremal flows obtained here for the Stiefel manifolds are similar to the Hamiltonian flows on the ‘extended Stiefel varieties’ as described in Federov (2005). Finally, we demonstrate that the natural symplectic manifold carrying the extremals of the optimal control problem is symplectomorphic to the cotangent bundle of the Stiefel manifold. In the next section, we pose our problem on a general Stiefel manifold and give the extremal flows obtained in the limiting cases of the sphere/ellipsoid ( $n = 1$ ) and the  $N$ -dimensional rigid body ( $n = N$ ). Section 3 presents the extremal solution to the variational problem for the general case and connects this solution to the extremal flow on the cotangent bundle, given by equation (46). Section 4 gives the extremal solution of the optimal control problem restricted to a symplectic submanifold of  $R^{nN} \times R^{nN}$  that has the dimension of the cotangent bundle. This section also gives the correspondence between this extremal solution and the extremal solution of the variational problem. Section 5 presents the structure of the tangent and cotangent bundles of the Stiefel manifold and establishes a symplectomorphism between the manifold carrying the extremal solutions of the optimal control problem and the cotangent bundle. Section 6 presents a few applications and some future research issues of interest regarding geodesic flows on Stiefel manifolds, while section 7 presents some concluding remarks.

## 2. Background and limiting cases

We introduce the variational and optimal control problems on a Stiefel manifold in this section, based on minimizing the time integral of the kinetic energy. The metric on the manifold is given by the kinetic energy expression. We also give the extremal flows obtained in the limiting cases of the sphere/ellipsoid ( $n = 1$ ) and the  $N$ -dimensional rigid body ( $n = N$ ). The extremal flows in these cases are well known and integrable and have been given in several earlier works such as Knorrer (1982), Moser (1980) and Bloch *et al* (2002).

2.1. Variational and optimal control problems on a Stiefel manifold

The Stiefel manifold  $V(n, N) \subset \mathbb{R}^{nN}$  consists of orthogonal  $n$  frames in  $N$ -dimensional real Euclidean space,

$$V(n, N) = \{Q \in \mathbb{R}^{nN}; \quad QQ^T = I_n\}.$$

Introduce the pairing in  $\mathbb{R}^{rs}$  given by

$$\langle A, B \rangle = \text{Tr}(A^T B), \tag{1}$$

where  $\text{Tr}(\cdot)$  denotes trace of a matrix and the left-invariant metric on  $\mathbb{R}^{nN}$  given by

$$\langle\langle W_1, W_2 \rangle\rangle = \langle W_1 \Lambda, W_2 \rangle = \langle W_1, W_2 \Lambda \rangle, \tag{2}$$

where  $\Lambda$  is a positive definite  $N \times N$  diagonal matrix. The pairing (1) was used in Ratiu (1980) as a positive definite bilinear form on  $\mathfrak{so}(n)$ . We are interested in the variational problem given by

$$\min_{Q(\cdot)} \int_0^T \frac{1}{2} \langle\langle \dot{Q}, \dot{Q} \rangle\rangle dt, \tag{3}$$

subject to  $QQ^T = I_n, Q \in \mathbb{R}^{nN}, 1 \leq n \leq N, Q(0) = Q_0, Q(T) = Q_T, I_n$  denotes the  $n \times n$  identity matrix. This is a variational problem defined on the Stiefel manifold  $V(n, N)$ . The dimension of this manifold is given by

$$\text{Dim } V(n, N) = nN - \frac{n(n+1)}{2} = n(N-n) + \frac{n(n-1)}{2}.$$

The corresponding optimal control problem is given by

$$\min_{U(\cdot)} \int_0^T \frac{1}{2} \langle\langle QU, QU \rangle\rangle dt, \tag{4}$$

subject to  $\dot{Q} = QU; QU^T = I_n, Q(0) = Q_0, Q(T) = Q_T$  where  $U \in \mathfrak{so}(N)$ . Note that the quantity to be minimized is invariant with respect to the left action of  $SO(n)$  on  $V(n, N)$  since the metric (2) is left invariant.

2.2. The rigid body equations

For the special case when  $n = N, V(N, N) \equiv SO(N)$  and the extremal trajectories of the optimal control problem (4) give the  $N$ -dimensional rigid body equations. The usual system of rigid body equations on  $T^*SO(N)$  are

$$\dot{Q} = QU, \quad \dot{M} = [M, U], \tag{5}$$

where  $M = J(U) \triangleq \Lambda U + U \Lambda$  is the body momentum and  $[\cdot, \cdot]$  denotes the matrix commutator. Equations (5) can also be obtained directly from the variational problem (3) in the case  $n = N$ . As shown by Manakov (1976) and Ratiu (1980), the Euler equations are integrable with a parameter-dependent Lax pair representation

$$\frac{d}{dt}(M + \lambda \Lambda^2) = [M + \lambda \Lambda^2, U + \lambda \Lambda].$$

In the optimal control approach, the costate  $P$  is a vector of Lagrange multipliers used to enforce the equality constraint  $\dot{Q} = QU$  (see Gelfand and Fomin (2000)). The extremal trajectories for the optimal control problem in this case are given by

$$\dot{Q} = QU, \quad \dot{P} = PU, \tag{6}$$

where  $U \in \mathfrak{so}(N)$  and  $[\cdot, \cdot]$  denotes the matrix commutator. In the symmetric representation of the rigid body equations given in Bloch *et al* (2002), the states  $Q$  and the costates  $P$  are both orthogonal  $N \times N$  matrices, and the extremal trajectories are on  $SO(N) \times SO(N)$ . For the symmetric representation of the rigid body equations (6),  $U$  is regarded as a function of  $Q$  and  $P$  via the equations

$$U \triangleq J^{-1}(M) \in \mathfrak{so}(N) \quad \text{and} \quad M \triangleq Q^T P - P^T Q.$$

The inverse of the mapping  $\Phi : (Q, P) \mapsto (Q, M) = (Q, Q^T P - P^T Q)$  is given by

$$P = Q(e^{\sinh^{-1}(M/2)}), \quad (7)$$

where  $\sinh: \mathfrak{so}(N) \rightarrow \mathfrak{so}(N)$  and its inverse are defined in Bloch *et al* (2002). The spatial momentum given by

$$m = QM Q^T = P Q^T - Q P^T$$

is conserved along the flow given by (6). Equivalence between (6) and (5) was established by Bloch *et al* (2002) on the sets  $S$  and  $S_M$  where

$$S = \{(Q, P) \in SO(N) \times SO(N); \quad \|M\|_{OP} < 2\},$$

$$S_M = \{(Q, M) \in SO(N) \times \mathfrak{so}(N); \quad \|M\|_{OP} < 2\},$$

and  $\|A\|_{OP} = \sup\{\|Ax\| : \|x\| = 1\}$  is the operator norm.  $\Phi^{-1}$  is then well defined on  $S_M$ .

Note that for the extremal flow to the optimal control problem given by (6), one need not choose the costates to be orthogonal. One can choose a costate vector  $P_0$  such that  $Q^T P_0$  is skew symmetric, for example. In that case, the extremal flow is given by

$$\dot{Q} = QJ^{-1}(Q^T P_0), \quad \dot{P}_0 = P_0 J^{-1}(Q^T P_0), \quad (8)$$

and these equations restrict to the invariant submanifolds defined by  $Q^T P_0 \in \mathfrak{so}(N)$ . These are the McLachlan–Scovel equations (see McLachlan and Scovel (1995)). Comparing these equations with (6), we see that  $P_0 = QM = P - QP^T Q$ . In section 4, we obtain a generalization of the McLachlan–Scovel equations to the Stiefel manifolds  $V(n, N)$ , where  $1 < n < N$ .

### 2.3. Geodesic flow on the ellipsoid

For the other extreme case, when  $n = 1$ , we obtain the equations for the geodesic flow on the sphere  $V(1, N) \equiv \mathbb{S}^{N-1}$  with  $Q = q^T, q^T q = 1$ . This can be also be regarded as the geodesic flow on the ellipsoid

$$\bar{q}^T \Lambda^{-1} \bar{q} = 1,$$

where  $q = \Lambda^{-1/2} \bar{q}$ . The costate variable  $P = p^T$  is used to enforce the constraint  $\dot{q} = -Uq$  for the optimal control problem (4) when  $n = 1$ . The extremal solutions to this optimal control problem are

$$\dot{q} = -Uq, \quad \dot{p} = -Up + Aq, \quad (9)$$

where  $A = qq^T U \Lambda U - U \Lambda U qq^T$ . These extremal solutions have the same form as the extremal solutions for the general case (when  $1 < n < N$ ) given in section 3. The body momentum is obtained as

$$M = qp^T - pq^T, \quad (10)$$

in terms of the solution  $(q, p)$  to the optimal control problem. Equations (9) can than be expressed in terms of the body momentum as

$$\dot{q} = -Uq, \quad \dot{M} = [M, U] - A. \quad (11)$$

The body momentum can also be expressed in terms of the solution to the variational problem  $(q, \dot{q})$ , which we write in the form  $(q, s)$ , where  $s = \Lambda \dot{q} = -\Lambda Uq$ , as

$$M = qs^T - sq^T = qq^T U \Lambda + \Lambda Uqq^T.$$

Pre- and post-multiplying the above expression with  $\Lambda^{-1}$ , we obtain

$$\Lambda^{-1} M \Lambda^{-1} = \Lambda^{-1} qq^T U + Uqq^T \Lambda^{-1}. \quad (12)$$

Since  $U$  is skew-symmetric,  $q^T Uq = 0$ , and post-multiplying both sides of equation (12) with the vector  $q$ , we get

$$\Lambda^{-1} M \Lambda^{-1} q = Uq(q^T \Lambda^{-1} q) \Leftrightarrow Uq = \frac{\Lambda^{-1} M \Lambda^{-1}}{q^T \Lambda^{-1} q} q. \quad (13)$$

Note that equation (13) specifies  $U$  upto the equivalence class

$$[U], V \in [U] \Leftrightarrow Uq = Vq.$$

Equation (13) gives our choice for  $U$  as

$$U = \frac{\Lambda^{-1} M \Lambda^{-1}}{q^T \Lambda^{-1} q}.$$

From this expression for  $U$ , we obtain

$$U \Lambda U = \frac{1}{(q^T \Lambda^{-1} q)^2} \Lambda^{-1} (p^T \Lambda^{-1} q (qp^T + pq^T) - (q^T \Lambda^{-1} q) pp^T - (p^T \Lambda^{-1} p) qq^T) \Lambda^{-1}$$

and

$$qq^T U \Lambda U = \frac{1}{(q^T \Lambda^{-1} q)^2} qq^T \Lambda^{-1} \Delta,$$

where

$$\Delta = (p^T \Lambda^{-1} q)^2 - (p^T \Lambda^{-1} p)(q^T \Lambda^{-1} q).$$

This gives us  $A$  in equation (9) as

$$A = [qq^T, U \Lambda U] = \frac{1}{(q^T \Lambda^{-1} q)^2} (qq^T \Lambda^{-1} - \Lambda^{-1} qq^T) \Delta. \quad (14)$$

Equations (9), (13) and (14) define extremal trajectories of the optimal control problem on  $V(1, N) = \mathbb{S}^{N-1}$ .

The Lagrangian (variational) formulation for this problem gives us the equations for the geodesic flow on the sphere. To obtain these equations, we take reduced variations (see Marsden and Ratiu, (1999)) on  $V(1, N) = \mathbb{S}^{N-1}$ . The equation of motion can be written as

$$\Lambda \ddot{q} = bq, \quad (15)$$

where  $b$  is a real scalar in this case. We determine  $b$  from the constraint  $q^T q = 1$ . Differentiating this constraint with respect to time twice, we get

$$\ddot{q}^T q + \dot{q}^T \dot{q} = 0.$$

Substituting for  $\ddot{q}$  from equation (15) into the above expression, we get

$$b = -\frac{\dot{q}^T \dot{q}}{q^T \Lambda^{-1} q}.$$

Thus, we get the Lagrangian (variational) equations for the geodesic flow on the sphere  $(\mathbb{S}^{N-1})$  as

$$\ddot{q} = -\frac{\dot{q}^T \dot{q}}{q^T \Lambda^{-1} q} \Lambda^{-1} q. \quad (16)$$

Integrability of these extremal flows were proven by Jacobi with relation to the Neumann problem of motion on the sphere with a quadratic potential, as shown by Knorrer (1982). A contemporary version of integrability of the geodesic flow on an ellipsoid was demonstrated by Moser (1980) using the theorem of Chasles and geometry of quadrics. We now obtain a Manakov Lax pair formulation for this system. In this case, we have

$$\Lambda U - U \Lambda = (M \Lambda^{-1} - \Lambda^{-1} M) / q^T \Lambda^{-1} q.$$

Using the body momentum equation in (11), we now get

$$\frac{d}{dt} M(\lambda) = [M(\lambda), U(\lambda)], \quad (17)$$

where

$$M(\lambda) = M + \lambda \Lambda + q q^T \lambda^{-1} \Delta / q^T \Lambda^{-1} q, \quad (18)$$

$$U(\lambda) = U - \lambda \Lambda^{-1} / q^T \Lambda^{-1} q. \quad (19)$$

From (17), we see that the coefficients of  $\lambda^j$  in the expansion of  $(1/2k) \text{tr}(M(\lambda))^k$  are conserved along the geodesic flow.

### 3. The variational problem

Now we look at the solution to the general problem (3) posed on the Stiefel manifold  $V(n, N)$  where  $1 \leq n \leq N$ . The Lagrangian for this variational problem is

$$\mathcal{L}(Q, \dot{Q}) = \frac{1}{2} \langle \dot{Q}, \dot{Q} \rangle = \frac{1}{2} \text{Tr}(\Lambda \dot{Q}^T \dot{Q}), \quad (20)$$

where  $Q \in V(n, N)$ , i.e.  $Q$  satisfies  $Q Q^T = I_n$ .

#### 3.1. Solution to the variational problem

The solution to the variational problem (3) posed on  $V(n, N)$  is given by the following result.

**Theorem 1.** *The variational problem (3) on the Stiefel manifold  $V(n, N)$  has the solution*

$$\dot{Q} = QU, \quad U \in \mathfrak{so}(N), \quad (21)$$

$$Q^T Q \dot{U} \Lambda + \Lambda \dot{U} Q^T Q = \Lambda U^2 Q^T Q - Q^T Q U^2 \Lambda. \quad (22)$$

**Proof.** We take reduced variations on the Stiefel manifold given by  $Q Q^T = I_n$ ; this ensures that the variation vector field is always locally tangent to the manifold. The reduced variations are given by

$$\delta Q = Q \Sigma, \quad \delta \dot{Q} = \dot{Q} \Sigma + Q \dot{\Sigma}, \quad (23)$$

where  $Q \Sigma \in T_Q V(n, N)$  and  $\Sigma$  is an  $N \times N$  skew-symmetric matrix. These variations have fixed end points, i.e.  $\Sigma(0) = \Sigma(T) = 0$ . The kinematic expression (21) is obtained easily from the constraint  $Q Q^T = I_n$ . On taking the first variation of the integral quantity in (3) and setting it to zero, we obtain

$$\int_0^T \{ \langle \delta \dot{Q} \Lambda, \dot{Q} \rangle + \langle \dot{Q} \Lambda, \delta \dot{Q} \rangle \} dt = 0.$$

On carrying out integration by parts using fixed end point variations, the above expression simplifies to

$$\begin{aligned}
 & - \int_0^T \{ \langle Q \Sigma \Lambda, \ddot{Q} \rangle + \langle \ddot{Q} \Lambda, Q \Sigma \rangle \} dt = 0 \\
 \Leftrightarrow & \int_0^T \langle \ddot{Q} \Lambda, Q \Sigma \rangle dt = 0 \\
 \Leftrightarrow & \text{Tr}(\Lambda \ddot{Q}^T Q \Sigma) = 0 \\
 \Leftrightarrow & \Lambda \ddot{Q}^T Q - Q^T \ddot{Q} \Lambda = 0, \tag{24}
 \end{aligned}$$

i.e.  $\Lambda \ddot{Q}^T Q$  is symmetric. Taking the time derivative of  $\dot{Q}$  in (21), we get

$$\ddot{Q} = QU^2 + Q\dot{U}.$$

Substituting this for  $\ddot{Q}$  in equation (24) gives the solution of the variational problem as given by equation (22).  $\square$

Note that the flow on the Stiefel manifold induced by equation (22) is invariant to variations within the equivalence class  $[U] \subset \mathfrak{so}(N)$ , defined by

$$V \in [U] \Leftrightarrow QV = QU. \tag{25}$$

Since  $QQ^T = I_n$ , the left-multiplication of equation (22) by  $Q$  gives

$$Q\dot{U}\Lambda Q^T + Q\Lambda\dot{U}Q^T = Q[\Lambda, U^2]Q^T. \tag{26}$$

Here  $W = \dot{U}\Lambda + \Lambda\dot{U} \in \mathfrak{so}(N)$  is a skew-symmetric matrix that also satisfies  $QWQ^T = Q[\Lambda, U^2]Q^T$ . Equation (26) is a statement of the conservation of the quantity

$$m = QU\Lambda Q^T + Q\Lambda U Q^T = SQ^T - QS^T, \tag{27}$$

where  $S = QU\Lambda = \dot{Q}\Lambda$ , along the flow of the vector field given by equations (21) and (22). Note that

$$\dot{m} = QWQ^T - Q[\Lambda, U^2]Q^T = 0$$

by equation (26). The quantity  $m$  is analogous to the spatial momentum of the rigid body, which is conserved along the flow of the rigid body equations. The equations (21) and (22) can also be written as

$$\dot{Q} = QU = S\Lambda^{-1}, \quad U \in \mathfrak{so}(N), \tag{28}$$

$$Q^T \dot{S} - \dot{S}^T Q = 0. \tag{29}$$

The above formulation of the variational problem, equations (28) and (29) in particular, does not give an implicit or explicit equation for the rate  $\dot{S}$ . To obtain such an expression, we need the following result.

**Proposition 1.** Define the linear map  $L_Q : \mathbb{R}^{nN} \rightarrow \mathbb{R}^{nN}$ , given by

$$L_Q(X) = Q^T X - X^T Q, \quad Q \in V(n, N). \tag{30}$$

If  $X \in \mathbb{R}^{nN}$  is in the kernel of this map, then  $X$  is of the form

$$X = BQ, \quad B = B^T. \tag{31}$$

**Proof.** If  $X$  is in the kernel of the  $L_Q$  defined by (30), then  $Q^T X = X^T Q$  and hence  $Q^T X$  is symmetric. Let us denote the rows of  $Q$  by the orthonormal vectors  $q_1^T, q_2^T, \dots, q_n^T$ . We complete a right-handed orthonormal basis of  $\mathbb{R}^N$  from this set of orthonormal vectors; the additions to this set are denoted by the vectors  $q_{n+1}^T, \dots, q_N^T$ . The vectors of this orthonormal basis are arranged in columns to form the orthogonal matrix

$$C = [q_1 \ q_2 \ \cdots \ q_n \ q_{n+1} \ \cdots \ q_N] \in SO(N).$$

Thus, on right-multiplication of  $Q$  by  $C$  we get

$$R = QC = [I_n \ 0]. \quad (32)$$

We express  $C^T$  in the block diagonal form

$$C^T = \begin{bmatrix} C_1 & C_3 \\ C_2 & C_4 \end{bmatrix},$$

where the diagonal blocks  $C_1$  and  $C_4$  are  $n \times n$  and  $(N-n) \times (N-n)$  matrices, respectively. Hence, from equation (32), we get

$$Q = [I_n \ 0]C^T = [C_1 \ C_3]. \quad (33)$$

We apply the same transformation to  $X \in \mathbb{R}^{nN}$  to get

$$Y = XC = [Y_1 \ Y_2], \quad (34)$$

where  $Y_1 \in \mathbb{R}^{mn}$  and  $Y_2 \in \mathbb{R}^{n(N-n)}$ . Since  $Q^T X = CR^T Y C^T$ , the relation  $Q^T X = X^T Q$  can be expressed as

$$R^T Y = Y^T R \Leftrightarrow Y_1 = Y_1^T = B, \quad Y_2 = 0. \quad (35)$$

Then we have

$$\begin{aligned} X &= [Y_1 \ 0] \begin{bmatrix} C_1 & C_3 \\ C_2 & C_4 \end{bmatrix} \\ &= [BC_1 \ BC_3] \\ &= B[C_1 \ C_3] = BQ, \end{aligned}$$

using the expression in equation (33) for  $Q$ . This gives us the result in equation (31), where  $B = Y_1 = Y_1^T$  is a symmetric  $n \times n$  matrix.  $\square$

Since from (29),  $\dot{S}$  is in the kernel of  $L_Q$ , using the above result we get

$$\dot{S} = BQ, \quad B = B^T. \quad (36)$$

The following lemma is necessary to obtain an expression for the flow of the quantity  $S$  and leads to an implicit equation for  $B$ .

**Lemma 1.** *If  $K$  is symmetric and positive definite and  $B$  is symmetric, the map  $J : \text{Sym}(n) \rightarrow \text{Sym}(n)$  given by  $J : B \mapsto KB + BK$  has kernel zero and is hence an isomorphism.*

**Proof.** Since  $B$  is symmetric, there exists an orthogonal matrix  $C$ , such that  $CBC^T = \Psi$ , where  $\Psi$  is a real diagonal matrix. Thus if  $KB + BK = 0$ , then

$$C^T K C \Psi + \Psi C^T K C = 0,$$

where  $C^T K C$  is a positive definite symmetric matrix. If  $e$  is an eigenvector of  $\widehat{K} = C^T K C$ , then

$$\widehat{K}e = \lambda e, \quad \lambda > 0,$$



and

$$\widehat{K}\Psi + \Psi\widehat{K} = 0 \Rightarrow \widehat{K}(\Psi e) + \lambda(\Psi e) = 0.$$

Hence,  $\Psi e$  is also an eigenvector of  $\widehat{K}$  with eigenvalue  $-\lambda < 0$ . But  $\widehat{K}$  is positive definite and so all its eigenvalues are strictly positive. Thus, we have a contradiction, unless  $\Psi = 0$  and hence  $B = 0$ .  $\square$

Here  $\text{Sym}(n)$  denotes the space of  $n \times n$  symmetric matrices. Now we state the main result of this subsection, which expresses the geodesic flows on the Stiefel manifold in terms of the quantities  $Q$  and  $S = QU\Lambda$ ; the pair  $(Q, S)$  can be used to parametrize the tangent bundle since  $S = \dot{Q}\Lambda^{-1}$ .

**Proposition 2.** *The geodesic flow on the Stiefel manifold given by the variational problem (3) is of the form*

$$\dot{Q} = S\Lambda^{-1}, \quad \dot{S} = BQ, \tag{37}$$

where

$$B = \widetilde{J}_Q^{-1}(E), \quad E = -2\dot{Q}\dot{Q}^T = -2S\Lambda^{-2}S^T, \tag{38}$$

and

$$\widetilde{J}_Q(B) \triangleq (Q\Lambda^{-1}Q^T)B + B(Q\Lambda^{-1}Q^T); \quad \widetilde{J}_Q : \text{Sym}(n) \rightarrow \text{Sym}(n). \tag{39}$$

**Proof.** The proof of this result makes use of the simple observation that

$$Q\Lambda^{-1}S^T = -QUQ^T$$

is skew-symmetric. We already know from proposition 1 and equation (29) that  $\dot{S} = BQ$  where  $B$  is a  $n \times n$  symmetric matrix. To obtain an expression for  $B$ , we take a time derivative of

$$Q\Lambda^{-1}S^T + S\Lambda^{-1}Q^T = 0$$

along the extremal trajectories. This gives us

$$\begin{aligned} Q\Lambda^{-1}Q^T\dot{B} + BQ\Lambda^{-1}\dot{Q}^T &= S\Lambda^{-1}U\dot{Q}^T - QU\Lambda^{-1}\dot{S}^T \\ \Rightarrow \widetilde{J}_Q(B) &= -2\dot{Q}\dot{Q}^T = -2S\Lambda^{-2}S^T, \end{aligned}$$

since  $S = QU\Lambda$ . Since  $Q\Lambda^{-1}Q^T$  is positive definite, we know from lemma 1 that  $\widetilde{J}_Q : \text{Sym}(n) \rightarrow \text{Sym}(n)$  is an isomorphism, and hence the inverse  $B = \widetilde{J}_Q^{-1}(-2QU^2Q^T)$  exists and is unique. This proves the proposition.  $\square$

Note that the quantity  $S = QU\Lambda$  depends on the equivalence class defined by (25); so equations (37)–(39) in terms of  $(Q, S)$  also have the advantage that they *uniquely* express the extremal flows, whereas  $U$  in equations (21) and (22) is not unique. Since  $S = \dot{Q}\Lambda^{-1}$ , the above proposition also expresses the extremal flow in terms of the tangent bundle pair  $(Q, \dot{Q}) \in TV(n, N)$ .

Now we give another method to obtain the above geodesic flow from the constrained Lagrangian

$$\mathcal{L}_c(Q, \dot{Q}) = \mathcal{L}(Q, \dot{Q}) + \frac{1}{2}\langle B, QQ^T - I_n \rangle = \frac{1}{2}\langle \langle \dot{Q}, \dot{Q} \rangle \rangle + \frac{1}{2}\langle B, QQ^T - I_n \rangle, \tag{40}$$

where  $B \in \text{Sym}(n)$  is a Lagrange multiplier matrix. We form the Hamiltonian for this problem by applying the Legendre transform to the Lagrangian (20), which gives the same result as applying the Legendre transform to the constrained Lagrangian (40). Just as  $TV(n, N)$  can be viewed as a submanifold of  $V(n, N) \times \mathbb{R}^{nN} \subset \mathbb{R}^{nN} \times \mathbb{R}^{nN}$ ,  $T^*V(n, N)$  can also via  $\langle \cdot, \cdot \rangle$  be viewed as a submanifold of  $V(n, N) \times \mathbb{R}^{nN} \subset \mathbb{R}^{nN} \times \mathbb{R}^{nN}$  and endowed with the symplectic structure induced from the canonical symplectic structure of  $\mathbb{R}^{nN} \times \mathbb{R}^{nN}$ .

**Proposition 3.** *The Legendre transform  $\mathbb{F}\mathcal{L} : TV(n, N) \rightarrow T^*V(n, N)$  is obtained from the fibre derivative of the Lagrangian  $\mathcal{L} : TV(n, N) \rightarrow \mathbb{R}$  as*

$$\mathbb{F}\mathcal{L}(Q, \dot{Q}) = S. \quad (41)$$

**Proof.** Let  $U, V \in \mathfrak{so}(N)$ , so that  $QU, QV \in T_QV(n, N)$ . The fibre derivative of the Lagrangian (20) is given by

$$\begin{aligned} \mathbb{F}\mathcal{L}(Q, \dot{Q})(\Gamma) &= \left. \frac{d}{ds} \right|_{s=0} \mathcal{L}(\dot{Q} + s\Gamma) \\ &= \left. \frac{d}{ds} \right|_{s=0} \frac{1}{2} \langle \dot{Q} + s\Gamma, \dot{Q} + s\Gamma \rangle \\ &= \frac{1}{2} \langle \Gamma, \dot{Q} \rangle + \frac{1}{2} \langle \dot{Q}, \Gamma \rangle = \langle \dot{Q}, \Gamma \rangle \\ \Leftrightarrow \mathbb{F}\mathcal{L}(Q, \dot{Q})(\Gamma) &= \langle QU\Lambda, \Gamma \rangle = \langle S, \Gamma \rangle, \end{aligned}$$

or by identifying  $T^*V(n, N)$  with  $TV(n, N)$  through  $\langle \cdot, \cdot \rangle$ ,  $\mathbb{F}\mathcal{L}(Q, \dot{Q}) = S$ . Thus, the map given by (41) is the Legendre transform  $\mathbb{F}\mathcal{L} : TV(n, N) \rightarrow T^*V(n, N)$ .  $\square$

Proposition 3 can be used to obtain the Hamiltonian from the constrained Lagrangian as follows:

$$\begin{aligned} H_c(Q, S) &= \langle \dot{Q}, S \rangle - \mathcal{L}_c(Q, \dot{Q}) \Big|_{\dot{Q}=\mathbb{F}\mathcal{L}^{-1}(S)} \\ &= \langle \langle S\Lambda^{-1}, S\Lambda^{-1} \rangle \rangle - \frac{1}{2} \langle \langle S\Lambda^{-1}, S\Lambda^{-1} \rangle \rangle - \frac{1}{2} \langle B, QQ^T - I_n \rangle \Big|_{\dot{Q}=S\Lambda^{-1}} \\ &= \frac{1}{2} \langle \langle S\Lambda^{-1}, S\Lambda^{-1} \rangle \rangle - \frac{1}{2} \langle B, QQ^T - I_n \rangle \Big|_{\dot{Q}=S\Lambda^{-1}} \\ &= \frac{1}{2} \text{Tr}(S^T S\Lambda^{-1}) - \frac{1}{2} \text{Tr}(BQQ^T - B). \end{aligned} \quad (42)$$

The extremal (geodesic) flows are then obtained from this Hamiltonian once the flow is restricted to  $V(n, N)$ , as

$$\begin{aligned} \dot{Q} &= \text{Grad}_S H_c(Q, S) = S\Lambda^{-1}, \\ \dot{S} &= -\text{Grad}_Q H_c(Q, S) = BQ, \end{aligned} \quad (43)$$

which are identical to equations (37) in proposition 2. Thus, we can parametrize  $T_Q^*V(n, N)$  by  $(Q, S)$ , and equations (37) then express the extremal flows in terms of this parametrization.

In the following subsection, we define a momentum quantity that generalizes the body momentum of the  $N$ -dimensional rigid body, and that can be expressed in terms of  $Q$  and  $S$ . This expression of the body momentum is then used to solve for the unique equivalence class  $[U]$ , which corresponds to a given extremal solution pair  $(Q, S)$  of the variational problem.

### 3.2. The momentum equation and its solution

Using equation (21), we can write the Lagrangian in the form

$$\mathcal{L}(Q, U) = -\frac{1}{2} \text{Tr}(\Lambda U Q^T Q U) = \frac{1}{2} \langle QU, QU\Lambda \rangle.$$

We define the (body) momentum as the  $U$ -gradient of the Lagrangian with respect to the pairing  $\langle \cdot, \cdot \rangle$  in (1):

$$M = 2\text{Grad}_U \mathcal{L}(Q, U) = Q^T Q U \Lambda + \Lambda U Q^T Q = Q^T S - S^T Q. \quad (44)$$

We call this the body momentum, since in the case  $n = N$  ( $N$ -dimensional rigid body), this quantity is the momentum expressed in the body coordinate frame. The body and spatial momenta quantities are related by

$$m = QMQ^T, \tag{45}$$

as can be verified using equations (27) and (44) and the constraint  $QQ^T = I_n$ . One can verify that with the body momentum defined as in (44), equation (22) is equivalent to

$$\dot{M} = [M, U] - A, \quad \text{where } A = Q^T S U + U S^T Q = [Q^T Q, U \Lambda U]. \tag{46}$$

We show how these equations generalize the classical rigid body in  $N$  dimensions, i.e. the case  $n = N$ . In this case, the Euler–Arnold equations are given by

$$\dot{Q} = QU, \quad \dot{M} = [M, U], \quad \text{where } M = U \Lambda + \Lambda U,$$

and  $Q \in SO(N)$ . From equation (29), we obtain in the case  $n = N$

$$\dot{U} \Lambda + \Lambda \dot{U} = \Lambda U^2 - U^2 \Lambda.$$

Taking a time derivative of  $M = U \Lambda + \Lambda U$  and substituting the above equation, we see that  $M$  satisfies the Euler–Arnold equation. This is also equivalent to the flow of  $S$  being represented by (36).

Now we present a solution of the algebraic equation (44), which we rewrite below as

$$M = Q^T S - S^T Q = Q^T Q U \Lambda + \Lambda U Q^T Q \triangleq J_Q(U), \tag{47}$$

where  $QQ^T = I$ ,  $U = -U^T$  and  $\Lambda > 0$  (diagonal and positive definite). The map  $J_Q : \mathfrak{so}(N) \rightarrow \mathfrak{so}(N)$  is defined by equation (47). We first present a few lemmas, which are necessary to prove the main result.

**Lemma 2.** *If  $K$  is symmetric and positive definite and  $X \in \mathfrak{so}(N)$ , the map  $J : \mathfrak{so}(n) \rightarrow \mathfrak{so}(n)$  given by  $J : X \mapsto KX + XK$  has kernel zero and is hence an isomorphism.*

**Proof.** The proof of this statement is similar to the proof of lemma 1. Since  $X$  is skew, there exists a unitary matrix  $L$ , i.e.  $L\bar{L}^T = \bar{L}^T L = I_N$ , such that  $LX\bar{L}^T = \iota\Sigma$ , where  $\Sigma$  is a real diagonal matrix. Thus if  $KX + XK = 0$ , then

$$LK\bar{L}^T\Sigma + \Sigma LK\bar{L}^T = 0,$$

where  $LK\bar{L}^T$  is a positive definite Hermitian matrix. If  $e$  is an eigenvector of  $\widehat{K} = LK\bar{L}^T$ , then

$$\widehat{K}e = \lambda e, \quad \lambda > 0,$$

and

$$\widehat{K}\Sigma + \Sigma\widehat{K} = 0 \Rightarrow \widehat{K}(\Sigma e) + \lambda(\Sigma e) = 0.$$

Hence,  $\Sigma e$  is also an eigenvector of  $\widehat{K}$  with eigenvalue  $-\lambda < 0$ . But  $\widehat{K}$  is positive definite and so all its eigenvalues are strictly positive. Thus, we have a contradiction, unless  $\Sigma = 0$  and hence  $X = 0$ . □

**Lemma 3.** *If  $J_Q(U) = 0$ , then  $QU = 0$ .*

**Proof.** From the proof of lemma 1, we know that  $LU\bar{L}^T = \iota\Sigma$  where  $L$  is unitary. We define the Hermitian matrices

$$R = LQ^T Q \bar{L}^T, \quad K = L\Lambda \bar{L}^T.$$

Then the result to be proved becomes equivalent to the following

$$R\Sigma K + K\Sigma R = LJ_Q(U)\bar{L}^T = 0 \Rightarrow R\Sigma = 0. \quad (48)$$

This is because

$$R\Sigma = 0 \Rightarrow Q^T QU = 0 \Rightarrow QU = 0,$$

since  $Q^T$  has full (column) rank. Thus we need to show that equation (48) is satisfied, to complete the proof. Note that  $\Sigma$  is real and diagonal,  $K$  is positive definite and  $R^2 = R$ , i.e.  $R$  is a projection matrix. Since  $R$  is a projection onto a subspace of dimension  $n$ , there exists a unitary matrix  $T$  ( $\bar{T}^T T = T\bar{T}^T = I_N$ ), such that

$$TR\bar{T}^T = \left[ \begin{array}{c|c} \overbrace{I_n}^{n \times n} & 0 \\ \hline 0 & \underbrace{0}_{(N-n) \times (N-n)} \end{array} \right].$$

Relative to this decomposition, we write

$$\begin{aligned} T\Sigma\bar{T}^T &= \begin{bmatrix} \Sigma_1 & \bar{\Sigma}_2^T \\ \Sigma_2 & \Sigma_3 \end{bmatrix}, & \bar{\Sigma}_1^T &= \Sigma_1, & \bar{\Sigma}_3^T &= \Sigma_3, \\ TK\bar{T}^T &= \begin{bmatrix} K_1 & \bar{K}_2^T \\ K_2 & K_3 \end{bmatrix}, & \bar{K}_1^T &= K_1, & \bar{K}_3^T &= K_3. \end{aligned}$$

We may then express the equation  $R\Sigma K + K\Sigma R = 0$  in the form

$$\begin{bmatrix} \Sigma_1 & \bar{\Sigma}_2^T \\ 0 & 0 \end{bmatrix} \begin{bmatrix} K_1 & \bar{K}_2^T \\ K_2 & K_3 \end{bmatrix} + \begin{bmatrix} K_1 & \bar{K}_2^T \\ K_2 & K_3 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ \Sigma_2 & 0 \end{bmatrix} = 0. \quad (49)$$

Note that

$$R\Sigma = 0 \Leftrightarrow \Sigma_1 = \Sigma_2 = 0.$$

Since  $K$  is Hermitian and positive definite,

$$x^T K_3 x = [0 \ x^T] \begin{bmatrix} K_1 & \bar{K}_2^T \\ K_2 & K_3 \end{bmatrix} \begin{bmatrix} 0 \\ x \end{bmatrix} > 0$$

for all  $x \neq 0$ . Hence,  $K_3$  is Hermitian and positive definite. Also, since  $Z = TK\bar{T}^T$  is positive definite, this implies that  $Z^{-1}$  is positive definite, and we denote

$$Z^{-1} = \begin{bmatrix} L_1 & \bar{L}_2^T \\ L_2 & L_3 \end{bmatrix},$$

where  $L_1$  and  $L_3$  are also positive definite (by the above argument). From the relations

$$K_2 L_1 + K_3 L_2 = 0, \quad K_1 L_1 + \bar{K}_2^T L_2 = I_n,$$

we obtain  $L_1$  as

$$L_1 = (K_1 - \bar{K}_2^T K_3^{-1} K_2)^{-1}.$$

Hence  $(K_1 - \bar{K}_2^T K_3^{-1} K_2)$  is also positive definite. Now equation (49) is equivalent to the following two independent equations:

$$\begin{aligned} \Sigma_1 K_1 + \bar{\Sigma}_2^T K_2 + K_1 \Sigma_1 + \bar{K}_2^T \Sigma_2 &= 0, \\ K_2 \Sigma_1 + K_3 \Sigma_2 &= 0. \end{aligned}$$

We know  $K_3$  is invertible so  $\Sigma_2 = -K_3^{-1}K_2\Sigma_1$  and

$$\begin{aligned} \Sigma_1 K_1 + (-K_3^{-1}K_2\Sigma_1)^T K_2 + K_1 \Sigma_1 + \bar{K}_2^T (-K_3^{-1}K_2\Sigma_1) &= 0 \\ \Rightarrow \Sigma_1 (K_1 - \bar{K}_2^T K_3^{-1}K_2) + (K_1 - \bar{K}_2^T K_3^{-1}K_2) \Sigma_1 &= 0. \end{aligned}$$

Here we note that  $K_1 - \bar{K}_2^T K_3^{-1}K_2$  is a Hermitian positive definite matrix. Using the result of lemma 2, we see that equation (50) implies that  $\Sigma_1 = 0$ , and hence  $\Sigma_2 = -K_3^{-1}K_2\Sigma_1 = 0$ . Thus  $R\Sigma = 0$ , which implies as we have shown that  $QU = 0$ .  $\square$

Now we present the solution to the algebraic equation (44) or (47).

**Theorem 2.** All solutions of the equation (47) have the form

$$U = U_1 + U_2 + V, \tag{50}$$

where

$$\begin{aligned} U_1 &= \Lambda^{-1}(Q^T(Q\Lambda^{-1}Q^T)^{-1}S - S^T(Q\Lambda^{-1}Q^T)^{-1}Q)\Lambda^{-1}, \\ U_2 &= \Lambda^{-1}(Q^T\widehat{J}_Q^{-1}(R)Q)\Lambda^{-1} \end{aligned} \tag{51}$$

and

$$\begin{aligned} R &= (Q\Lambda^{-1}S^T)(Q\Lambda^{-1}Q^T)^{-1} - (Q\Lambda^{-1}Q^T)^{-1}(S\Lambda^{-1}Q^T), \\ \widehat{J}_Q(X) &= (Q\Lambda^{-1}Q^T)X + X(Q\Lambda^{-1}Q^T); \widehat{J}_Q : \mathfrak{so}(n) \rightarrow \mathfrak{so}(n) \end{aligned} \tag{52}$$

and  $QV = 0$ .

**Proof.** We evaluate  $J_Q(U_1)$  as follows

$$\begin{aligned} J_Q(U_1) &= Q^T Q\Lambda^{-1}(Q^T(Q\Lambda^{-1}Q^T)^{-1}S - S^T(Q\Lambda^{-1}Q^T)^{-1}Q) + (Q^T(Q\Lambda^{-1}Q^T)^{-1}S \\ &\quad - S^T(Q\Lambda^{-1}Q^T)^{-1}Q) \\ &= Q^T S - Q^T(Q\Lambda^{-1}S^T)(Q\Lambda^{-1}Q^T)^{-1}Q + Q^T(Q\Lambda^{-1}Q^T)^{-1}(S\Lambda^{-1}Q^T)Q - S^T Q \\ &= M - Q^T R Q. \end{aligned}$$

Let  $U_2 = \Lambda^{-1}Q^T X Q\Lambda^{-1}$ ; then

$$J_Q(U_2) = Q^T(Q\Lambda^{-1}Q^T)XQ + Q^T X(Q\Lambda^{-1}Q^T)Q = Q^T\widehat{J}_Q(X)Q.$$

Thus if  $\widehat{J}_Q(X) = R$ , then  $J_Q(U_2) = Q^T R Q$ . From lemma 2, we know that  $\widehat{J}_Q : \mathfrak{so}(n) \rightarrow \mathfrak{so}(n)$  is an isomorphism. It follows that  $\widehat{J}_Q$  is invertible,  $\widehat{J}_Q^{-1}(R) = X$  is unique and  $U_2 = \Lambda^{-1}Q^T X Q\Lambda^{-1}$  satisfies  $J_Q(U_2) = Q^T R Q$ . Thus,

$$J_Q(U_1 + U_2) = J_Q(U_1) + J_Q(U_2) = M - Q^T R Q + Q^T R Q = M,$$

and hence all solutions of (47) have the form

$$U = U_1 + U_2 + V, \quad J_Q(V) = 0.$$

From lemma 3, we know that  $J_Q(V) = 0 \Rightarrow QV = 0$ . Thus, we have proved this theorem.  $\square$

Note that this solution to the algebraic equation (44) or (47) uses the decomposition of  $M$  in the  $(Q, S)$  variables.

Let  $X = \widehat{J}_Q^{-1}(R)$ , where  $R$  and  $\widehat{J}_Q$  are as defined in equation (52). The following lemma gives an expression for  $X$  in terms of  $Q$  and  $S$ .

**Lemma 4.** If  $\widehat{J}_Q$  and  $R$  are as given by equation (52), then we can express

$$X = \widehat{J}_Q^{-1}(R) = (Q\Lambda^{-1}Q^T)^{-1}Q\Lambda^{-1}S^T(Q\Lambda^{-1}Q^T)^{-1}. \tag{53}$$

**Proof.** If  $U$  is a solution of the algebraic equation (47), then the quantity  $S$  can be expressed in terms of  $U$  of as

$$\begin{aligned} S &= QU\Lambda = Q\Lambda^{-1}\{Q^T(Q\Lambda^{-1}Q^T)^{-1}S - S^T(Q\Lambda^{-1}Q^T)^{-1}Q\} + Q\Lambda^{-1}Q^TXQ \\ &\Rightarrow S = S - Q\Lambda^{-1}S^T(Q\Lambda^{-1}Q^T)^{-1}Q + Q\Lambda^{-1}Q^TXQ \\ &\Rightarrow Q\Lambda^{-1}Q^TXQ = Q\Lambda^{-1}S^T(Q\Lambda^{-1}Q^T)^{-1}Q \\ &\Rightarrow XQ = (Q\Lambda^{-1}Q^T)^{-1}Q\Lambda^{-1}S^T(Q\Lambda^{-1}Q^T)^{-1}Q. \end{aligned}$$

Post-multiplying both sides of the above expression with  $Q^T$ , we get the expression in (53) for  $X$ .  $\square$

One can verify that  $\widehat{J}_Q(X) = R$ , where  $X$  is as given by (53), as follows

$$\begin{aligned} \widehat{J}_Q(X) &= (Q\Lambda^{-1}Q^T)X + X(Q\Lambda^{-1}Q^T) \\ &= (Q\Lambda^{-1}S^T)(Q\Lambda^{-1}Q^T)^{-1} + (Q\Lambda^{-1}Q^T)^{-1}(Q\Lambda^{-1}S^T) \\ &= (Q\Lambda^{-1}S^T)(Q\Lambda^{-1}Q^T)^{-1} - (Q\Lambda^{-1}Q^T)^{-1}(S\Lambda^{-1}Q^T) = R, \end{aligned}$$

where we used  $Q\Lambda^{-1}S^T = -QUQ^T$  is skew-symmetric in the last step above.

The following statement is a corollary of theorem 2 and lemma 4.

**Corollary 1.** For a given  $Q$ , the map  $\mathcal{Z} : T_Q^*V(n, N) \rightarrow T_QV(n, N)$  where  $\mathcal{Z} : S \mapsto [U]$  is given by theorem 2 is an isomorphism.

**Proof.** Clearly, the map  $\mathcal{Z}^{-1} : [U] \mapsto S = QU\Lambda$  is a linear isomorphism. Observe from theorem 2 and lemma 4 that we get the following expressions for  $U_1$  and  $U_2$ :

$$\begin{aligned} U_1 &= \Lambda^{-1} [Q^T(Q\Lambda^{-1}Q^T)^{-1}S - S^T(Q\Lambda^{-1}Q^T)^{-1}Q] \Lambda^{-1}, \\ U_2 &= \Lambda^{-1} [Q^T(Q\Lambda^{-1}Q^T)^{-1}Q\Lambda^{-1}S^T(Q\Lambda^{-1}Q^T)^{-1}Q] \Lambda^{-1}, \end{aligned}$$

which depend linearly on  $S$ .

It is now easy to verify that

$$\begin{aligned} QU_1 &= [S - Q\Lambda^{-1}S^T(Q\Lambda^{-1}Q^T)^{-1}Q] \Lambda^{-1}, \\ QU_2 &= [Q\Lambda^{-1}S^T(Q\Lambda^{-1}Q^T)^{-1}Q] \Lambda^{-1}, \end{aligned}$$

which gives us  $Q(U_1 + U_2) = S\Lambda^{-1}$ , validating the relation  $S = QU\Lambda$ . Hence, the map  $\mathcal{Z} : S \mapsto [U]$  is an isomorphism.  $\square$

Thus, theorem 2 and lemma 4 describe the exact relationship between the extremal solutions expressed in terms of  $(Q, [U])$  in equations (21) and (22) and those expressed in terms of  $(Q, S)$  in equations (37)–(39).

### 3.3. The discrete variational problem

The discrete counterpart of the variational problem (3) and the discrete extremal trajectories obtained thereof were given by Moser and Veselov (1991). The discrete variational problem is given by

$$\min_{Q_k} \sum_k \frac{1}{2} \langle Q_{k+1}\Lambda, Q_k \rangle, \quad (54)$$

subject to  $Q_k Q_k^T = I_n$ . The extremal trajectories to this discrete variational problem are given by Moser and Veselov (1991)

$$Q_{k+1}\Lambda + Q_{k-1}\Lambda = B_k Q_k, \quad k \in \mathbb{Z}, \quad (55)$$

where  $B_k = B_k^T \in \mathbb{R}^{nm}$  is a (symmetric) Lagrange multiplier matrix for the symmetric constraint  $Q_k Q_k^T = I_n$ . The above equation is the discrete counterpart of equation (37). The discrete body momentum is defined as

$$M_k = Q_{k-1}^T Q_k \Lambda - \Lambda Q_k^T Q_{k-1}. \tag{56}$$

Since  $B_k$  is symmetric, equation (55) is equivalent to the conservation of the discrete spatial momentum

$$m_{k+1} \triangleq Q_{k+1} \Lambda Q_k^T - Q_k \Lambda Q_{k+1}^T = Q_k \Lambda Q_{k-1}^T - Q_{k-1} \Lambda Q_k^T \triangleq m_k. \tag{57}$$

Thus,  $m_k = Q_{k-1} M_k Q_{k-1}^T$  is conserved along the discrete extremal trajectories. Let us define

$$U_k \triangleq Q_{k-1}^T Q_k,$$

which implies that

$$Q_k = Q_{k-1} U_k, \quad M_k = U_k \Lambda - \Lambda U_k^T.$$

The following proposition gives the discrete extremal trajectories in terms of  $U_k$  and the discrete body momentum  $M_k$ .

**Proposition 4.** *The extremal trajectories of the discrete variational problem (54) on the Stiefel manifold  $V(n, N)$  in terms of  $(M_k, U_k)$  are given by*

$$M_{k+1} = U_k^T M_k U_k - A_k, \tag{58}$$

where

$$A_k = U_k^T \Lambda (I_N - U_k^T U_k) - (I_N - U_k^T U_k) \Lambda U_k. \tag{59}$$

**Proof.** One can obtain this result from the second order difference equation (55) which gives the extremal trajectories for (54). Using this equation, we can represent the body momentum at the  $(k + 1)$ th step as

$$\begin{aligned} M_{k+1} &= Q_k^T (B_k Q_k - Q_{k-1} \Lambda) - (B_k Q_k - Q_{k-1} \Lambda)^T Q_k \\ &= \Lambda U_k - U_k^T \Lambda. \end{aligned} \tag{60}$$

From equation (57) expressing conservation of the spatial momentum, we get

$$\begin{aligned} Q_k M_{k+1} Q_k^T &= Q_{k-1} M_k Q_{k-1}^T = Q_k U_k^T M_k U_k Q_k^T \\ \Leftrightarrow M_{k+1} &= U_k^T M_k U_k - A_k \quad \text{where } Q_k A_k Q_k^T = 0. \end{aligned}$$

From the above expression, we get

$$A_k = M_{k+1} - U_k^T M_k U_k,$$

and now using equation (60), we obtain  $A_k$  as given in equation (59). This proves the given result.  $\square$

Note that equation (60) is equivalent to equations (58) and (59) for the discrete body momentum. These equations are therefore the discrete counterpart of equation (46) for the continuous case.

Proposition 4 can also be used to prove theorem 4 of Moser and Veselov (1991), which gives the following set of isospectral deformations for the discrete extremal flows

$$L_{k+1}(\lambda) = C_k(\lambda) L_k(\lambda) C_k^{-1}(\lambda), \tag{61}$$

where  $L_k(\lambda) = \Lambda^2 + \lambda M_k - \lambda^2 Q_{k-1}^T Q_{k-1}$  and  $C_k(\lambda) = \Lambda - \lambda Q_k^T Q_{k-1}$ . Note that  $L_k(\lambda)$  can be factored as

$$L_k(\lambda) = (\Lambda + \lambda Q_{k-1}^T Q_k)(\Lambda - \lambda Q_k^T Q_{k-1}) = C_k^T(-\lambda) C_k(\lambda).$$

Hence, the determinant of  $L_k$  is an even polynomial in  $\lambda$  of degree  $2n$ . Using equations (58) and (59), we can express the left-hand side of equation (61) as

$$\begin{aligned} \Lambda^2 + \lambda(Q_k^T Q_{k-1} M_k Q_{k-1}^T Q_k - A_k) - \lambda^2 Q_k^T Q_k \\ = \Lambda^2 + \lambda(\Lambda Q_{k-1}^T Q_k - Q_k^T Q_{k-1} \Lambda) - \lambda^2 Q_k^T Q_k. \end{aligned}$$

The right-hand side of equation (61) is obtained from the factorization of  $L_k(\lambda)$  as follows

$$\begin{aligned} C_k(\lambda)(C_k^T(-\lambda)C_k(\lambda))C_k^{-1}(\lambda) &= C_k(\lambda)C_k^T(-\lambda) \\ &= (\Lambda - \lambda Q_k^T Q_{k-1})(\Lambda + \lambda Q_{k-1}^T Q_k) \\ &= \Lambda^2 + \lambda(\Lambda Q_{k-1}^T Q_k - Q_k^T Q_{k-1} \Lambda) - \lambda^2 Q_k^T Q_k. \end{aligned}$$

Using this set of isospectral deformations, Moser and Veslov (1991) gave a method to reconstruct the discrete flow under further conditions, and the discrete flow is integrable in this sense.

#### 4. The optimal control problem

We now study the Hamiltonian approach to the variational problem (4). The Hamiltonian for the optimal control problem (4) is given by

$$\begin{aligned} H(P, Q, U) &= \langle P, QU \rangle - \mathcal{L}(Q, U) \\ &= \text{Tr}(P^T QU) + \frac{1}{2} \text{Tr}(\Lambda U Q^T QU), \end{aligned} \quad (62)$$

where  $P \in \mathbb{R}^{nN}$  denotes the costates (Lagrange multipliers). This optimal control problem is nominally posed on  $W_{n,N} = \mathbb{R}^{nN} \times \mathbb{R}^{nN}$ , on which the symplectic structure is given by the symplectic form

$$\omega((A_1, A_2), (B_1, B_2)) = \langle A_1, B_2 \rangle - \langle A_2, B_1 \rangle. \quad (63)$$

We restrict the solutions of this optimal control problem to those which are governed by extremals that leave the submanifolds  $W_{n,N}^k$  invariant, where the  $W_{n,N}^k$  are level sets of  $W_{n,N}$  specified by

$$W_{n,N}^k = \{(Q, P) \in W_{n,N}; Q Q^T = I_n; Q P^T + P Q^T = k\}, \quad (64)$$

where  $k$  is some constant symmetric  $n \times n$  matrix. Note that the dimension of  $W_{n,N}^k$  is given by

$$\text{Dim } W_{n,N}^k = \text{Dim } V(n, N) + nN - \frac{n(n+1)}{2} = 2\text{Dim } V(n, N), \quad (65)$$

which is equal to the dimension of the (co)tangent bundle of  $V(n, N)$ .

##### 4.1. Space of extremal solutions to the optimal control problem

Consider the vector space  $L$  of vector fields on  $W_{n,N}$  characterized by the differential equations

$$\begin{aligned} \dot{Q} &= QU, & U &\in \mathfrak{so}(N), \\ \dot{P} &= PU + QV, & V &\in \mathfrak{so}(N). \end{aligned} \quad (66)$$

Let  $X_1 = (QU_1, PU_1 + QV_1)$  and  $X_2 = (QU_2, PU_2 + QV_2)$  be vectors in  $L$ . This vector space is seen to be a Lie algebra, since

$$\begin{aligned} [X_1, X_2] &= [(QU_1, PU_1 + QV_1), (QU_2, PU_2 + QV_2)] \\ &= (Q[U_1, U_2], P[U_1, U_2] + Q[V_1, U_2] + Q[U_1, V_2]) \end{aligned}$$

also belongs to  $L$ . We now show that the submanifolds  $W_{n,N}^k$  are integral manifolds to the involutive distribution of vector fields in the Lie algebra  $L$ .



**Lemma 5.** *The submanifolds  $W_{n,N}^k \subset W_{n,N}$  are integral submanifolds to the involutive distribution on  $W_{n,N}$  defined by the Lie algebra of vector fields,  $L$ .*

**Proof.** Differentiating the constraints  $QQ^T = I_n$  and  $PQ^T + QP^T = k$ , along trajectories of vector fields in  $L$  defined by the system (66), we get

$$\begin{aligned} QUQ^T - QUQ^T &= 0, \\ (PU + QV)Q^T - PUQ^T + QU P^T + Q(-UP^T - VQ^T) &= 0. \end{aligned}$$

This shows that all vector fields in  $L$  are tangent to each of the submanifolds  $W_{n,N}^k$  of  $W_{n,N} = \mathbb{R}^{nN} \times \mathbb{R}^{nN}$ . We find the dimension of the subspace  $L(Q, P)$  of  $T_{(Q,P)}W_{n,N}^k$  spanned by the vector fields in  $L$ . Since  $U$  and  $V$  are independent

$$\begin{aligned} \text{Dim} \{ (QU, PU + QV), (Q, P) \in W_{n,N}^k, U, V \in \mathfrak{so}(N) \} \\ &= \text{Dim} \{ (QU, PU), U \in \mathfrak{so}(N) \} + \text{Dim} \{ (0, QV), V \in \mathfrak{so}(N) \} \\ &= \text{Dim} \{ (QU, 0), U \in \mathfrak{so}(N) \} + \text{Dim} \{ (0, QV), V \in \mathfrak{so}(N) \} \\ &= 2\text{Dim} \{ QU, U \in \mathfrak{so}(N), QQ^T = I_n \} = 2\text{Dim} V(n, N), \end{aligned}$$

since the dimension of the tangent space to  $V(n, N)$  at  $Q$  is the same as the dimension of  $V(n, N)$  itself. From (65), we see that this is also the dimension of  $W_{n,N}^k$  and hence that of  $T_{(Q,P)}W_{n,N}^k$ . Thus, we conclude that at every  $(Q, P) \in W_{n,N}^k$ ,

$$L(Q, P) = T_{(Q,P)}W_{n,N}^k.$$

This proves the given result. □

Now that we have shown that the  $W_{n,N}^k$  are integral submanifolds of the Lie algebra  $L$ , we next state and prove the following important result.

**Theorem 3.** *The space  $W_{n,N}^k$  defined by (64) is a symplectic submanifold of  $W_{n,N} = (\mathbb{R}^{nN} \times \mathbb{R}^{nN}, \omega)$ .*

**Proof.** To show this result, we need to show that the symplectic form  $\omega$  on  $W_{n,N}$  given by (63) is non-degenerate on  $W_{n,N}^k$  when restricted to  $W_{n,N}^k$ . Let us denote the restriction of the symplectic form on  $W_{n,N}$  to  $W_{n,N}^k$  by

$$\Omega = \omega|_{W_{n,N}^k}.$$

Tangent vectors to  $W_{n,N}^k$  are given by (66). Thus we must show that

$$\Omega((QU_1, PU_1 + QV_1), (QU_2, PU_2 + QV_2)) = 0$$

for all  $(U_2, V_2) \in \mathfrak{so}(N) \times \mathfrak{so}(N)$  implies that  $(QU_1, PU_1 + QV_1) = (0, 0)$ . By definition of  $\Omega$ , we have

$$\langle QU_1, PU_2 + QV_2 \rangle - \langle QU_2, PU_1 + QV_1 \rangle = 0, \quad \forall (U_2, V_2) \in \mathfrak{so}(N) \times \mathfrak{so}(N). \tag{67}$$

Setting  $U_2 = 0$ , we get

$$\text{Tr}(U_1 Q^T Q V_2) = 0, \quad \forall V_2.$$

Since  $V_2$  is skew-symmetric, this in turn implies that  $U_1 Q^T Q$  is symmetric or

$$QU_1 Q^T Q = -QU_1, \quad \text{since } QQ^T = I_n.$$

Thus, each row of  $QU_1$  is an eigenvector of  $Q^T Q$  with an eigenvalue of  $-1$ .

Since  $Q^T Q$  is positive semi-definite, this is a contradiction unless  $QU_1 = 0$ . Now setting  $QU_1 = 0$  we have from (67) that

$$\langle QU_2, PU_1 + QV_1 \rangle = \text{Tr}(U_2 Q^T (PU_1 + QV_1)) = 0, \quad \forall U_2.$$

Since  $U_2$  is skew-symmetric, this implies that

$$\begin{aligned} Q^T(PU_1 + QV_1) &= -(U_1 P^T + V_1 Q^T)Q \\ \Rightarrow PU_1 Q^T + QV_1 Q^T &= -QU_1 P^T - QV_1 Q^T \\ \Rightarrow QV_1 Q^T &= -QU_1 P^T \quad \text{since } QU_1 = 0 \\ \Rightarrow QV_1 Q^T &= 0 \\ \Rightarrow Q^T(PU_1 + QV_1)Q^T &= 0 = -(U_1 P^T + V_1 Q^T)QQ^T \\ \Rightarrow PU_1 + QV_1 &= 0. \end{aligned}$$

Thus,  $(\Omega, W_{n,N}^k)$  for any value of  $k$  is a symplectic submanifold of  $(\omega, W_{n,N})$ .  $\square$

We restrict the extremal flows of the optimal control problem to the symplectic manifolds  $W_{n,N}^k$ , for a value of  $k$  given by the initial conditions.

#### 4.2. Solution to the optimal control problem

The extremal solutions on  $W_{n,N}^k$  to the optimal control problem are characterized by the following result.

**Theorem 4.** *The extremal trajectories of the optimal control problem (4),  $(Q, P) \in W_{n,N}^k$ , are given by*

$$\dot{Q} = QU, \quad U \in \mathfrak{so}(N), \quad (68)$$

$$\dot{P} = PU - QA, \quad (69)$$

where  $A = Q^T QU \wedge U - U \wedge U Q^T Q$ .

**Proof.** The Hamiltonian (62) can be written in the alternate form

$$H(P, Q, U) = \langle P, QU \rangle + \frac{1}{2} \langle Q^T Q, U \wedge U \rangle. \quad (70)$$

This optimal control problem may be restricted (with possible loss of generality) so that extremal trajectories lie on the symplectic manifold  $(W_{n,N}^k, \Omega)$ , where  $\Omega = \omega|_{W_{n,N}^k}$  is the symplectic two-form given by

$$\Omega((\dot{Q}_1, \dot{P}_1), (\dot{Q}_2, \dot{P}_2)) = \langle \dot{Q}_1, \dot{P}_2 \rangle - \langle \dot{Q}_2, \dot{P}_1 \rangle. \quad (71)$$

Hence, we have

$$dH(X_2) = \Omega(X_1, X_2), \quad (72)$$

where  $X_1 = (\dot{Q}_1, \dot{P}_1)$  and  $X_2 = (\dot{Q}_2, \dot{P}_2)$  are vector fields such that  $\dot{Q}_i = QU_i$ ,  $\dot{P}_i = PU_i - QV_i$ ,  $i = 1, 2$ ,  $U_i, V_i \in \mathfrak{so}(N)$  and  $X_1 = X_H$ . Thus, from (72), we get

$$\begin{aligned} \langle PU_2 - QV_2, QU_1 \rangle - \langle PU_1 - QV_1, QU_2 \rangle \\ = \langle PU_2 - QV_2, QU \rangle + \langle P, QU_2 U \rangle + \frac{1}{2} \langle [Q^T Q, U_2], U \wedge U \rangle. \end{aligned}$$

Setting  $U_2 = 0$  in the above expression, we get  $QU_1 = QU$ . Thus, from above, we get

$$\begin{aligned} \langle PU_2, QU \rangle - \langle PU - QV_1, QU_2 \rangle \\ = \langle PU_2, QU \rangle + \langle P, QU_2 U \rangle + \frac{1}{2} \langle [Q^T Q, U_2], U \wedge U \rangle. \end{aligned}$$

So

$$\begin{aligned} \langle QV_1, QU_2 \rangle &= \frac{1}{2} \langle [Q^T Q, U_2], U \Lambda U \rangle \\ &= \frac{1}{2} \langle Q^T QU_2 - U_2 Q^T Q, U \Lambda U \rangle \\ &= \frac{1}{2} \langle QU_2, QU \Lambda U \rangle + \frac{1}{2} \langle QU_2, QU \Lambda U \rangle \\ &= \langle QU_2, QU \Lambda U \rangle. \end{aligned}$$

But

$$\langle QU_2, QU \Lambda U \rangle = \langle QU_2, Q(Q^T QU \Lambda U - U \Lambda U Q^T Q) \rangle,$$

since  $QQ^T = I_n$  and

$$\langle QU_2, QU \Lambda U Q^T Q \rangle = \langle QU_2 Q^T, QU \Lambda U Q^T \rangle = 0,$$

as  $U_2$  is skew-symmetric and  $QU \Lambda U Q^T$  is symmetric. Hence, we get  $V_1 = [Q^T Q, U \Lambda U]$  up to equivalence, and  $X_H$  is the vector field given by

$$\dot{Q} = QU, \quad \dot{P} = PU - QA, \quad \text{where } A = [Q^T Q, U \Lambda U].$$

Thus, the Hamiltonian vector field  $X_H$  prescribes the flow given by equations (68) and (69).  $\square$

We determine the optimal control  $U$  applying Pontryagin's maximum principle (see Bloch *et al* (2003), Gelfand and Fomin (2000) and Kirk (2004)). The Hamiltonian in (62) can also be expressed as

$$H(Q, P, U) = \text{Tr} \frac{1}{2} (P^T Q - Q^T P) U + \frac{1}{2} \text{Tr} Q^T QU \Lambda U. \tag{73}$$

Then  $\text{Grad}_U H(Q, P, U^*) = 0$  with respect to the pairing in (1) is equivalent to

$$Q^T P - P^T Q = Q^T QU^* \Lambda + \Lambda U^* Q^T Q = M, \tag{74}$$

where  $U^*$  is the optimal control. This equation gives the momentum  $M$  in terms of the states  $Q$  and costates  $P$  and also in terms of  $Q$  and  $U^*$  (or  $Q$  and  $S$ ), as given by the solution to the variational problem in equation (44). We now appeal to theorem 2 to give an explicit representation for  $U^*$  in terms of  $(Q, P)$ , by replacing  $S$  with  $P$ .

Taking the time derivative of the equation  $M = Q^T P - P^T Q$  along the vector field given by equations (68) and (69), we get

$$\dot{M} = [M, U] - A, \tag{75}$$

which is identical to equation (46) obtained from the solution to the variational problem.

Since  $QAQ^T = 0$ , one can right-multiply (69) with  $Q^T$  to get

$$\dot{P} Q^T = PU Q^T,$$

which implies (from 68) that  $\lambda = P Q^T$  is conserved. Hence the symmetric quantity

$$k = P Q^T + Q P^T = \lambda + \lambda^T \tag{76}$$

is conserved along the trajectories of (68) and (69).

Since the spatial momentum is

$$m = QMQ^T = P Q^T - Q P^T = \lambda - \lambda^T, \tag{77}$$

as originally defined in (27), this quantity is also conserved along the flow of the extremal solution (68) and (69) to the optimal control problem. Hence, for each initial condition set, the solution trajectory is confined to the level set

$$W_m^k = \{(Q, P) \in W_{n,N}; \quad QQ^T = I_n; \quad QP^T + PQ^T = k, \quad P Q^T - Q P^T = m\}, \tag{78}$$

which is a submanifold of  $W_{n,N}^k$ . The manifold  $W_{n,N}^k$  is formed by an union of the  $W_m^k$  over all values of  $m$ . Clearly, the Hamiltonian vector fields given by (68) and (69) are tangent to the  $W_m^k$  which are level momentum sets of this Hamiltonian flow. Note that  $PQ^T = \lambda = \frac{1}{2}(k+m)$  is constant for  $(P, Q) \in W_m^k \subset W_{n,N}^k$ .

The quantity that corresponds to the body momentum in the full-ranked case is according to (74)

$$M = Q^T P - P^T Q, \tag{79}$$

which has been previously defined in equation (44) in terms of the solution to the variational problem. In the case  $n = N$  we have obtained (Bloch *et al* 2002) the following relation when  $\|m\| < 2$

$$P = Qe^{\sinh^{-1}(M/2)} = e^{\sinh^{-1}(m/2)} Q.$$

For the general case,  $1 \leq n < N$ , when  $PQ^T = \lambda$  is orthogonal and  $\|m\| < 2$ , the extremal solutions (68) and (69) in  $W_m^k \subset W_{n,N}^k$  can be expressed as

$$Q \in V(n, N), \quad P = e^{\sinh^{-1}(m/2)} Q + QR,$$

where  $k = c^T + c$ ,  $c = e^{-\sinh^{-1}(m/2)}$ , and

$$R \in \mathfrak{so}(N), \quad QRQ^T = 0, \quad \dot{R} = [R, U] - A.$$

This can be easily verified by direct substitution into equations (68) and (69).

Note that the quantity  $k$  specifies the symplectic submanifold on which the extremal solution lies, while the spatial momentum  $m$  specifies the momentum level set. We want to express the costates  $P$  of this optimal control problem in terms of  $Q$  and the momentum quantities  $m$  and  $M$ , given by (77) and (79), respectively. Note also that if  $(Q, P)$  is an extremal trajectory to the optimal control problem, i.e.

$$\dot{Q} = QU, \quad \dot{P} = PU - QA, \tag{80}$$

then  $(Q, P+hQ)$ , where  $h$  is a constant  $n \times n$  matrix, is also an extremal trajectory satisfying (80). If  $(Q, P)$  lies in the solution submanifold  $W_{n,N}^k$  given by (64), then  $(Q, P+hQ) \in W_{n,N}^{k+b}$  where  $b = h + h^T$ . Further, if  $h$  is symmetric and  $(Q, P)$  lies in the momentum level set  $W_m^k \subset W_{n,N}^k$ , then  $(Q, P+hQ)$  lies in the momentum level set  $W_m^{k+2h} \subset W_{n,N}^{k+2h}$ ; i.e. the spatial momentum value remains unchanged. One can verify that in this case, the body momentum value  $M$  also remains unchanged. This brings us to the following remarkable result.

**Proposition 5.** *We have a map  $\Xi : W_{n,N}^k \rightarrow W_{n,N}^0$ , defined by  $\Xi : (Q, P) \mapsto (Q, P_0) \in W_{n,N}^0$ , where*

$$P_0 = P - \frac{1}{2}kQ, \quad (Q, P) \in W_{n,N}^k. \tag{81}$$

*Further, if  $(Q, P) \in W_m^k \subset W_{n,N}^k$ , then  $(Q, P_0) \in W_m^0 \subset W_{n,N}^0$ , i.e. it satisfies equations (80) and*

$$\begin{aligned} P_0 Q^T - Q P_0^T &= m = P Q^T - Q P^T, \\ Q^T P_0 - P_0^T Q &= M = Q^T P - P^T Q. \end{aligned}$$

*Hence, the map  $\Xi$  leaves the spatial and body momenta unchanged.*

**Proof.** Let  $(Q, P) \in W_{n,N}^k$ . Then we have

$$\begin{aligned} P_0 Q^T + Q P_0^T &= (P - \frac{1}{2}kQ) Q^T + Q (P - \frac{1}{2}kQ)^T \\ &= P Q^T - \frac{1}{2}k + Q P^T - \frac{1}{2}k \\ &= k - k = 0. \end{aligned}$$

Thus  $(Q, P)$  is mapped to  $(Q, P_0) \in W_{n,N}^0$  by  $\Xi$ . If  $(Q, P) \in W_m^k$ , then  $PQ^T - QP^T = m$  in addition to  $PQ^T + QP^T = k$ , and we have

$$\begin{aligned} P_0Q^T - QP_0^T &= (P - \frac{1}{2}kQ)Q^T - Q(P - \frac{1}{2}kQ)^T \\ &= PQ^T - \frac{1}{2}k - QP^T + \frac{1}{2}k \\ &= PQ^T - QP^T = m. \end{aligned}$$

This proves the second part of the statement. □

The following result is a corollary of proposition 5 and gives the costate  $P = P_0$  as a function of  $Q, m$  and  $M$ , such that  $(Q, P_0) \in W_m^0$ . As we show later, this costate is a natural choice since a direct relation exists between the symplectic forms on  $T^*V(n, N)$  and  $W_{n,N}^0$ .

**Corollary 2.** *The pair  $(Q, P_0) \in W_m^0$  satisfying equations (80) may be expressed as*

$$P_0 = -\frac{1}{2}mQ + QM. \tag{82}$$

For this solution,  $P_0 = Q\bar{M}$  where

$$\bar{M} = -\frac{1}{2}Q^TmQ + M \tag{83}$$

is a momentum-like quantity.

**Proof.** The proof of this statement can be carried out in two stages. In the first stage, we consider an extremal solution pair  $(Q, P) \in W_m^k \subset W_{n,N}^k$  and obtain an expression for  $P$  in terms of  $Q$  and  $M$ . We observe that if

$$P = \mu Q + QM,$$

where  $\mu \in \mathbb{R}^{nn}$  is constant and  $M$  is the body momentum, then  $(Q, P) \in W_{n,N}^k$ . This can be easily verified by taking a time derivative of  $P$  along the extremal solutions given by equations (68) and (75). We also observe that  $PQ^T = \mu + m$  for this solution pair. This solution is in  $W_m^k \subset W_{n,N}^k$  if and only if

$$\begin{aligned} PQ^T + QP^T &= \mu + \mu^T = k, & \text{and} \\ PQ^T - QP^T &= \mu - \mu^T + 2m = m \Leftrightarrow \mu^T - \mu = m. \end{aligned}$$

This gives  $\mu$  uniquely as

$$\mu = \frac{1}{2}(k - m),$$

and thus  $(Q, P) \in W_m^k$  where

$$P = \frac{1}{2}(k - m)Q + QM. \tag{84}$$

Now applying proposition 5 in the second stage of this proof, we have  $(Q, P_0) \in W_m^0$  where  $P_0 = \bar{P} - \frac{1}{2}kQ$ , which when applied to (84) gives the result expressed in equation (82). As a result of proposition 5, we also know that the momenta  $m$  and  $M$  are left unchanged by this transformation in the costate variable. That  $P_0 = Q\bar{M}$  is easily verified by substituting  $\bar{M}$  from equation (83) and comparing with equation (82). □

Note that, with the costate variable  $P = P_0$ , the solution to the optimal control problem satisfies

$$k = QP_0^T + P_0Q^T = -\frac{1}{2}m + \frac{1}{2}m = 0. \tag{85}$$

Also note that, for the special case of the rigid body in  $N$  dimensions ( $n = N$ ), the momentum quantity  $\bar{M} = \frac{1}{2}M$  is half the body momentum.

This result has further important implications for the symplectic structure of  $W_{n,N}^0$  and its relation to the symplectic structure on  $T^*V(n, N)$ , which we will explore in the next section. Note that equations (68) and (69) conserve  $PQ^T$  and  $QP^T$  separately. In the symmetric representation of the rigid body equations given in Bloch *et al* (2002),  $PQ^T$  and hence  $PQ^T - QP^T$  and  $PQ^T + QP^T$  are constant, and this generalizes to the extremal solution (in  $W_{n,N}^k$ ) of the optimal control problem on the Stiefel manifold  $V(n, N)$  for  $1 \leq n < N$ .

#### 4.3. Correspondence of the variational and optimal control solutions

We now give the correspondence between the variational (or Lagrangian) and optimal control (or Hamiltonian) representations of the extremal solutions to this problem. The extremal solutions to the variational problem can be defined in terms of the pair  $(Q, \dot{Q}) \in TV(n, N)$  or  $(Q, S) \in T^*V(n, N)$  and extremal solutions to the optimal control problem are defined in terms of the pair  $(Q, P_0) \in W_{n,N}^0$ , a symplectic manifold. Here we describe a correspondence between them,  $\mathfrak{M} : T^*V(n, N) \rightarrow W_{n,N}^0$  such that  $\mathfrak{M} : (Q, S) \mapsto (Q, P_0) \in W_{n,N}^0$ .

From the solutions to the variational and optimal control problems, we know that the body momentum  $M$  satisfies

$$M = Q^T S - S^T Q = Q^T P_0 - P_0^T Q. \quad (86)$$

In addition, on pre-multiplying equation (86) by  $Q$  and post-multiplying it by  $Q^T$  on both sides, we also get

$$m = QMQ^T = SQ^T - QS^T = P_0Q^T - QP_0^T, \quad (87)$$

which is derived from (86). Equation (86) leads us to the following result.

**Corollary 3.** *The extremal solutions  $(Q, S)$  and  $(Q, P_0)$  to the variational and optimal control problems, respectively, are related by*

$$P_0 = S + X, \quad (88)$$

where

$$X = DQ, \quad D = D^T. \quad (89)$$

This statement is a corollary of proposition 1 and is obtained from inspection of equation (86). The difference  $P_0 - S = X$  in equation (88) obviously lies in the kernel of the linear map  $L_Q : \mathbb{R}^{nN} \rightarrow \mathbb{R}^{nN}$  defined in (30). From proposition 1, we know that  $X$  is going to have the form given by equation (89).

For the extremal solution to the optimal control problem  $(Q, P_0) \in W_m^0 \subset W_{n,N}^0$ , we know from corollary 2 that  $P_0 = -\frac{1}{2}mQ + QM$ . The extremal solution of the variational problem in section 3 was given in terms of  $(Q, S) \in T^*V(n, N)$ . The following result gives the relation between these two solution pairs.

**Proposition 6.** *The map  $\mathfrak{M} : T^*V(n, N) \rightarrow W_{n,N}^0$  defined by  $\mathfrak{M}((Q, S)) = (Q, P_0)$  is a diffeomorphism and  $P_0$  is given in terms of  $Q$  and  $S$  by*

$$P_0 = S + DQ, \quad (90)$$

where

$$D = D^T = \frac{1}{2}Q[\Lambda, U]Q^T = -\frac{1}{2}(QS^T + SQ^T). \quad (91)$$

The inverse of this diffeomorphism is given by  $\mathfrak{M}^{-1} : (Q, P_0) \mapsto (Q, S)$  where

$$S = P_0 - [(Q\Lambda^{-1}P_0^T)(Q\Lambda^{-1}Q^T)^{-1} - Q\Lambda^{-1}Q^T\widehat{J}_Q^{-1}(R_0)]Q, \quad (92)$$

where  $\widehat{J}_Q^{-1}(R_0)$  is given by equations (52) with  $S$  replaced by  $P_0$ .

**Proof.** From corollary 3, we know that  $P_0 = S + DQ$  where  $D$  is symmetric. Thus, we have

$$S = P_0 - DQ = -\left(\frac{1}{2}m + D\right)Q + QM,$$

where  $M = Q^T S - S^T Q$  and  $m = QM Q^T$ . This gives us

$$\begin{aligned} & -\left(\frac{1}{2}m + D\right)Q + Q(Q^T S - S^T Q) = S, \\ \Leftrightarrow & -\frac{1}{2}(SQ^T - QS^T)Q - DQ = QS^T Q, \\ \Leftrightarrow & -DQ = \frac{1}{2}(QS^T + SQ^T)Q, \\ \Leftrightarrow & D = -\frac{1}{2}(QS^T + SQ^T) = -\frac{1}{2}Q[\Lambda, U]Q^T, \end{aligned}$$

since  $S = QU\Lambda$ , where  $QU$  is obtained from  $Q$  and  $S$  using theorem 2. This establishes the relation given by equations (90) and (91) between the variables  $S$  and  $P_0$ , and  $P_0$  is seen to be linearly dependent on  $S$ . Clearly, if  $S = 0$  then  $P_0 = 0$ . For the converse, if  $P_0 = 0$  then  $S = -DQ$ . Since  $D$  is symmetric, we have  $Q^T S = S^T Q = -Q^T DQ$ . Since  $S = QU\Lambda$ , we have  $Q^T S - S^T Q = J_Q(U)$ , as defined in equation (47). From lemma 3, we know that if  $J_Q(U) = 0$ , then  $QU = 0$ . Thus, when  $P_0 = 0$ , we have  $J_Q(U) = 0$  and hence  $S = QU\Lambda = 0$ . Thus  $\mathfrak{M}((Q, S)) = (Q, P_0)$  as given by (90) is a diffeomorphism. For the inverse of this map, it is clear from theorem 2 and its proof that  $U$  can be expressed in terms of  $Q$  and  $P_0$  in the same manner that it is expressed in terms of  $Q$  and  $S$  in that theorem. This is true because that representation is based on the decomposition  $M = Q^T S - S^T Q$ , which has the same form as  $M = Q^T P_0 - P_0^T Q$ . Hence, we also have a diffeomorphism  $\mathcal{Z}_0 : W_{n,N}^0 \rightarrow TV(n, N)$  defined by  $\mathcal{Z}_0((Q, P_0)) = (Q, [U]) = (Q, [U_1 + U_2])$  as follows

$$\begin{aligned} U_1 &= \Lambda^{-1}(Q^T(Q\Lambda^{-1}Q^T)^{-1}P_0 - P_0^T(Q\Lambda^{-1}Q^T)^{-1}Q)\Lambda^{-1}, \\ U_2 &= \Lambda^{-1}(Q^T\widehat{J}_Q^{-1}(R_0)Q)\Lambda^{-1}, \end{aligned}$$

where

$$R_0 = (Q\Lambda^{-1}P_0^T)(Q\Lambda^{-1}Q^T)^{-1} - (Q\Lambda^{-1}Q^T)^{-1}(P_0\Lambda^{-1}Q^T).$$

Thus, the inverse of  $\mathfrak{M}$  is given by  $\mathfrak{M}^{-1} = \mathcal{Z}^{-1} \circ \mathcal{Z}_0$ , where  $\mathcal{Z}$  is as defined in corollary 1. From this corollary, we also know that  $S = Q(U_1 + U_2)\Lambda$  uniquely determines  $S$ , and this gives us the relation (92) for  $S$  in terms of  $P_0$  and  $Q$ .  $\square$

Note that if  $\Lambda = I_N$ , the  $N \times N$  identity matrix, then  $S = P_0$ . Also note that  $D = (Q\Lambda^{-1}P_0^T)(Q\Lambda^{-1}Q^T)^{-1} - Q\Lambda^{-1}Q^T\widehat{J}_Q^{-1}(R_0)$  in (92) is a symmetric matrix. This can be shown using equations (52) of theorem 2 as follows:

$$\begin{aligned} D - D^T &= (Q\Lambda^{-1}P_0^T)(Q\Lambda^{-1}Q^T)^{-1} - Q\Lambda^{-1}Q^T\widehat{J}_Q^{-1}(R_0) - \widehat{J}_Q^{-1}(R_0)Q\Lambda^{-1}Q^T \\ &\quad - (Q\Lambda^{-1}Q^T)^{-1}(Q\Lambda^{-1}P_0^T) \\ &= (Q\Lambda^{-1}P_0^T)(Q\Lambda^{-1}Q^T)^{-1} - Q\Lambda^{-1}Q^T X_0 - X_0 Q\Lambda^{-1}Q^T \\ &\quad - (Q\Lambda^{-1}Q^T)^{-1}(Q\Lambda^{-1}P_0^T) \\ &= R_0 - \widehat{J}_Q(X_0) = R_0 - R_0 = 0, \end{aligned}$$

where  $X_0 = \widehat{J}_Q^{-1}(R_0)$ . This expresses the symmetric matrix  $D$  in terms of  $Q$  and  $P_0$ , while  $D$  is also given as a function of  $Q$  and  $S$  by equation (91).

### 5. The tangent and cotangent bundles of the Stiefel manifold

In this section, we explore the structure of the tangent and cotangent bundles of the Stiefel manifold, which is the homogeneous space  $V(n, N) = SO(N)/SO(N - n)$ . The Stiefel

manifold is parametrized by  $Q \in \mathbb{R}^{nN}$  such that  $QQ^T = I_n$ . Consider the point  $Q_0 = [I_n \ 0]$  on  $V(n, N)$ , where  $0$  denotes the  $(N - n) \times n$  matrix of zeros.  $SO(N)$  acts on  $Q_0$  on the right as  $\mathcal{R} : Q_0 \mapsto Q_0 R$  where  $R \in SO(N)$ , and the isotropy group is the subgroup

$$\left\{ R \in SO(N) \mid R = \begin{bmatrix} I_n & 0 \\ 0 & \bar{R} \end{bmatrix}; \bar{R} \in SO(N - n) \right\}.$$

This gives the dimension of  $V(n, N)$  as the difference of the dimensions of  $SO(N)$  and  $SO(N - n)$

$$\dim V(n, N) = N(N - 1)/2 - (N - n)(N - n - 1)/2 = nN - n(n + 1)/2.$$

### 5.1. Symplectic structure of the cotangent bundle of the Stiefel manifold

The tangent space to  $V(n, N)$  at  $Q$  is parametrized by  $QU, U \in \mathfrak{so}(N)$ . However,  $U$  is unique only upto the equivalence class  $[U]$  defined by equation (25). If  $Q = [I_n \ 0]$  as before, then the set  $V \in SO(N)$  such that  $QV = 0$  is given by

$$\left\{ V \in SO(N) \mid V = \begin{bmatrix} 0 & 0 \\ 0 & \bar{V} \end{bmatrix}, \bar{V} \in \mathfrak{so}(N - n) \right\}.$$

This determines  $T_Q V(n, N)$  as the vector space

$$\mathfrak{so}(N)/\mathfrak{so}(N - n).$$

We occasionally refer to elements of  $T_Q V(n, N)$  as the equivalence classes  $Q[U]$ , where  $[U]$  is as defined by equation (25).

We may parametrize the cotangent space  $T_Q^* V(n, N)$  by

$$P \in T_Q^* V(n, N), \quad P = 2\langle QM, \cdot \rangle, \quad M \in \mathfrak{so}(N).$$

Let  $X \in T_Q V(n, N)$ , hence  $X = QU, U \in \mathfrak{so}(N)$ . Then we have

$$P(X) = 2\langle QM, QU \rangle = 2\langle Q^T QM, U \rangle.$$

In the case  $n = N$ , this pairing gives the Killing form on  $\mathfrak{so}(N)$ , which is non-degenerate. The following result shows that in the equivalence classes  $[\cdot]$  defined on  $\mathfrak{so}(N)$ , this pairing between  $T_Q V(n, N)$  and  $T_Q^* V(n, N)$  is non-degenerate.

**Lemma 6.** *The pairing  $\mathfrak{D} : T_Q^* V(n, N) \times T_Q V(n, N) \rightarrow \mathbb{R}$  given by  $\mathfrak{D}(Q[M], Q[U]) \mapsto \langle Q[M], Q[U] \rangle$  is non-degenerate.*

**Proof.** For this pairing to be non-degenerate on  $T_Q V(n, N)$ , we should have

$$\langle QM, QU \rangle = 0 \forall U \Rightarrow Q[M] = 0.$$

Evaluating the pairing on the left-hand side gives us

$$\begin{aligned} \langle QM, QU \rangle &= \langle Q^T QM, U \rangle = 0 \forall U \in \mathfrak{so}(N) \\ &\Leftrightarrow Q^T QM + MQ^T Q = 0. \end{aligned}$$

Now, by lemma 3, the above equality is satisfied if and only if  $QM = 0$  (substituting  $U = M$  and  $\Lambda = I_N$  in lemma 3).

Thus, we have shown that the bilinear pairing  $\mathfrak{D}(Q[M], Q[U]) \mapsto \langle Q[M], Q[U] \rangle$  is non-degenerate on  $T_Q V(n, N)$ .  $\square$

Hence, this representation of  $T_Q^* V(n, N)$  is well defined.

Identifying  $T^* V(n, N)$  and  $T\check{V}(n, N)$  by this pairing, we may parametrize  $TT^* V(n, N)$  by vectors  $(Q, QM, QU, QUM + QZ)$ , where  $M, U, Z \in \mathfrak{so}(N)$ . Viewing the cotangent bundle



$T^*V(n, N)$  as a subset of  $\mathbb{R}^{nN} \times \mathbb{R}^{nN}$ , we may pull back the symplectic form on  $\mathbb{R}^{nN} \times \mathbb{R}^{nN}$  given by

$$\omega((A_1, B_1), (A_2, B_2)) = \langle B_2, A_1 \rangle - \langle B_1, A_2 \rangle$$

to  $T^*V(n, N)$  via the inclusion map. If

$$\begin{aligned} X_1 &= (Q, QM, QU_1, QU_1M + QZ_1), \\ X_2 &= (Q, QM, QU_2, QU_2M + QZ_2) \in T_{(Q, QM)}T^*V(n, N), \end{aligned}$$

then the two form we obtain on  $T^*V(n, N)$  is

$$\begin{aligned} \omega_{(Q, QM)}(X_1, X_2) &= \langle QU_2M + QZ_2, QU_1 \rangle - \langle QU_1M + QZ_1, QU_2 \rangle \\ &= \langle Q^T QU_1, Z_2 \rangle - \langle Q^T QU_2, Z_1 \rangle + \langle M, U_1 Q^T QU_2 - U_2 Q^T QU_1 \rangle. \end{aligned} \quad (93)$$

It is simple to check using lemma 3 that this is indeed nondegenerate and hence a symplectic form. This expression for the symplectic form can also be obtained using the canonical structure on  $T^*V(n, N)$ . In general, for a smooth manifold  $M$  and using the projection  $\pi : T^*M \rightarrow M$ , we define a one form  $\theta$  on  $T^*M$  by

$$\theta_p(X) = p\pi_*X,$$

where  $p \in T^*M$  and  $X$  is a vector field on  $T^*M$ . We define the canonical symplectic form  $\omega_c$  on  $T^*M$  by setting  $\omega_c = -d\theta$ , where  $d\theta$  is the exterior derivative of the one form  $\theta$ . We may simplify this expression using the identity (given in Bloch *et al* (2003))

$$d\theta(X, Y) = X(\theta(Y)) - Y(\theta(X)) - \theta([X, Y]),$$

so

$$\omega_c(X, Y) = Y(\theta(X)) - X(\theta(Y)) + \theta([X, Y]). \quad (94)$$

We apply equation (94) to the case where the manifold  $M = V(n, N)$ . We parametrize  $T^*V(n, N)$  by pairs  $(Q, M)$ ,  $Q \in V(n, N)$  and  $M \in \mathfrak{so}(N)$ . Then we can write elements of  $T_{(Q, M)}T^*V(n, N)$  at the point  $(Q, M)$  as

$$X_k = (QU_k, Z_k), \quad k = 1, 2.$$

Hence, if  $\pi$  is the projection  $\pi : TT^*V(n, N) \rightarrow TV(n, N)$ , then  $\pi_*X_k = QU_k$ , and if  $p \in T^*V(n, N)$ ,  $p = \langle QM, \cdot \rangle$ , then

$$\theta(X_k) = p\pi_*X_k = \langle QM, QU_k \rangle.$$

Thus, from equation (94) we have

$$\omega_c(X_1, X_2) = X_2\langle QM, QU_1 \rangle - X_1\langle QM, QU_2 \rangle + \langle QM, \pi_*[X_1, X_2] \rangle. \quad (95)$$

The last term in the above equation can be simplified as follows

$$\begin{aligned} \pi_*[X_1, X_2] &= [\pi_*X_1, \pi_*X_2] \circ \pi \\ &= [QU_1, QU_2] \quad (\text{as vector fields}) \\ &= QU_1U_2 - QU_2U_1 \\ &= Q[U_1, U_2]. \end{aligned}$$

The first two terms are given by

$$X_2\langle QM, QU_1 \rangle = (QU_2, Z_2)\langle QM, QU_1 \rangle = \langle QZ_2, QU_1 \rangle + \langle QU_2M, QU_1 \rangle + \langle QM, QU_2U_1 \rangle$$

and

$$\begin{aligned} X_1\langle QM, QU_2 \rangle &= (QU_1, Z_1)\langle QM, QU_2 \rangle \\ &= \langle QZ_1, QU_2 \rangle + \langle QU_1M, QU_2 \rangle + \langle QM, QU_1U_2 \rangle. \end{aligned}$$

Thus, we obtain

$$\begin{aligned}\omega_c(X_1, X_2) &= \langle QZ_2, QU_1 \rangle - \langle QZ_1, QU_2 \rangle - \langle M, U_2 Q^T QU_1 \rangle + \langle M, U_1 Q^T QU_2 \rangle \\ &\quad + \langle M, Q^T Q(U_2 U_1 - U_1 U_2) \rangle + \langle QM, Q[U_1, U_2] \rangle \\ &= \langle QZ_2, QU_1 \rangle - \langle QZ_1, QU_2 \rangle + \langle M, U_1 Q^T QU_2 - U_2 Q^T QU_1 \rangle,\end{aligned}\quad (96)$$

which is identical to equation (93) obtained from restricting the two-form on  $\mathbb{R}^{nN} \times \mathbb{R}^{nN}$  to  $T^*V(n, N)$ .

We now have a formula for the natural symplectic form on  $T^*V(n, N)$  and know that it can be formulated as the restriction of the symplectic form on the product  $\mathbb{R}^{nN} \times \mathbb{R}^{nN}$ . We want to use this formula to recover the Hamiltonian flow corresponding to the geodesic problem (4) on  $V(n, N)$ . The Hamiltonian is given by (62) as  $H(P, Q, U) = \langle P, QU \rangle - \langle QU, QU \Lambda \rangle$  and the optimal control  $U^*$  is given by the maximum principle as in (74). We replace  $P$  by the parametrization  $QM$ , where we expect  $M$  to be the momentum. Thus, we can write the Hamiltonian as

$$H(M, Q, U) = \langle QM, QU \rangle - \frac{1}{2} \langle QU \Lambda, QU \rangle. \quad (97)$$

The Hamiltonian flow on  $T^*V(n, N)$  is obtained from the solution of

$$dH_{(M, Q, U)} X_2 = \omega_c(X_1, X_2), \quad \forall X_2 = (QU_2, Z_2),$$

where  $X_1 = X_H$  is the Hamiltonian vector field corresponding to  $H$ . We calculate

$$\begin{aligned}dH(X_2) &= \langle QU_2 M, QU \rangle + \langle QM, QU_2 U \rangle + \langle QZ_2, QU \rangle \\ &\quad - \frac{1}{2} \langle QU_2 U \Lambda, QU \rangle - \frac{1}{2} \langle QU \Lambda, QU_2 U \rangle,\end{aligned}$$

and this is equated to  $\omega_c(X_H, X_2)$  in (96). From equating these expressions after replacing  $Z_1$  and  $U_1$  by  $Z$  and  $U$ , respectively, in (96), we get

$$\langle QU_2, QZ \rangle = \langle QU_2, QMU \rangle - \langle QU_2, QUM \rangle - \langle QU_2, QU \Lambda U \rangle. \quad (98)$$

Note that  $\langle QU_2 Q^T, QU \Lambda U Q^T \rangle = 0$  since  $QU_2 Q^T$  is skew-symmetric and  $QU \Lambda U Q^T$  is symmetric. Therefore,  $\langle QU_2, QU \Lambda U Q^T Q \rangle = 0$  and we can express the last term in equation (98) as

$$\langle QU_2, QU \Lambda U \rangle = \langle QU_2, Q(Q^T QU \Lambda U - U \Lambda U Q^T Q) \rangle.$$

If  $F$  is skew-symmetric and  $\langle QU_2, QU \Lambda U \rangle = \langle QU_2, QF \rangle$ , then

$$\langle QU_2, Q(F - [Q^T Q, U \Lambda U]) \rangle = \langle QU_2, QU \Lambda U Q^T Q \rangle = 0 \quad \forall U_2.$$

Thus,  $F = [Q^T Q, U \Lambda U]$  up to an equivalence class. Hence, from (98) we get

$$\begin{aligned}\langle QU_2, Q(Z_1 - [M, U] + [Q^T Q, U \Lambda U]) \rangle &= 0 \quad \forall U_2 \\ \Rightarrow Z_1 &= [M, U] - [Q^T Q, U \Lambda U],\end{aligned}$$

up to equivalence class. It follows that the geodesics on  $T^*V(n, N)$  with the metric in (4) are given by the Hamiltonian flow

$$\begin{aligned}\dot{Q} &= QU, \\ \dot{M} &= [M, U] - A, \quad A = [Q^T Q, U \Lambda U],\end{aligned}$$

as given previously by (46) and (75).

### 5.2. Symplectomorphism between the cotangent bundle and $W_{n,N}^0$

Finally we explore the relation between the symplectic structures on the cotangent bundle  $T^*V(n, N)$  and  $W_{n,N}^0$ . Consider the Hamiltonian vector fields defined by (68) and (69):

$$\begin{aligned} X_1 &= (QU_1, P_0U_1 - QA_1), \\ X_2 &= (QU_2, P_0U_2 - QA_2), \end{aligned}$$

where  $A_i = [Q^T Q, U_i \Lambda U_i]$ ,  $i = 1, 2$ . The symplectic form  $\Omega$  on  $W_{n,N}^0$  evaluated along these vector fields on  $W_{n,N}^0$  is obtained using equation (71) as

$$\Omega_{(Q, P_0)}(X_1, X_2) = \langle P_0U_2 - QA_2, QU_1 \rangle - \langle P_0U_1 - QA_1, QU_2 \rangle. \quad (99)$$

From corollary 2, we know that if  $(Q, P_0) \in W_m^0 \subset W_{n,N}^0$ , then

$$P_0 = -\frac{m}{2}Q + QM = Q\bar{M},$$

where  $\bar{M}$  is given by (83). Now we show the following relationship between  $T^*V(n, N)$  and  $W_{n,N}^0$ .

**Theorem 5.** *The map  $\Phi : (\Omega, W_{n,N}^0) \rightarrow (\omega_c, T^*V(n, N))$ , given by  $(Q, P_0) \mapsto (Q, Q\bar{M})$  where*

$$\bar{M} = -\frac{1}{2}Q^T(P_0Q^T - QP_0^T)Q + Q^T P_0 - P_0^T Q, \quad (100)$$

*is a symplectomorphism.*

**Proof.** Coordinates for  $T^*V(n, N)$  are given by  $(Q, Q\bar{M})$ . Consider the Hamiltonian vector fields on  $T^*V(n, N)$  corresponding to the extremal flows for the optimal control problem considered in the last section

$$\begin{aligned} \chi_1 &= \partial_1(Q, Q\bar{M}) = (QU_1, QU_1\bar{M} + QZ_1), \\ \chi_2 &= \partial_2(Q, Q\bar{M}) = (QU_2, QU_2\bar{M} + QZ_2), \end{aligned}$$

where

$$\begin{aligned} Z_1 &= \partial_1\bar{M} = [\bar{M}, U_1] - A_1, \\ Z_2 &= \partial_2\bar{M} = [\bar{M}, U_2] - A_2. \end{aligned}$$

We show that the push-forwards of the Hamiltonian vector fields  $X_i$  on  $W_{n,N}^0$  defined earlier give the Hamiltonian vector fields  $\chi_i$  on  $T^*V(n, N)$ . We have

$$\begin{aligned} (\Phi_* X_i)(\Phi(Q, P_0)) &= (T_{(Q, P_0)}\Phi \cdot X_i)(\Phi(Q, P_0)) \\ &= (QU_i, Q\bar{M}U_i - QA_i) \\ &= (QU_i, QU_i\bar{M} + Q[\bar{M}, U_i] - QA_i) \\ &= (QU_i, QU_i\bar{M} + QZ_i) = \chi_i(Q, Q\bar{M}) \\ \Leftrightarrow \Phi_* X_i &= \chi_i. \end{aligned}$$

The canonical symplectic form on  $T^*V(n, N)$  is evaluated as

$$\begin{aligned} \omega_c_{(Q, Q\bar{M})}(\chi_1, \chi_2) &= \langle QU_2\bar{M} + QZ_2, QU_1 \rangle - \langle QU_1\bar{M} + QZ_1, QU_2 \rangle \\ &= \langle QU_2\bar{M} + Q[\bar{M}, U_2] - QA_2, QU_1 \rangle \\ &\quad - \langle QU_1\bar{M} + Q[\bar{M}, U_1] - QA_1, QU_2 \rangle \\ &= \langle Q\bar{M}U_2 - QA_2, QU_1 \rangle - \langle Q\bar{M}U_1 - QA_1, QU_2 \rangle. \end{aligned} \quad (101)$$

The pull-back of this symplectic form from  $T^*V(n, N)$  to  $W_{n,N}^0$  gives us the symplectic form on  $W_{n,N}^0$ , as shown below:

$$\begin{aligned} (\Phi^*\omega_c)_{(Q, P_0)}(X_1, X_2) &= \omega_{c(Q, Q\bar{M})}(\Phi_*X_1, \Phi_*X_2) = \omega_{c(Q, Q\bar{M})}(X_1, X_2) \\ &= \langle Q\bar{M}U_2 - QA_2, QU_1 \rangle - \langle Q\bar{M}U_1 - QA_1, QU_2 \rangle \\ &= \langle P_0U_2 - QA_2, QU_1 \rangle - \langle P_0U_1 - QA_1, QU_2 \rangle = \Omega_{(Q, P_0)}(X_1, X_2). \end{aligned}$$

However, from lemma 5, we know that the vector fields  $X_i$  form a Lie algebra that spans the tangent space to  $W_{n,N}^0$  at every point  $(Q, P_0)$ . Hence, we have

$$\Phi^*\omega_c = \Omega. \quad (102)$$

The map  $\Phi : W_{n,N}^0 \rightarrow T^*V(n, N)$  given by  $(Q, P_0) \mapsto (Q, Q\bar{M})$ , where  $\bar{M}$  is given by (100), has an inverse which is simply given by  $\Phi^{-1} : (Q, Q\bar{M}) \mapsto (Q, P_0)$  where  $P_0 = Q\bar{M}$ . The inverse map  $\Phi^{-1}$  is clearly a diffeomorphism. Thus, the map  $\Phi$  is a diffeomorphism that also maps the symplectic form in  $W_{n,N}^0$  to the symplectic form in  $T^*V(n, N)$  and is hence a symplectomorphism.  $\square$

With  $\bar{M}$  expressed in terms of  $Q$  and  $P_0$  as in this theorem, we have that  $Q\bar{M} = P_0$  where  $(Q, P_0) \in W_{n,N}^0$ . Hence, theorem 5 shows that the solution space  $W_{n,N}^0$  for the optimal control problem can be identified with the cotangent bundle of the Stiefel manifold. The Hamiltonian formulation of the optimal control problem parametrized by  $(Q, P_0)$  or  $(Q, Q\bar{M})$  is related to the variational formulation of this problem parametrized by  $(Q, \dot{Q})$  or  $(Q, S)$ , via the relationship established in proposition 6. Combining theorem 5 with proposition 6, we get the following expression for  $\bar{M}$  in terms of  $(Q, S)$ :

$$\bar{M} = Q^T \bar{S} - \bar{S}^T Q, \quad \bar{S} = S(I_N - \frac{1}{2}Q^T Q). \quad (103)$$

Of course, in the case  $\Lambda = I_N$ , we have  $P_0 = S$  and these solutions are directly related.

## 6. Applications and open problems

Nowadays numerical linear algebra computations and numerical integration of ODEs are increasingly based on variational problems on manifolds. Problems on Stiefel (and Grassman) manifolds are finding increasing use in numerical linear algebra applications (see, for example, Edelman *et al* (1998), Elden and Park (1999) and references therein). The first of these papers develops Newton and conjugate gradient methods on these manifolds, while the latter paper deals with a problem related to regression analysis in psychometrics. Another application of numerical calculations on Stiefel manifolds is in computing Lyapunov exponents for finite-dimensional dynamical systems by time integration. The Lyapunov exponents are computed by a continuous orthonormalization (which is essential for stable numerical integration) of a set of solution vectors of the linearized system. This amounts to restricting the linearized system to the Stiefel manifold  $V(k, N)$  for computing the  $k$  largest Lyapunov exponents of an  $N$ -dimensional system. A sample of the literature on this topic can be obtained from Bridges and Reich (2001) and references therein.

It is known that the geodesic flow on the Stiefel manifold  $V(n, N)$  with a left-invariant metric is integrable for the extreme cases: when  $n = 1$ , which represents the geodesic flow on the sphere/ellipsoid, and when  $n = N$ , which represents the  $N$ -dimensional rigid body on  $SO(N)$ . Bolsinov and Jovanovic have shown that the extremal flows on Stiefel manifolds and other homogeneous spaces with bi-invariant metrics are integrable. There are two sets of integrals for such flows. The first set of integrals are the Noether integrals

$$\mathcal{F}_1 = \{h \circ \mathfrak{G}, \mathfrak{G} : TV(n, N) \rightarrow \mathfrak{so}^*(N), h : \mathfrak{so}^*(N) \rightarrow \mathbb{R}\},$$

where  $\mathfrak{G}(Q, QU) = J_Q(U) = M$ . In this case, it is easy to verify that  $M$  is conserved along the extremal flows. Without loss of generality, we may take  $\Lambda = I_N$  as the bi-invariant metric, and the extremal flows are then given by

$$\dot{Q} = QU, \quad \ddot{Q} = BQ, \quad B = B^T.$$

The body momentum is  $M = Q^T QU + U Q^T Q$ , and its derivative along the extremal flows is

$$\dot{M} = \dot{Q}^T \dot{Q} + Q^T \ddot{Q} - \ddot{Q}^T Q - \dot{Q}^T \dot{Q} = 0.$$

This is a generalization of the case of the symmetric rigid body, where  $\Lambda$  is a scalar multiple of the identity matrix. The second set of integrals for this bi-invariant case is

$$\mathcal{F}_2 = \text{all } SO(N)\text{-invariant functions on } TV(n, N).$$

The most important issue for future research is the integrability of the extremal solutions in the general case ( $1 < n < N$ ) with a left-invariant metric. Another research issue of considerable interest for numerical applications is discretization of the optimal control problem (4) based on the maximum principle, to obtain the corresponding discrete extremal flow in the states and costates. Treatment in more detail of special cases such as the rank 2 case ( $n = 2$ ) could also be carried out in the future.

### 7. Conclusions

In this paper we have presented the (continuous) geodesic flow on Stiefel manifolds with left-invariant metrics. The geodesic equations were obtained from two approaches; a variational approach taking reduced variations on the Stiefel manifold and an optimal control approach using costate variables. We have attempted to generalize the symmetric representation of the  $N$ -dimensional rigid body flow given by Bloch *et al* (2002) to geodesic flows on Stiefel manifolds.

The solution manifold in the symmetric representation of the rigid body was  $SO(N) \times SO(N)$ . We found it easier to restrict the geodesic flows on Stiefel manifolds with a left-invariant metric to the  $W_{n,N}^k$ , which are symplectic submanifolds of  $\mathbb{R}^{nN} \times \mathbb{R}^{nN}$ . The relation between  $SO(N) \times SO(N)$  and  $W_{N,N}^k$  for any fixed value of  $k$  is given by

$$SO(N) \times SO(N) = \bigcup_{m \in \{m' \mid \frac{k+m'}{2} \in SO(N)\}} W_m^k,$$

where the union is over all such  $m \in \mathfrak{so}(N)$  that satisfy  $(k + m)/2 \in SO(N)$ . We obtained the geodesic flows using the maximum principle of optimal control theory and related them to the Hamiltonian flows using the natural symplectic structure on the Stiefel manifold. We also related this optimal control formulation to the form naturally derived from variational calculus.

Note that the extremal solutions of the optimal control problem in  $W_{n,N}^0$  are a generalization of the McLachlan–Scovel equations (8) for the  $N$ -dimensional rigid body, where  $Q^T P_0$  is skew symmetric. However, these extremal solutions do not generalize the symmetric representation of the rigid body equations given in Bloch *et al* (2002), wherein  $Q^T P$  is orthogonal.

### Acknowledgments

The authors would like to thank Yuri Federov and Božidar Jovanović for helpful comments and discussions. Support from the NSF is gratefully acknowledged.

## References

- Bloch A M, Crouch P E, Marsden J E and Ratiu T S 2002 The symmetric representation of the rigid body equations and their discretization *Nonlinearity* **15** 1309–41
- Bloch A M, Baillieul J, Crouch P E and Marsden J E 2003 *Nonholonomic Mechanics and Control (Series in Interdisciplinary Applied Mathematics)* vol 24 (New York: Springer)
- Bolsinov A V and Jovanovic B 2004 Complete involutive algebras of functions on cotangent bundles of homogeneous spaces *Math. Z.* **246** 213–36
- Bridges T and Reich S 2001 Computing Lyapunov exponents on a Stiefel manifold *Physica D* **156** 219–38
- Edelman A, Arias T and Smith S T 1998 The geometry of algorithms with orthogonality constraints *SIAM J. Matrix Anal. Appl.* **20** 303–53
- Elden L and Park H 1999 A procrustes problem on the Stiefel manifold *Numer. Math.* **82** 599–619
- Federov Y N 1995 Various aspects of  $n$ -dimensional rigid body dynamics *Am. Math. Soc. Transl.* **168** 141–71
- Federov Y N 2005 Integrable flows and Backlund transformations on extended Stiefel varieties with application to the Euler top on the Lie group  $SO(3)$  *J. Nonlin. Math. Phys.* **12** 77–94 (Preprint [nlin.SI/0505045](http://nlin.SI/0505045))
- Gelfand I M and Fomin S V 2000 *Calculus of Variations* translated by R A Silverman (Mineola, NY: Dover)
- Kirk D E 2004 *Optimal Control Theory: An Introduction* (New York: Dover)
- Knorrer H 1980 Geodesics on the ellipsoid *Invent. Math.* **59** 119–43
- Knorrer H 1982 Geodesics on quadrics and a mechanical problem of C. Neumann *J. Reine Angew. Math.* **334** 69–78
- Manakov S V 1976 Note on the integration of Euler's equations of the dynamics of an  $n$ -dimensional rigid body *Funct. Anal. Appl.* **10** 253–99
- Marsden J E and Ratiu T S 1999 *Introduction to Mechanics and Symmetry* 2nd edn (New York: Springer) pp 345–8
- McLachlan R I and Scovel C 1995 Equivariant constrained symplectic integration *J. Nonlin. Sci.* **5** 233–56
- Mischenko A S and Fomenko A T 1978 Generalized Liouville method of integration of Hamiltonian systems *Funct. Anal. Appl.* **12** 113–21
- Moser J 1980 Geometry of quadrics and spectral theory *Chern Symposium 1979* (New York: Springer) pp 147–88
- Moser J and Veselov A P 1991 Discrete versions of some classical integrable systems and factorization of matrix polynomials *Commun. Math. Phys.* **139** 217–43
- Ratiu T 1980 The motion of the free  $n$ -dimensional rigid body *Indiana Univ. Math. J.* **29** 609–29