Journal of Global Optimization **23:** 309–327, 2002.

The Geometry of the Newton Method on Non-Compact Lie Groups

ROBERT MAHONY[1] and JONATHAN H. MANTON[2]
[1]*Department of Engineering, Australian National University, A.C.T., 0200, Australia (Robert.Mahony@anu.edu.au);* [2]*Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Victoria, 3010, Australia. (j.manton@ee.mu.oz.au)*

Abstract. An important class of optimization problems involve minimizing a cost function on a Lie group. In the case where the Lie group is non-compact there is no natural choice of a Riemannian metric and it is not possible to apply recent results on the optimization of functions on Riemannian manifolds. In this paper the invariant structure of a Lie group is exploited to provide a strong interpretation of a Newton iteration on a general Lie group. The paper unifies several previous algorithms proposed in the literature in a single theoretical framework. Local asymptotic quadratic convergence is proved for the algorithms considered.

1. Introduction

A fundamental optimization problem is to minimise a smooth cost function subject to a set of smooth equality constraints. Classically, such problems are solved using either a Lagrange multiplier approach (cf., for example, Fletcher, 1996) or using local coordinate charts on the manifold defined by the equality constraints (e.g., Gabay, 1982). The second approach is of particular interest for classes of constrained optimization problems in which the cost is relatively simple and much of the non-linear complexity of the problem is coded in a highly structured equality constraint. Recent work by (Udriste, 1994) considered a class of convex problems on Riemannian manifolds. General optimization algorithms on Riemannian manifolds have also been proposed by Smith (1994) and Edelman et al. (1998) among others. An important class of constrained optimization problems of this nature comes from the field of linear algebra. In the early eighties a connection between the QR-algorithm and a continuous-time dynamical system was discovered (Flashka, 1974; Deift et al. 1983) which sparked extensive research on using dynamical systems to solve linear algebraic problems (Brocket, 1989; Watkins and Elsner, 1988; Chu, 1988; Chu and Driessel, 1990; Helmke and Moore, 1994). These problems tend to involve simple, often linear or perhaps quadratic, cost functions on matrix Lie groups or homogeneous spaces. Problems of this nature are also important in the fields of linear systems theory and digital signal processing (Perkins et al., 1990; Yan et al., 1994; Gevers and Li 1993; Dehaene 1995; Liu et al., 1996; Madievski et al., 1994; Tseng et al., 1998; Bruyne et al., 1999). Following the early work on gradient flows, several authors have looked into the question of developing efficient numerical algorithms to solve such problems. A number of linearly convergent algorithms based on the geometry of the problem were developed in the early nineties (Brockett, 1993; Smith, 1993; Moore et al., 1994). Slightly later, Newton like algorithms displaying at least local quadratic convergence were developed (Smith, 1994; Mahony, 1996), although the applications considered involved either compact Lie groups or symmetric spaces. More recently, Newton like algorithms have been specialized to the important class of problems displaying orthogonality constraints (Edelman et al., 1998; Absil et al.; Manton, 2001) and to vector valued functions derived from implicit integration algorithms (Owren and Welfert, 2000). To the authors' knowledge there is no prior work aimed at unifying the Newton methods proposed or developing a full understanding of the situation on non-compact Lie groups.

In this paper we present a unifying analysis of Newton like methods on general Lie groups in terms of the geometry derived from invariant structures associated with the Lie group action. The approach taken shows the important connection between the canonical or normal coordinates on a Lie group or Riemannian manifold and the Newton iterate. In the case of a Riemannian manifold or compact Lie group this leads directly to a geometric interpretation of the Newton iteration that corresponds to the approach taken by Smith (1993) or Udriste (1994). On a non-compact Lie group it is less clear what approach should be taken. The approach we take is to define a Newton iterate with respect to local canonical coordinates of the first kind on a Lie group. These coordinates may be thought of as a replacement for normal coordinates on a Riemannian manifold. To provide a link with the geometric interpretation of the Newton method the three classical Cartan-Schouten connections are introduced. These are affine connections that ensure the canonical coordinates on a Lie group have the same geodesic properties as the normal coordinates have on a Riemannian manifold. It is shown that the symmetric Cartan-Schouten connection provides a geometric interpretation for the Newton iterate defined using local canonical coordinates. It is also shown that a Newton like algorithm defined using any of the Cartan-Schouten connections displays the local quadratic convergence characteristic of Newton algorithms. It is important to note that none of the Cartan-Schouten connections is a Levi-Civita connection associated with a Riemannian metric on a non-compact Lie group. To complete the paper, the possibility of a Riemannian geometry imposed on a non-compact Lie group is considered and it is shown that a Newton algorithm derived with respect to such a geometry is not linked to the canonical coordinates and must be analyzed independently of the Lie group structure. The theory developed unifies a class of Newton methods on general Lie groups and provides a geometric interpretation of several algorithms proposed in the literature (Gabay, 1982; Udriste, 1994; Smith, 1994; Mahony, 1996; Edelman et al., 1998; Owren and Welfert, 2000).

The paper comprises an introduction, four technical sections and a short conclusion. Section 2 reviews the Newton method on a Riemannian manifold, emphasizing the interpretation as a local iteration computed with respect to normal coordinates centred at the present estimate. Section 3 analyses the Newton method on a general Lie group. Firstly, the Newton iterate defined in terms of local canonical coordinates is reviewed. Following this the Cartan-Schouten connections are reviewed and it is shown that a Newton iterate computed with respect to local canonical coordinates corresponds to a Newton update for the torsion free (0) connection. Finally, it is shown that a Newton-like algorithm based on any of the Cartan-Schouten connections will always yield local quadratic convergence to a non-degenerate critical point. Section 4 presents the necessary results required to compute a Newton iterate with respect to a left invariant Riemannian metric on a non-compact Lie group. Section 5 provides a discussion of several of the Newton like algorithms proposed in prior literature.

2. The Newton Iteration on a Riemannian Manifold

In this section the Newton iteration on a Riemannian manifold is reviewed.

Let M be a Riemannian manifold with metric g and Levi-Civita connection ∇. Let ω be a 1-form and consider the question of finding a point $p \in M$ such that $\omega_p = 0$. If $f : M \to \mathbb{R}$ then setting $\omega = df$ and finding a zero of ω is equivalent to finding a critical point of f. The general Taylor expansion of a 1-form in multi-variable calculus is a complicated formulae (cf. Eq. (5)) where the higher order contributions depend on cross terms associated with the non-commutativity of differentiation on an arbitrary manifold. However, it is only necessary to compute the linear approximation of the 1-form ω in order to determine a Newton iteration. This corresponds to computing a single variable Taylor expansion of ω along a geodesic γ_X in M with initial velocity $X \in T_pM$. Denote parallel transport along a curve $\gamma_X : [0, 1] \to M$, $\gamma_x(o) = p$, by $\mathbf{P}_{\gamma_X} : T_{\gamma_X(0)}M \to T_{\gamma_X(1)}M$. Then one has (cf. Smith, 1993; Udriste, 1994)

$$\mathbf{P}_{\gamma_X}^{-1}\omega_{\gamma_X(1)} = \omega_p + (\nabla_X\omega)_p + \mathbf{O}(|X|^2). \tag{1}$$

Note that $\mathbf{P}_{\gamma_X}^{-1}\omega_{\gamma_X(1)} \in T_p^*M$ is a covector at p for all $X \in T_pM$. The Taylor expansion along γ_X may be computed using standard multi-variable calculus in the dual finite dimensional vector spaces T_pM and T_p^*M and corresponds, up to second order terms, to the full Taylor expansion on M. The fact that the derivatives correspond to covariant differentiation is a consequence of computing the Taylor expansion along geodesics. The Newton iteration is computed by discarding the $\mathbf{O}(|X|^2)$ term and solving for the vector $X \in T_pM$ for which the linear estimate of the 1-form ω is zero, namely

$$(\nabla_X\omega)_p = -\omega_p. \tag{2}$$

Since the Taylor expansion was computed along geodesics, the Newton iteration is a unit length step along the geodesic emanating from the present estimate $p \in M$ in direction X.

In the Euclidean case, it is usual to express the Newton method in terms of a Hessian. The concept of Hessian, however, does not generalize in a unique manner to an arbitrary manifold M. The following material discusses the two most natural geometric definitions of a Hessian on a Riemannian manifold and their interrelationship. Define the *geometric Hessian* with respect to a connection ∇ to be $\mathrm{Hess} f \in \mathfrak{T}_2$ (a (0,2) tensor field):

$$\mathrm{Hess} f(X, Y) = \nabla_X \nabla_Y f = (\nabla_X df)Y. \tag{3}$$

If the connection is symmetric then $\mathrm{Hess} f(X, Y) = \mathrm{Hess} f(Y, X)$. This is true for the Levi-Civita connection on a Riemannian manifold but is not necessarily the case for the affine connections considered in Section 3 on non-compact Lie-groups. Contracting the Hessian with a single vector field leads to a covector $\mathrm{Hess} f(X, \cdot) = \nabla_X df \in \mathfrak{T}_1$. In the case that $\omega = df$ is an exact 1-form the Newton iterate Eq. (2) may be rewritten as

$$\mathrm{Hess} f(X, \cdot) = -df. \tag{4}$$

This is the geometric derivation of the Newton iterate on a Riemannian manifold found in most developments (Smith 1994; Udriste, 1994).

The other Hessian commonly used on a Riemannian manifold is the $\mathfrak{T}_1^1$ tensor $\mathrm{Hess}_R f = \nabla \mathrm{grad} f$ (do Carmo, 1992). Since the definition of the gradient vector field depends on the existence of a Riemannian metric this Hessian can only be defined on a Riemannian manifold and is always computed using the Levi-Civita connection. We term this Hessian the *Riemannian Hessian* of a function f. The Riemannian Hessian may be thought of as a map

$$\mathrm{Hess}_R f : T_p M \to T_p M, \quad \mathrm{Hess}_R f(X) \mapsto \nabla_X \mathrm{grad} f.$$

This is simply a matrix vector space mapping. Since the metric is invariant with respect to the connection, the local coordinate representation of the Riemannian Hessian is computed by raising an index of the geometric Hessian, that is pre-multiplying the local coordinate expression for the geometric Hessian by the inverse of the matrix representing the metric. The gradient $\mathrm{grad} f$ is similarly obtained by pre-multiplication of the differential df by the inverse of the metric. Thus the geometric Hessian and the Riemannian Hessian are two sides of the same coin on a Riemannian manifold.

Pre-multiplying Eq. (4) by the inverse metric shows that the Newton iterate Eq. (4) with respect to the Levi-Civita connection can be rewritten in terms of the Riemannian Hessian as

$$\mathrm{Hess}_R f(X) = -\mathrm{grad} f.$$

This form of the Newton iterate, which is used in several contemporary works (Gabay, 1982; Udriste, 1994), is equivalent to Eqs (2) and (4).

The Newton iterate on a Riemannian manifold can also be expressed in terms of normal coordinates. Let $\{E_i\}$ be an orthonormal basis for T_pM. Denote the geodesic map on M by $\exp_p(X)$ that takes $X = \sum u^i E_i \in T_pM$ to the point $\gamma_X(1)$ where $\gamma_X(\tau)$ is the geodesic curve passing through p and having tangent vector $X = \sum u^i E_i$. The coordinates $\varphi(\exp_p(\sum u^i E_i)) \mapsto (u^1, \ldots, u^n)$ are termed *normal coordinates* on a Riemannian manifold. These local coordinates are closely related to the approximation Eq. (1). To see this, it is instructive to consider the case of an exact 1-form $\omega = df$ and compute local coordinate expressions for the Newton iterate Eq. (2). Due to the particular structure of normal coordinates around a point p one has (Boothby, 1986)

$$\left(\Gamma^k_{ij}\right)_p = 0, \quad (g_{ij})_p = \delta_{ij}$$

where $\left(\Gamma^k_{ij}\right)_p$ and $(g_{ij})_p$ are the local coordinate representation of the Christoffel symbols and the metric respectively and δ_{ij} is the Kronecker delta function. It follows that $\mathrm{Hess} f$ and df are the Euclidean Hessian and gradient respectively of the local coordinate representation $\tilde{f}(u) = f(\exp_p(\sum u^i E_i))$ of the function f with respect to the normal coordinates. We write

$$\mathrm{Hess} f = H\tilde{f}, \quad df = D\tilde{f}$$

where

$$H\tilde{f} = \left.\frac{\partial^2 \tilde{f}}{\partial u^i \partial u^j}\right|_{u=0}, \quad D\tilde{f} = \left.\frac{\partial \tilde{f}}{\partial u^i}\right|_{u=0}.$$

The Newton iterate Eq. (4) in local coordinates is the familiar matrix equation

$$H\tilde{f}X = -D\tilde{f}, \quad X \in \mathbb{R}^n.$$

A local coordinate interpretation of the Newton method is the classical alternative to a Lagrange multiplier approach for solving constrained optimization problems.

There are two complimentary interpretations of the Newton method on a Riemannian manifold. A geometric perspective is to consider Eq. (2), depending on the choice of affine connection as the key equation in the Newton iterate and the interpretation in terms of the normal coordinates as a secondary phenomena following from the properties of geodesics. Alternatively, one may directly define the Newton iterate with respect to a particular set of local coordinates (the normal coordinates). The representation of the iteration as a covariant derivative (Eq. (2)) is then considered as a secondary structure that is used to provide an elegant representation of the iteration. In fact, the full Taylor expansion associated with the approximation in Eq. (1) is just a coordinate free expression of the local coordinate Taylor expansion of $\tilde{f}$. Clearly, the two interpretations differ only in a qualitative sense and lead to the same Newton iterate and the same analysis. However, the second perspective is important and leads naturally to the discussion in Section 3.

REMARK 2.1. Taking the point of view that the geometry is secondary, then *a priori* there appears to be no particular reason why the normal coordinates will have better numerical properties than any other local coordinates. Following this line of reasoning leads one to consider what are the best local coordinates in which to express a function f in order to compute its minima. Such an approach motivated the work of Botsaris in the early eighties (Botsaris, 1978, 1981a,b). □

REMARK 2.2. Not only 1-forms can be used as the basic iterates of a Newton method. Let $Z \in \mathfrak{X}(M)$ be a smooth vector field on M. Then

$$\mathbf{P}_{\gamma_X}^{-1} Z_{\gamma_X(1)} = Z_p + (\nabla_X Z)_p + \mathbf{O}(|X|^2),$$

where $\gamma_X(t) := \exp_p(tX)$, $t \in [0, 1]$ and $X \in T_pM$. The object $\nabla_X Z$ is a $\mathfrak{T}_1^1$ tensor similar to the Riemannian Hessian. Indeed if $Z = \operatorname{grad} f$ is chosen to be the gradient of a cost function on a Riemannian manifold then this approach leads to the same equations considered above. On a non-compact Lie-group it is usually impossible to find a Riemannian metric and Levi-Civita connection that have 'nice' properties and the geometric Hessian interpretation will be of considerably more use to our development than that of the Riemannian Hessian. In the case where the Newton method is used as an algorithm to compute a stationary point of a vector field $Z \in \mathfrak{X}(M)$ then the above formulae is of direct interest (cf. §5.2). □

3. The Newton iteration on a Lie group

In this section several Newton algorithms on a general Lie group are proposed and analyzed. Since a non-compact Lie group does not have a natural Riemannian structure, the elegant correspondence of geometric and local coordinate approaches of the previous section is somewhat more complicated. A prototype Newton algorithm is proposed based on canonical local coordinates analogous to the normal coordinates considered on a Riemannian manifold in Section 2. This is shown to display the characteristic local quadratic convergence properties of a classical Newton method. To provide a geometric interpretation of the algorithm the three classical Cartan-Schouten connections are introduced. These connections ensure the canonical coordinates on a Lie group have the same geodesic properties as the normal coordinates have on a Riemannian manifold. It is shown that the symmetric Cartan-Schouten (0) connection provides a geometric interpretation for the Newton iterate defined using local canonical coordinates analogous to the situation on a Riemannian manifold with the Levi-Civita connection. However, the geometric Newton methods defined using any of the Cartan-Schouten connections are also of interest and it is shown that these algorithms also display local quadratic convergence. A discussion of Riemannian geometry on a Lie group is deferred to Section 4.

A Lie group G is an abstract group which is also a smooth manifold on which the operations of group multiplication ($\tau \mapsto \sigma\tau$, for $\sigma, \tau \in G$) and inversion

($\tau \mapsto \tau^{-1}$, for $\tau \in G$) are smooth diffeomorphisms of G onto G. For $g \in G$, let $L_g(\sigma) = g\sigma$ denote left multiplication by a fixed element g. Right translation R_g is defined in an analogous manner. Denote the vector space of smooth vector fields on G by $\mathfrak{X}(G)$ (or just $\mathfrak{X}$). A vector field $X \in \mathfrak{X}$ is said to be left (right) invariant if $X_\sigma = dL_\sigma x$ ($X_\sigma = dR_\sigma x$) for $x \in T_eG$ an element of the identity tangent space of G. The set of left (right) invariant vector fields is a finite dimensional Lie algebra, denoted by $\mathfrak{g}$, under the standard Lie-bracket operation for vector fields. There is a one-to-one correspondence between left invariant vector fields in $\mathfrak{g}$ (generally written in upper case) and elements of T_eG (generally written in lower case).

A 1-parameter subgroup of a Lie-group G is a smooth subgroup $\gamma_X(t)$ parameterized by a scalar $t \in \mathbb{R}$. The map

$$\exp : \mathfrak{g} \times \mathbb{R} \to G, \quad (X, t) \mapsto \gamma_X(t) := \exp(tX)$$

is the unique homomorphism from the Lie-algebra to 1-parameter subgroups of G (Warner, 1983, p. 101). If G is a matrix group then this corresponds to the algebraic matrix exponential

$$\exp(tX) = e^{tX} = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k.$$

where $X \in \mathbb{R}^{n\times n}$. Let $E_1, \dots, E_n$ be a basis for the Lie algebra $\mathfrak{g}$ and let $e_1, \dots, e_n \in T_eG$ be the corresponding basis in the identity tangent space of G. Canonical coordinates of the first kind provide a local coordinate chart around a point $\sigma \in G$:

$$\varphi\left(L_\sigma \exp\left(\sum_{i=1}^{n} u^i e_i\right)\right) \mapsto u = (u^1, \dots, u^n).$$

Let $f \in \mathcal{C}^\infty(G)$ be a smooth function on a Lie group G. The Taylor expansion of f with respect to canonical coordinates around the point σ may be computed using standard multi-variable calculus (cf., for example, Fleming, 1977). On a general Lie group one has (Varadarajan, 1984, p. 96)

$$f\left(\sigma \exp(u^1 e_1 + \cdots + u^n e_n)\right) = f(\sigma) + \sum_{i=1}^{n} u^i E_i f(\sigma) + \frac{1}{4}\sum_{i,j=1}^{n} u^i u^j (E_i E_j + E_j E_i) f(\sigma) + \mathbf{O}(|u|^3) \tag{5}$$

Equation (5) provides a quadratic model of f in terms of the canonical coordinates. The quadratic term is linked to the differential operators $H_{ij} : \mathcal{C}^\infty(G) \to \mathcal{C}^\infty(G)$, $i, j = 1, \dots, n$,

$$H_{ij} f(\sigma) := \frac{1}{2}(E_i E_j + E_j E_i) f(\sigma) = \frac{1}{2}\left[d\left(df \cdot E_j\right) \cdot E_i + d\left(df \cdot E_i\right) \cdot E_j\right].$$

In local coordinates this is simply the second derivative

$$H_{ij} f_\sigma = \frac{\partial^2}{\partial u^i \partial u^j} f(\sigma \exp(u^1 e_1 + \cdots + u^n e_n)).$$

For a fixed choice of left invariant vector fields $\{E_i\}$ define

$$Hf = \left[H_{ij} f\right] = \frac{1}{2}\left[(E_i E_j + E_j E_i) f\right] \in \mathbb{R}^{n\times n} \tag{6}$$

to be the matrix of differential operators $H_{ij} f$. We term this matrix the *differential Hessian* for the canonical coordinates at the point σ. Note that the definition depends on the choice of $\{E_i\}$ or equivalently on the representation of the canonical coordinates used. Based on the differential Hessian a Newton iteration is defined in the local canonical coordinates:

ALGORITHM 3.1 (Newton-Raphson Algorithm on a Lie-group G).
Given $\sigma_k \in G$ *compute* $df_{\sigma_k} = \frac{\partial}{\partial u^i} f(\sigma_k \exp(u^1 e_1 + \cdots + u^n e_n))\big|_{u=0}$.
Compute the differential Hessian matrix $(Hf)_{\sigma_k}$.
Set $u = -(Hf)^{-1}_{\sigma_k} df_{\sigma_k}$.
Set $\sigma_{k+1} = L_{\sigma_k} \exp(u_1 e_1 + \cdots + u_n e_n)$.
Set $k = k + 1$ *and repeat.* □

REMARK 3.2. Analogously to the familiar matrix version of the Newton algorithm on Euclidean space, Algorithm 3.1 is independent of the basis $\{e_1, \ldots, e_n\}$ of $\mathfrak{g}$ used. This follows from the correspondence of Hf to the geometric Hessian Hess$f$ for the Cartan-Schouten (0) connection (cf. Eq. (8)) discussed later in the section.

PROPOSITION 3.3. *Let* $f : G \to \mathbb{R}$ *be a smooth function on a Lie-group G and* $\mu \in G$ *be a non-degenerate critical point of* f. *Let* $e_1, \ldots, e_n \in \mathfrak{g}$ *be a basis for the Lie algebra of* G. *Then Algorithm 3.1 is locally quadratically convergent to* μ.

Sketch of Proof. Only a sketch of the proof is included in the present paper (based on the approach in Mahony (1996) (c.f. also Smith, 1994; Owren and Welfert, 2000)). In order to measure the rate of convergence of Algorithm 3.1 it is necessary to introduce a measure of distance, $|\cdot|_\mu$, around the point μ on the Lie group. The distance measure used is the Euclidean norm defined on the local canonical coordinates of the first kind centred at the point μ. It is known that for any $\mu \in G$ there exists a neighbourhood of μ such that for any σ_k in the neighbourhood there exists $X_k \in \mathfrak{g}$ with $\sigma_k = \mu \exp(X_k)$ (Warner, 1983). Moreover, if the exponential is restricted to a neighbourhood of the origin in $\mathfrak{g}$ then this correspondence is unique. Thus, for any point σ_k in the neighbourhood of μ we denote its distance $|\sigma_k|_\mu = |\mu \exp(X_k)|_\mu := |X_k|$ where $|X_k| = \sqrt{X_k^T X_k}$. Denote

$$Z_{k+1} = u_1(k+1)e_1 + \cdots + u_n(k+1)e_n$$

for $\{u(k+1)\}$ determined by Algorithm 3.1. Following the standard approach used in proving quadratic convergence of the Newton algorithm for Euclidean space (Fletcher, 1996; Dennis and Schnabel, 1983) one obtains the bounds

$$|Z_{k+1}| \leqslant 2|X_k| + \mathbf{O}(|X_k|^2), \quad |X_k + Z_{k+1}| = \mathbf{O}(|X_k|^2).$$

In Algorithm 3.1 the update step is

$$\sigma_{k+1} = \mu \exp(X_{k+1}) = \sigma_k \exp(Z_{k+1}) = \mu \exp(X_k) \exp(Z_{k+1}).$$

To link X_{k+1} to $X_k + Z_{k+1}$, Dynkins formulae may be used (Varadarajan, 1984, p. 7) (Helgason, 1978, p. 106) $\exp(X)\exp(Y) = \exp((X+Y) + \mathbf{O}(|X|\ |Y|))$, $X, Y \in \mathfrak{g}$. Combining this approximation with the standard bounds one obtains

$$|\sigma_{k+1}|_\mu = |X_{k+1}| = |X_k + Z_{k+1}| + \mathbf{O}(|X_k|\,|Z_{k+1}|) \leqslant K|X_k|^2 = K|\sigma_k|_\mu^2$$

for some constant $K > 0$. Choosing the neighbourhood around μ such that $|\sigma_k|_\mu = |X_k| \leqslant 1/\sqrt{2K}$ ensures that the quadratic bound obtained also leads to local convergence. □

It is of interest to determine whether the Newton iteration on Lie groups, proposed above, is connected to any 'natural' geometric structure on a Lie group. In particular, it is desirable to find an affine connection such that the Newton iteration may be written in terms of a geometric Hessian. In order that this is true it is important that the local canonical coordinates have the same geodesic properties with respect to the affine connection considered as the normal coordinates had with respect to the Levi-Civita connection. The Cartan-Schouten connections considered below have this property.

Let $\psi : G \to G$ be a diffeomorphism of a Lie group G. An affine connection ∇ on G is invariant under ψ if

$$d\psi \nabla_X Y = \nabla_{d\psi X} d\psi Y.$$

An affine connection ∇ is termed *left invariant* on G if it is invariant under left translation L_g for arbitrary $g \in G$. A similar definition holds for *right invariant* affine connections and an affine connection that is both left and right invariant is prosaically termed *bi-invariant*. There is a one-to-one correspondence between left invariant affine connections on G and bi-linear maps $\omega : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, given by

$$\omega(Y, Z) = (\nabla_{dL_\sigma Y} dL_\sigma Z)(e), \tag{7}$$

for $Y, Z \in \mathfrak{g}$. The unique bilinear map ω is termed the *connection function* for ∇. Let ∇ be a left invariant affine connection and consider a 1-parameter subgroup $\gamma_x(t) = \exp(tx)$. Then

$$\nabla_{\dot\gamma_x(t)} \dot\gamma_x(t) = \nabla_{dL_{\exp(tx)}x} dL_{\exp(tx)} x = dL_{\exp(tx)} \omega(x, x).$$

It follows that a one parameter subgroup of G is a geodesic with respect to ∇ if and only if $\omega(x, x) = 0$.

There are three important bi-invariant affine connections on any connected Lie group that have the property $w(x, x) = 0$. They are known as the Cartan-Schouten connections and are associated with certain invariant structures on the Lie group. The following definition is based on Nomizu (1954).

(i) The Cartan-Schouten (0) connection is the left-invariant affine connection associated with the connection function

$$\omega(x, y) = \frac{1}{2}[x, y].$$

It is the unique affine connection that has torsion tensor zero and for which all 1-parameter subgroups are geodesics.

(ii) The Cartan-Schouten (–) connection is the left-invariant affine connection associated with the connection function

$$\omega(x, y) = 0.$$

It is the unique affine connection for which parallel transport of a vector along a 1-parameter subgroup corresponds to transformation by left multiplication

$$\mathbf{P}_{\gamma_x} y = dL_{\gamma_x(1)} y$$

for $\gamma_x(t) := \exp(tx)$, $t \in [0, 1]$ and $x, y \in T_e G$.

(iii) The Cartan-Schouten (+) connection is the left-invariant affine connection associated with the connection function

$$\omega(x, y) = [x, y].$$

It is the unique left invariant affine connection for which parallel transport of a vector along a 1-parameter subgroup corresponds to transformation by right multiplication.

$$\mathbf{P}_{\gamma_x} y = dR_{\gamma_x(1)} y$$

for $\gamma_x(t) := \exp(tx)$, $t \in [0, 1]$ and $x, y \in T_e G$.

Define $\mathrm{Ad}_\sigma : \mathfrak{g} \to \mathfrak{g}$ to be the automorphism of $\mathfrak{g}$ given by $Ad_g x = dL_g dR_{g-1} x$ (Warner, 1983). To see that the Cartan-Schouten connections are bi-invariant it is sufficient to observe that for any $g \in G$ and $x, y \in \mathfrak{g}$

$$\begin{aligned}
\nabla_{dR_g x} dR_g y &= dL_g \nabla_{(dL_{g^{-1}} dR_g x)} (dL_{g^{-1}} dR_g y) \\
&= dL_g \omega(\mathrm{Ad}_{g^{-1}} x, \mathrm{Ad}_{g^{-1}} y) \\
&= dL_g \mathrm{Ad}_{g^{-1}} \omega(x, y) \\
&\qquad \text{(true for } (-), (0), (+) \text{ Cartan-Schouten connections)} \\
&= dR_g \omega(x, y) = dR_g \nabla_X Y,
\end{aligned}$$

where X, Y are the left invariant vector fields associated with $x, y \in \mathfrak{g}$.

On a general non-compact Lie group there is no bi-invariant Riemannian metric (Boothby, 1986) and consequently there is no natural choice of associated geometry given by the Levi-Civita connection. The three Cartan-Schouten connections combine bi-invariance with a link between geodesics and 1-parameter subgroups. Both aspects are important from a numerical perspective; the first since the derivative is invariant under adjoint automorphisms of the Lie algebra⋆ while the link between geodesics and 1-parameter subgroups is important from a numerical perspective during the implementation of a Newton algorithm.

It is clearly of interest to investigate whether the geometric Hessian with respect to the Cartan-Schouten connections is related to the differential Hessian with respect to canonical coordinates of the first kind. For any left invariant connection and left invariant vector fields $X, Y \in \mathfrak{g}$ one has

$$\begin{aligned} XYf = X(dfY) &= \nabla_X (dfY) = (\nabla_X df)\, Y + df \nabla_X Y \\ &= \mathrm{Hess} f(X, Y) + \omega(X, Y) f \end{aligned}$$

and hence

$$\mathrm{Hess} f(X, Y) = (XY - \omega(X, Y))\, f$$

as a differential operator on a function f.

Let ∇ be the Cartan-Schouten (0) connection then

$$\mathrm{Hess} f(X, Y) = \left(XY - \frac{1}{2}[X, Y] \right) f = \frac{1}{2} (XY + YX)\, f.$$

Writing $X = \sum_{i=1}^n x^i E_i$, $Y = \sum_{i=1}^n y^i E_i$ it is clear that

$$\mathrm{Hess} f(X, Y) = X^T H f Y \tag{8}$$

with respect to canonical coordinates of the first kind. Thus, the Taylor expansion Eq. (5) may be written

$$f\,(L_\sigma \exp(X)) = f(\sigma) + df\,(X) + \frac{1}{2}\mathrm{Hess} f(X, X) + \mathbf{O}(|u|^3)$$

where $X = u^1 E_1 + \cdots + u^n E_n$. The Cartan-Schouten (0) connection plays the same role for the Newton iteration Algorithm 3.1 on an arbitrary non-compact Lie group as the Levi-Civita connection played for the Newton iteration Eq. (4) on a Riemannian manifold.

The other two Cartan-Schouten connections are also of considerable interest. A direct consequence of the simple expression for parallel transport is that both the

⋆ This property is important when the concepts studied in this paper are extended to homogeneous spaces. Such an extension is beyond the scope of the present paper.

$(-)$ and $(+)$ connections have zero curvature on the Lie group. The torsion of a connection, given by

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y],$$

is an anti-symmetric tensor $T(X, Y) = -T(Y, X)$. The Torsion measures the non-symmetry of a connection. For the (0) connection the torsion $T(X, Y) = 0$ and the connection is said to be symmetric. In the case of the $(-)$ connection the torsion is $T(X, Y) = +[X, Y]$ while the $(+)$ connection has torsion $T(X, Y) = +[X, Y]$.

For the $(-)$ connection one has

$$\begin{aligned}\text{Hess}^{(-)} f(X, Y) = XYf &= \frac{1}{2}(XY + YX) f + \frac{1}{2}[X, Y]f \\ &= X^T H f Y - \frac{1}{2} T(X, Y) f.\end{aligned}$$

It follows that the geometric Hessian computed with respect to this connection will not be symmetric. The linear Taylor expansion of df (along a 1-parameter subgroup $\gamma_x(t) = \exp(tx)$) constructed using the parallelism of the $(-)$ connection is

$$\begin{aligned}\mathbf{P}_{\gamma_x}^{-1} df &= dL_{\exp(-X)} df \\ &= f(\sigma) + df \cdot X + \frac{1}{2}\text{Hess}^{(-)} f(X, X) + \mathbf{O}(|X|^3). \\ &= f(\sigma) + df \cdot X + \frac{1}{2}\left(X^T H f X - \frac{1}{2} df \cdot T(X, X)\right) + \mathbf{O}(|X|^3). \\ &= f(\sigma) + df \cdot X + \frac{1}{2} X^T H f X + \mathbf{O}(|X|^3). \end{aligned} \tag{9}$$

Due to the anti-symmetry of the torsion it does not contribute to the quadratic approximation. The quadratic approximation obtained above, based on the parallelism of the (–) connection, is equivalent to that based on the (0) connection. An analogous relation to Eq. (9) is valid for the Cartan-Schouten (+) connection. Indeed, any connection with geodesics given by the 1-parameter subgroups varies from the (0) connection only in its torsion. Thus, the underlying quadratic approximation, based on the symmetric part of the connection, depends only on the canonical coordinates.

From the above discussion it appears that the (0) connection may be regarded as the natural geometry to use when minimizing a function on a Lie group. However, this does not necessarily mean that the (0) connection is best from a numerical perspective. Indeed, in situations where the symmetry of the problem is closely linked to the geometry of the $(-)$ or (+) connection the torsion may act to improve the numerical performance of the scheme. An example of such a situation is discussed in Section 5.2. In such situations it is possible to work with a non-symmetric connection directly and compute a ‘Newton’ update according to the geometric formulae

$$Z_{k+1} = \text{Hess} f^{-1} df\big|_{\sigma_k}, \quad \sigma_{k+1} = L_{\sigma_k} \exp(Z_{k+1}). \tag{10}$$

For a connection with torsion this update formulae does not minimise the truncated Taylor expansion (5) due to the non-symmetric component of the Hessian. Nevertheless, the following lemma shows that the iterate given by Eq. (10) displays local quadratic convergence.

LEMMA 3.4. *Let $f : G \to \mathbb{R}$ be a smooth function on a Lie-group G and $\mu \in G$ be a non-degenerate critical point of f. Let ∇ be any affine connection on G such that the 1-parameter subgroups of G are geodesics. Then the iteration Eq. (10) generates a sequence that is locally quadratically convergent to μ.*

Proof. Write the update direction $Z_{k+1} \in \mathfrak{g}$ as

$$Z_{k+1} = Z'_{k+1} + \Delta$$

where Z'_{k+1} is the minimizer of the quadratic approximation of f with respect to canonical coordinates (cf. Alg. 3.1). We prove that $|\Delta| = \mathbf{O}(|X_k|^2)$ where $\sigma_k = \mu \exp(X_k)$. For all $Y \in \mathfrak{g}$ one has

$$\begin{aligned}
&\text{Hess} f_{\sigma_k}(Z_{k+1}, Y) = -df_{\sigma_k} \cdot Y,\\
&Y^T Hf Z_{k+1} + \frac{1}{2} df T(Z_{k+1}, Y) = -df_{\sigma_k} \cdot Y,\\
&Y^T Hf \Delta + \frac{1}{2} df T(\Delta, Y) = -\frac{1}{2} df T(Z'_{k+1}, Y).
\end{aligned} \tag{11}$$

Since μ is a critical point and f is smooth then in a neighbourhood of μ one has $df = \mathbf{O}(|X_k|)$. Exploiting the fact that μ is non-degenerate ($Hf > 0$) the linear operator acting on Δ on the left hand side of Eq. (11) is non-degenerate in a neighbourhood of μ and its inverse is well defined and bounded. Considering the right hand side of Eq. (11) it is clear that

$$|\Delta| = \mathbf{O}(|df|\,|Z'_{k+1}|) = \mathbf{O}(|X_k|^2).$$

where the final asymptotic approximation uses the asymptotic bound $|Z'_{k+1}| = \mathbf{O}(|X_k|)$ derived in the proof of Proposition 3.3.

Considering the iteration in canonical coordinates (and using Dynkin's formulae) one has

$$\begin{aligned}
|X_{k+1}| &= |X_k + Z_{k+1}| + \mathbf{O}(|X_k|\,|Z_{k+1}|)\\
&\leqslant |X_k + Z'_{k+1}| + |\Delta| + \mathbf{O}(|X_k|\,|Z'_{k+1}| + |X|\,|\Delta|) \leqslant K(|X_k|^2),
\end{aligned}$$

for some $K > 0$. Here the first term was shown to be $\mathbf{O}(|X_k|^2)$ in the proof of Proposition 3.3, the second term was considered above and the final two terms are direct from the earlier discussion. Choosing the neighbourhood around μ such that $|X_k| \leqslant 1/\sqrt{2K}$ ensures that the quadratic bound obtained also leads to local convergence and completes the proof. □

4. Riemannian geometry on a Lie group

In the previous section, a Newton iterate on an arbitrary Lie group G was proposed and analyzed based on an interpretation in terms of canonical local coordinates of the first kind. In this section, the case where a left or right invariant Riemannian metric is assigned to a Lie group is considered. An arbitrary Riemannian structure will have no connection to the Lie group structure and is not relevant to the goals of this paper. The reader is referred to Nomizu (1954) for background material on Riemannian geometry on a Lie group.

Let G be a general Lie group. Any inner product on $\mathfrak{g}$ generates a Riemannian metric on G via

$$\langle X_\sigma, Y_\sigma \rangle = \langle dL_{\sigma^{-1}} X_\sigma, dL_{\sigma^{-1}} Y_\sigma \rangle_e .$$

A similar construction may be made using right invariance. If a Riemannian metric is both left and right invariant it is termed bi-invariant. A compact Lie group admits a bi-invariant metric (Boothby, 1986), whereas a non-compact Lie group does not in general admit a bi-invariant metric.

For an arbitrary left invariant Riemannian metric on G let ∇ be the Levi-Civita connection. It is easily verified that ∇ is itself left invariant. Let the connection function of ∇ be denoted

$$\alpha(x, y) := U(x, y) + \frac{1}{2}[x, y] \tag{12}$$

where $U(x, y)$ is a bi-linear function that measures the difference between the Levi-Civita connection function and the Cartan-Schouten (0) connection. Since the Levi-Civita connection has zero torsion $\alpha(x, y) - \alpha(y, x) = [x, y]$ it follows that $U(x, y) = U(y, x)$. Let $X, Y, Z \in \mathfrak{X}$ be left invariant vector fields on G corresponding to elements $x, y, z \in \mathfrak{g}$. Then, as a consequence of the left invariance of the metric

$$\nabla_Z \langle X, Y \rangle = 0 = \langle \alpha(z, x), y \rangle + \langle x, \alpha(z, y) \rangle .$$

Substituting for $\alpha(x, y)$ one obtains

$$\langle U(z, x), y \rangle + \langle x, U(z, y) \rangle = \frac{1}{2}\Big(\langle [x, z], y \rangle + \langle [y, z], x \rangle \Big).$$

Permuting the indices of this identity cyclically and adding and subtracting yields

$$\langle U(x, y), z \rangle = \frac{1}{2} \left(\langle [z, x], y \rangle + \langle x, [z, y] \rangle \right) . \tag{13}$$

This uniquely determines the bi-linear function $U(x, y)$ and consequently the connection function Eq. (12). In particular, the Levi-Civita connection is equal to the Cartan-Schouten (0) connection if and only if $U(x, y)$ is zero.

An important observation is that if U is non-zero then there exists an $x \in \mathfrak{g}$ such that the one parameter subgroup $\exp(tx)$ is not a geodesic on G (choose x a non-degenerate eigenvector of U, then $\alpha(x, x) \neq 0$). As a consequence, canonical coordinates do not provide a useful structure in which to analyze the performance of a Newton algorithm derived with respect to a Riemannian geometric structure on a non-compact Lie-group. Once the geodesics of ∇ have been computed, however, the Newton algorithm with respect to the geometric Hessian will be equivalent to the Newton method in the normal Riemannian coordinates centred on the present iterate. The situation is entirely analogous to that considered in Section 2.

In the case where the Cartan-Schouten (0) connection corresponds to the Levi-Civita connection one has that $U(x, y) \equiv 0$. A necessary and sufficient condition that $U(x, y) \equiv 0$ is that

$$\langle [z, x], y\rangle + \langle x, [z, y]\rangle = 0. \tag{14}$$

This is true for any bi-invariant metric since such metrics are invariant under the Adjoint action

$$\langle \mathrm{Ad}_{\exp(tz)}x, \mathrm{Ad}_{\exp(tz)}y\rangle = \langle x, y\rangle. \tag{15}$$

For a given $y \in \mathfrak{g}$ the differential of $\mathrm{Ad}_\sigma y$ with respect to σ is denoted $\mathrm{ad}_\star y$ and corresponds to the Lie bracket of the Lie algebra $\mathfrak{g}$, $\mathrm{ad}_x y = D\mathrm{Ad}_\star y[x] = [x, y]$. Thus, taking the time-derivative of Eq. (15) yields condition Eq. (14). In particular, the connection function for the Levi-Civita connection associated with a bi-invariant metric is always $\alpha(x, y) = \frac{1}{2}[x, y]$.

From the above analysis it follows that on a general Lie group, the Riemannian approach corresponds to the approach proposed in Section 3 if and only if the metric considered is bi-invariant. In general, the computational cost of evaluating geodesics for a Levi-Civita connection on a non-compact Lie group is excessive. For this reason the authors recommend working with the Cartan-Schouten connections unless there are very specific reasons for using a Riemannian structure on a non-compact Lie group.

5. Some examples

Recent literature contains a number of applications of Newton like methods to problems defined on Lie groups. Two examples of such problems are reviewed in Sections 5.1 and 5.2 to provide context for the results presented in Sections 2–4. Although the first example is an application on a compact Lie group it provides an excellent comparison of a Newton method derived using the (0) connection to an algorithm derived using the (–) connection with links to published results. The second example has direct applications to algorithms on non-compact Lie-groups, for example, implicit Euler iterations for rigid body motion on $SE(3)$.

5.1. THE SYMMETRIC EIGENVALUE PROBLEM

Let $SO(n)$ denote the set of special orthogonal matrices, $U \in \mathbb{R}^{n\times n}$ such that $U^T U = I_n$ the identity and det(U)=1. The determination of the eigenvalues of a symmetric real matrix H may be achieved by minimizing the cost

$$\phi(U) := \mathrm{tr}(U^T H U N), \quad N = \mathrm{diag}(1, 2, \dots, n)$$

over $U \in SO(n)$ (Brockett, 1989; Chu, 1988; Chu and Driessel, 1990; Helmke and Moore, 1994). Two different Newton methods for this problem were proposed by Smith (1993, 1994) and Mahony (1994, 1996). The algorithm proposed by Smith is based on the Levi-Civita connection associated with the bi-invariant Killing form on $SO(n)$. The geometric Hessian generated in this manner is symmetric and equal to the differential Hessian computed with respect to the canonical coordinates of the first kind.

The development proposed by Mahony (1994, 1996) may be interpreted in terms of the (–) connection. To recover the equations proposed by Mahony (1996) observe that any left invariant metric is invariant with respect to the (–) connection.* The geometric Hessian for the (-) connection may be written

$$\mathrm{Hess}^{(-)}\phi(Y, X) = Y(X\phi) = d(X\phi)\cdot Y = g(\mathrm{grad}(X\phi), Y).$$

This leads directly to the Newton iteration equation proposed in Mahony (1996): solve $0 = \mathrm{grad}\phi + \mathrm{grad}(X\phi)$ for $X \in T_\sigma G$. The algorithm proposed by Mahony was shown to display local quadratic convergence.

An interesting observation is that the cost ϕ is symmetric about the critical point (Smith, 1994) and the cubic terms of the approximation in canonical local coordinates are zero. As a consequence the Newton algorithm based on the (0) connection, and related to the differential Hessian, displays cubic convergence (Smith, 1994). In contrast, for the algorithm based on the (–) connection the Hessian contains non-symmetric terms (related to the torsion) that decrease only quadratically in the neighbourhood of the critical point. In practice, the convergence observed for the (0) connection algorithm (Smith, 1994) is cubic while that observed for the (–) connection algorithm (Mahony, 1996) is only quadratic.

5.2. IMPLICIT INTEGRATION ALGORITHMS

In recent work, Owren and Welfert (2000) have proposed two Newton type meth-

* Let g be a left invariant metric on G and choose $X, Y, Z \in \mathfrak{X}$ left invariant vector fields. One has

$$Zg(X,Y) = 0 = \nabla_Z(g(X,Y)) = \nabla_Z g(X,Y) + g(\nabla_Z X, Y) + g(X, \nabla_Z Y) = \nabla_Z g(X,Y). \quad (16)$$

The (–) connection is not a Levi-Civita connection since it has non zero torsion.

ods for a simple implicit Euler integration scheme. The Newton algorithm is used to determine the stationary point of a vector field $dL_y f(y) \in T_y G$ where $f : G \to \mathfrak{g}$ is a vector valued function on G derived from an implicit Euler iterate. In both algorithms proposed, Owren and Welfert introduce canonical local coordinates by considering the lift $\tilde{f} := f(L_y \exp(v))$ of f, $\tilde{f} : \mathfrak{g} \to \mathfrak{g}$. The Newton iteration is computed using the Jacobian of $\tilde{f}$ as a vector map $\mathfrak{g} \to \mathfrak{g}$

$$d\tilde{f} = -\tilde{f}.$$

Much of the additional detail in the paper is concerned with the determination of $f(y)$ from the implicit iteration formulae and finding formulae to compute its differential.

This iteration may be interpreted in terms of the (–) connection geometry on a Lie group. Let $X \in \mathfrak{X}(G)$ be a vector field and let $x = dL_{y^{-1}} X_y \in \mathfrak{g}$ be the element of the Lie algebra corresponding to the left translate of X_y back to the identity tangent space. The Jacobian of a vector field $dL_y f(y)$ with respect to each of the Cartan-Schouten connections is

$$\nabla_X dL_y f(y) = dL_y d\tilde{f} x + dL_y \omega(x, f(y)), \quad \omega(x, y) \in \left\{0, \frac{1}{2}[x, y], [x, y]\right\}.$$

For example, the Newton algorithm with respect to the (0) connection leads to an iterate

$$d\tilde{f} x + \frac{1}{2}[x, f(y)] = -\tilde{f}, \quad x \in \mathfrak{g}.$$

From the above discussion, it is clear that the algorithm considered in Owren and Welfert (2000) is based on the (–) connection. Observe that in the vicinity of a critical point then $|[x, f(y)]| = \mathbf{O}(|v_k|^2)$ where $y_k = y \exp(v_k)$ is the canonical local coordinates representation of the present estimate in the Newton iteration. Using this property it is straightforward (cf. Lemma 3.4) to show that any Newton iteration computed using a Cartan-Schouten connection will display quadratic convergence to a local critical point.

The above two examples provide a good contrast of the advantages and possible disadvantages of deriving optimization algorithms according to different geometric structures on a Lie group. In the example of symmetric eigenvalue determination it appears that the better approach is to use the (0) connection linked to the Riemannian structure on the compact manifold $SO(n)$. In particular, the symmetry properties of the connection lead to an improvement in the asymptotic rate of convergence. In contrast, in the second example the natural symmetry of the problem is linked to the definition of the objective function as a smooth vector function $f : G \to \mathfrak{g}$ transported to each tangent space via left multiplication. The structure is closely linked to the geometry of parallel transport by left multiplication that is associated with the (–) connection. Thus, in this example it is natural to use the (–) connection as the basic geometric structure for the problem.

6. Conclusions

In this paper, we have provided a geometric framework for the analysis of Newton methods in terms of the invariant geometry of the Cartan-Schouten connections. We have shown that a Newton method defined with respect to any of the three possible connections will retain the quadratic convergence properties of a classical Newton method. It is the authors' opinion that the framework provided should aid in the choice of suitable numerical algorithms for the class of problems considered based on the properties of the optimization problem under consideration and the natural geometry of the Lie group on which it is posed.

7. Acknowledgements

This work was supported by the Australian Research Council and the Special Research Centre for Ultra-Broadband Information Networks.

References

Absil, P.-A., Mahony, R., Sebulchre, R. and VanDooren, P., (2002). A Grassman-Rayleigh quotient iteration for computing invariant subspaces. SIAM Review 44(1), 57–73.

Boothby, W.M. (1986), *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, London.

Botsaris, C.A. (1978), Differential gradient methods. *Journal of Mathematical Analysis and Applications* 63, 177–198.

Botsaris, C.A. (1981a), A class of differential descent methods for constrained optimization. *Journal Mathematical Analysis and Applications* 79, 96–112.

Botsaris, C.A. (1981b), Constrained optimization along geodesics. *Journal Mathematical Analysis and Applications* 79, 295–306.

Brockett, R.W. (1989), Least squares matching problems. *Linear Algebra and its Applications*, 122–124, 761–777.

Brockett, R.W. (1993), Differential geometry and the design of gradient algorithms. *Proceedings of Symposia in Pure Mathematics* 54, 69–92.

Bruyne, F. De, Anderson, B.D.O., Gevers, M. and Linard, N. (1999), Gradient expressions for a closed-loop identification scheme with a tailor-made parametrization. *Automatica*, 35(11), 1867–1871.

Chu, M.T., (1988). On the continuous realization of iterative processes. *SIAM Review* 30(3), 375–387.

Chu, M.T. and Driessel, K.R. (1990), The projected gradient method for least squares matrix approximations with spectral constraints. *SIAM Journal of Numerical Analysis* 27(4), 1050–1060.

Dehaene, J. (1995), *Continuous-time matrix algorithms systolic algorithms and adaptive neural networks*. Ph.D. thesis, Faculteit Toegepaste Wetenschappen, Department Elektrotechniek-ESAT, Katholieke Universiteit Leuven.

Deift, P., Nanda, T. and Tomei, C., (1983). Ordinary differential equations for the symmetric eigenvalue problem. *SIAM Journal of Numerical Analysis* 20(1), 1–22.

Dennis, J.E. and Schnabel, R.B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall Series in Computational Mathematics. Prentice-Hall, Englewood Cliffs, NJ.

do Carmo, M.P. (1992), *Riemannian Geometry. Mathematics: Theory and Applications*. Birkhäuser, Boston, MA.

Edelman, A., Arias, T.A. and Smith, S.T. (1998), The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications* 20(2), 303–353.

Flashka, H. (1974), The Toda lattice, II. Existence of integrals. *Physical Review b* 9(4), 1924–1925.

Fleming, W. (1977), Functions of Several Variables. *Undergraduate Texts in Mathematics*, Springer, NY.

Fletcher, R. (1996), Practical Methods of Optimization, 2nd edn. Wiley, Oxford.

Gabay, D. (1982), Minimizing a differentiable function over a differentiable manifold, *Journal of Optimization Theory and Applications* 37(2), 177–219.

Gevers, M.R. and Li, G. (1993), Parametrizations in control, estimation and filtering problems: accuracy aspects. In: *Communications in Control Engineering*. Springer, NY.

Helgason, S. (1978), Differential Geometry, Lie Groups and Symmetric Spaces. Academic Press, NY.

Helmke, U. and Moore, J.B. (1994), Optimization and dynamical systems. In: *Communications and Control Engineering*. Springer, London.

Liu, W.Q., Yan, W.Y., Sreeram, V. and Teo, K.L. (1996), Gradient flow approach to LQ cost improvement for simultaneous stabilization problem, *Optimal Control Applications and Methods* 17, 367–375.

Madievski, A.G., Anderson, B.D.O. and Gevers, M.R. (1994), Optimum realizations of sampled data controllers for FWL sensitivity minimization, Automatica 31(3), 367–379.

Mahony, R.E. (1994), *Optimization algorithms on homogeneous spaces: with applications in linear systems theory*. Ph.D. thesis, Department of Systems Engineering, Canberra, Australia.

Mahony, R.E. (1996), The constrained Newton method on a Lie-group and the symmetric eivenvalue problem. *Linear Algebra and its Applications* 248, 67–89.

Manton, J. (2001), Optimisation algorithms exploiting unitary constraints, *IEEE Transactions on Signal Processing*. Accepted for publication.

Moore, J.B., Mahony, R.E. and Helmke, U. (1994), Numerical gradient algorithms for eigenvalue and singular value calculations. *SIAM Journal of Matrix Analysis* 15(3), 881–902.

Nomizu (1954), Invariant affine connections on homogeneous spaces. *American Journal of Mathematics* 76, 33–65.

Owren, B. and Welfert, B. (2000), The Newton method on Lie groups. *BIT* 40(1), 121–145.

Perkins, J.E., Helmke, U. and Moore, J.B. (1990), Balanced realizations via gradient flow techniques *Systems and Control Letters* 14, 369–380.

Smith, S. (1994), Optimization techniques on Riemannian manifolds. *Fields Institute Communications* 3, 113–136. Proceedings Fields Inst. Workshop on Hamiltonian and Gradient Flows, Algorithms and Control.

Smith, S.T. (1993), *Geometric optimization methods for adaptive filtering*. Ph.D. thesis, Division of Applied Science.

Tseng, C.H., Teo, K.L., Cantoni, A. and Zang, Z. (1998), Iterative methods for optimal envelope-constrained filter design using gradient flow approach. *Proceedings of Optimization Techniques and Applications* 1, 534–541.

Udriste, C. (1994), *Convex Functions and Optimization Methods on Riemannian Manifolds*. Kluwer Academic Publishers, Dordrecht.

Varadarajan, V.S. (1984), Lie groups, Lie algebras and their representations. In: *Graduate Texts in Mathematics*. Springer, NY.

Warner, F.W. (1983), Foundations of differentiable manifolds and Lie groups. *Graduate Texts in Mathematics*. Springer, NY.

Watkins, D.S. and Elsner, L. (1988), Self similar flows. *Linear Algebra and its Applications* 110, 213–242.

Yan, W.-Y., Helmke, U. and Moore, J.B. (1994), Global analysis of Oja's flow for a neural networks. *IEEE Transactions on Neural Networks*, 5(5).