The Continuous-time Rayleigh Quotient Flow on the Sphere

R. Mahony,
Dep. of Elec. and Comp. Sys. Eng.,
Monash University,
Clayton, Victoria,
3800, AUSTRALIA.
email: `mahony@ieee.org`

P.-A. Absil,
Institut d'Electricité Montefiore,
Université de Liège,
Bât. B28 Systèmes, Grande Traverse 10,
4000 Liège, BELGIUM.
email: `absil@montefiore.ulg.ac.be`

Submitted to Linear Algebra and Applications,
February 2001.

Abstract

A continuous-time differential equation analogous to the Rayleigh Quotient Iteration for a symmetric matrix A is studied. The set of all continuous solutions, termed the Rayleigh quotient flow, is shown to be a scaled version of the Newton flow for Rayleigh quotient cost functional. The scaling factor ensures that the rate of variation of the Rayleigh quotient is constant and positive along solutions. This interpretation leads to a precise phase portrait for Rayleigh quotient flow. In particular, it is shown that complete solutions of the Rayleigh quotient flow visits the eigenvectors of A in ascending order.

Keywords: Rayleigh quotient, inverse iteration, Newton's method, symmetric eigenvalue problem, finite time convergence.

Submitted to Linear Alg. and App., February 2001.

1 Introduction

One of the 'classical' numerical methods used to compute a single eigenvalue, eigenvector pair for a symmetric matrix A is the *Rayleigh Quotient Iteration* (Wielandt, 1944; Golub & F.Van Loan, 1989). The Rayleigh quotient iteration is also important since it is closely related to the shifted QR-algorithm, an important tool in most numerical routines used to compute eigen-decompositions of symmetric matrices (cf. Watkins (Watkins, 1982) for an excellent review of the QR algorithm). Since the early eighties there has been considerable interest in studying continuous-time flows related to algebraic iterations. The result that ignited world-wide interest in such flows was when iterates of the unshifted QR-algorithm were shown to be unit time samples of a particular Lax-pair equation (Flashka, 1974; Symes, 1982; Deift *et al.*, 1983; Nanda, 1985). This work sparked extensive research on using dynamical systems to solve linear algebraic problems (Brockett, 1989; Watkins & Elsner, 1988; Chu, 1988; Chu & Driessel, 1990; Helmke & Moore, 1994). Despite a strong understanding of the QR algorithm and power method in terms of continuous-time flows (Ammar & Martin, 1986; Chu, 1986; Smith, 1993) an analogous interpretation of the shifted QR-algorithm and the Rayleigh quotient iteration has proved elusive. Chu (Chu, 1986) studied flows on the sphere that compute spectral components of a symmetric matrix. As a remark at the end of this paper, Chu proposed a flow based on the projected Rayleigh quotient direction and showed that the rate of increase of the Rayleigh quotient along solutions was constant. More recently, Smith (Smith, 1993) observed that Newton tangent vector field of the Rayleigh quotient cost (computed with respect to the natural geometry on S^{n-1}) was co-linear with the projected Rayleigh quotient direction and indeed equal to the stereographic projection of the Rayleigh quotient update. The connection between the Rayleigh quotient iteration and the Newton method has been studied by a number of authors since the sixties (Osborne & Michaelson, 1964; Stewart, 1973; Peters & Wilkinson, 1979). To the authors knowledge no prior work has provided a detailed analysis of the Rayleigh quotient flow generated by the projection of the Rayleigh quotient direction on the sphere.

In this paper the continuous-time flow generated by projecting the Rayleigh quotient update direction onto the tangent space of the sphere is fully analysed. The major contribution of the paper is to fully understand the topology and singularities of this flow, termed the *Rayleigh quotient flow*. To do this the relationship between the Rayleigh quotient vector field and the Newton vector field is derived. The connection has been discussed in several prior works (Osborne & Michaelson, 1964; Stewart, 1973; Peters & Wilkinson, 1979; Smith, 1993), however, the two vector fields are not equal and in this paper the difference is expressed in terms of a scaling factor that may be interpreted in terms of the constant increase condition for the Rayleigh quotient flow first noted by Chu (Chu, 1986). The Rayleigh quotient flow is shown to be a singular flow with singularities occurring at eigenvectors. A simple method to continue solutions of the flow at singularities is proposed and leads to complete solutions of the flow on S^{n-1}. A full phase portrait of the Rayleigh quotient flow is derived for general symmetric $n \times n$ matrices.

The paper is divided into five sections including the introduction. In Section 2, continuous-time versions of the power method and Rayleigh quotient algorithms are presented. Section 3 proves two lemmas that show the links between the Newton and the Rayleigh quotient flows. In Section 4 these results are used to derive the full phase portrait of the Rayleigh quotient flow for a general $n \times n$ matrix. The simple cases of 2×2 and 3×3 matrices are explicitly considered and figures displaying the relevant phase portraits are given to provide insight into the general case. Some conclusions are given in Section 5.

2 The Power Method Flow and the Rayleigh Quotient Flow

In this section the well known power method algorithm along with its associated continuous-time flow is reviewed. Using this as a template the Rayleigh quotient iteration is considered and an associated continuous-time flow is proposed.

Let $A = A^T$ be a symmetric matrix in $\mathbb{R}^{n \times n}$. The Rayleigh quotient is a function $r_A : \mathbb{R}^n/\{0\} \to \mathbb{R}$,

$$r_A(x) = \frac{x^T A x}{x^T x}. \tag{1}$$

The Rayleigh quotient can be thought of as providing an estimate of the eigenvalue of A most closely corresponding to the vector direction given by $x \in \mathbb{R}^n$. Indeed, if $Ax = \lambda x$ then $r_A(x) = \lambda$. Moreover, if $A = A^T$ and $x \in \mathbb{R}^n$ are fixed then $\lambda = r_A(x)$ minimises the cost functional $|(A - \lambda I)x|$, where $|\cdot|$ denotes the 2-norm in $\mathbb{R}^n$ (Golub & F.Van Loan, 1989, pg. 440). The Rayleigh quotient is invariant with respect to scaling of the vector x, $r_A(sx) = r_A(x)$ for any scalar $s \in \mathbb{R}$. Consequently, any information about the matrix A 'encoded' in the Rayleigh quotient may be obtained by studying the Rayleigh quotient constrained to the sphere $S^{n-1} = \{x \in \mathbb{R}^n \mid x^T x = 1\}$. In this case the Rayleigh quotient may be written $r_A(x) = x^T A x$.

The power method is the most geometrically natural of the classical eigenvector/eigenvalue methods and (almost) always converges to the maximal (in absolute value) eigenvalue, eigenvector pair of A without any numerical modification. The iterative power method is given by

$$x_{k+1} = \frac{A x_k}{|A x_k|}, \quad x_0 \in S^{n-1}. \tag{2}$$

where $|\cdot|$ denotes the Euclidean 2-norm in $\mathbb{R}^n$.

Let $\pi_x : \mathbb{R}^n \to \mathbb{R}^n$ be the projection of $\mathbb{R}^n$ onto the tangent space of S^{n-1} at a point $x \in S^{n-1}$. Explicitly,

$$\pi_x(Z) := Z - \langle x, Z\rangle x.$$

where $\langle x, Z\rangle = x^T Z$. It is easily verified that $x^T \pi_x(Z) = \langle x, \pi_x(Z)\rangle = 0$. The normal space at $x \in S^{n-1}$ is the span of x. Define the normal projection to be $\pi_x^{\perp} : T_x\mathbb{R}^n \to \mathrm{sp}\{x\}$

$$\pi_x^{\perp}(Z) := \langle x, Z\rangle x.$$

The continuous-time flow associated with the power method is

$$\dot{x} = \pi_x(Ax) = Ax - x^T Axx = (A - r_A(x)I_n)x, \tag{3}$$

for $x(0) = x_0 \in S^{n-1}$. This flow may be interpreted as a gradient ascent flow for the Rayleigh quotient cost

$$\dot{x} = \frac{1}{2}\pi_x(2Ax) = \frac{1}{2}\pi_x(\frac{\partial}{\partial x}r_A) = \frac{1}{2}\mathrm{grad}r_A(x), \tag{4}$$

where the gradient is taken with respect to the natural Riemannian metric on S^{n-1} induced by its embedding in $\mathbb{R}^n$ (Helmke, 1992, Theorem 2.2).

The Rayleigh quotient iteration comes from a class of algorithms known as shifted inverse iterations. That is, the update direction is generated by inverting a given matrix (or in practice by solving the associated linear system of equations) which is modified by a shift of origin of the linear matrix operator to strongly emphasise the desired objective of the iterative procedure. Since the linear systems obtained in this manner tend to be numerically ill-conditioned it is necessary to renormalise after each iteration. For a given estimate $x_k \in S^{n-1}$ the Rayleigh quotient iterate is generated by solving a linear system for Z_{k+1} and normalising

$$\begin{aligned} (A - r_A(x_k)I_n)Z_{k+1} &= x_k & (5)\\ x_{k+1} &= \frac{Z_{k+1}}{|Z_{k+1}|}. & (6) \end{aligned}$$

If x_k is close to an eigenvector $\hat{v}$ with eigenvalue $\hat{\lambda}$ then $r_A(x_k)$ is close to $\hat{\lambda}$ and the column space solution to Eq. 5 will be dominated by the component of Z_{k+1} lying in the eigen-direction $\hat{v}$. After normalisation it is clear that x_{k+1} will be a better approximation of $\hat{v}$ than x_k. It follows that $r_A(x_{k+1})$ will be a better estimate of the eigenvalue $\hat{\lambda}$ and the iterates $x_k \to \hat{v}$. The Rayleigh quotient scheme exhibits cubic convergence to an eigenvector of A (Parlett, 1974). The fast convergence rate is a consequence of the shift of origin $(A - r_A(x_k)I_n)$ and is linked directly to the ill-conditioning of the linear system Eq. 5 (Peters & Wilkinson, 1979). The Rayleigh quotient iterate does not guarantee convergence to the maximal eigenvector of A. In most circumstances (although not always) it will converge to the eigenvector of A associated with the closest eigenvalue $\hat{\lambda}$ of A to the initial value $r_A(x_0)$ of the Rayleigh quotient (Batterson & Smillie, 1989).

The continuous-time flow analysed, associated with the Rayleigh quotient iteration, is defined in an analogous manner to the power method flow Eq. 3:

$$\dot{x} = \pi_x(A - r_A(x)I_n)^{-1}x, \quad x(0) = x_0 \in S^{n-1}. \tag{7}$$

The flow is defined on all S^{n-1} except those vectors such that the $r_A(x)$ is an eigenvalue of A. The flow generated by the singular ODE Eq. 7 is termed the *Rayleigh quotient flow.* Although Eq. 7 is analogous to the power method flow as a projected vector field it does not admit an interpretation as a gradient flow.

Rather than work with the inverse matrix operation it is preferable to rewrite Eq. 7 as a differential/algebraic system

$$\dot{x} = \pi_x(Z) = Z - (x^T Z)x, \quad x(0) = x_0 \tag{8}$$

$$x = (A - r_A(x)I_n)Z. \tag{9}$$

The differential/algebraic structure emphasises the connection to the classical form (Eqn's 5 and 6) of the Rayleigh quotient iteration.

3 Connections between the Rayleigh Quotient flow and the Newton flow.

In this section we give an alternative derivation of Eqn's 8-9 as a scaled version of the Newton flow for the Rayleigh quotient cost. The scaling factor ensures that the Rayleigh quotient along solutions is monotonically increasing with constant rate.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a cost function on $\mathbb{R}^n$ and denote the Euclidean gradient as $Df(\xi)$ and the Hessian as $Hf := D^2 f(\xi)$. The Newton flow is defined to be

$$\dot{\xi} = Z$$

$$(Hf)_\xi Z = -Df(\xi).$$

Locally around any non-degenerate critical point the limiting behaviour of the flow is that of homogeneous linear system

$$\dot{\xi} = -\xi + \mathbf{O}(|\xi|^2). \tag{10}$$

That is, the linearization of the Newton flow at a non-degenerate critical point is structurally stable with independent eigenspaces and all eigenvalues equal to unity. This qualitative behaviour is the continuous-time analogue of the discrete-time quadratic convergence displayed by the classical Newton method.

A number of authors have proposed generalisations of the Newton algorithm to a Riemannian manifold (Botsaris, 1981; Gabay, 1982), (Udriste, 1994, pg. 279), the theses of Smith and Mahony (Smith, 1993; Mahony, 1994) and related papers (Smith, 1994; Mahony, 1996), and the recent paper (Edelman *et al.*, 1998). The problem is not as simple as that of the Euclidean case since the Hessian is not truly defined independently of the local coordinates used. The choice of Hessian $\mathrm{Hess} f = \nabla.\mathrm{grad} f$ is the accepted standard (do Carmo, 1992, pg. 141) and corresponds to using the normal coordinates on the Riemannian manifold as local coordinates for the computation of the Hessian (Mahony & Manton, 2001).

The continuous-time Newton flow on an arbitrary Riemannian manifold that we consider is given by

$$\dot{p} = N,$$

$$\nabla_N \mathrm{grad} f(p) = -\mathrm{grad} f(p), \quad N \in T_p M.$$

It is straightforward to show that the linearization of generalised continuous-time Newton flow around non-degenerate critical point has the form Eq. 10.

Let ∇ denote the Levi-Civita connection on the sphere S^{n-1} induced by the natural Riemannian structure inherited from the embedding of S^{n-1} in $\mathbb{R}^n$. Thus, if X, Z are smooth vector fields on S^{n-1} then (Boothby, 1986, pp. 307-317)

$$(\nabla_X Z)_x = \pi_x(DZ[X])$$

where $DZ[X]$ denotes the Euclidean vector derivative on $\mathbb{R}^n$ (computed by multiplying the vector X by the Jacobian matrix derived from the vector mapping Z).

Lemma 3.1 *Let $A = A^T$ be an $n \times n$ symmetric matrix and let r_A be the Rayleigh quotient given by Eq. 1. For all $x \in S^{n-1}$ such that $r_A(x)$ is not an eigenvalue of A, there exists a unique solution $Z \in \mathbb{R}^n$ to*

$$(A - r_A(x)I_n)Z = x. \tag{11}$$

Let $Z = Z_T + Z_N$ be decomposed into tangential and normal components $Z_T := \pi_x(Z) \in T_xS^{n-1}$ and $Z_N := \pi_x^\perp(Z) = \langle Z, x\rangle x$ in the normal space $T_x^\perp S^{n-1}$. Then

$$\nabla_{Z_T} \mathrm{grad} r_A(x) = -\mu \mathrm{grad} r_A(x)$$

where

$$\mu = \langle Z, x\rangle.$$

Proof Substitute the decomposition of $Z = Z_T + Z_N$ into Eq. 11

$$(A - r_A(x)I_n)Z_T + (A - r_A(x)I_n)Z_N = x. \tag{12}$$

Recalling Eq. 4 note that

$$\begin{aligned}(A - r_A(x)I_n)Z_N &= \langle Z, x\rangle (A - r_A(x)I_n)x \\ &= \frac{1}{2}\langle Z, x\rangle \pi_x(2Ax) \\ &= \frac{1}{2}\langle Z, x\rangle \mathrm{grad} r_A(x).\end{aligned}$$

It follows that $(A - r_A(x)I_n)Z_N \in T_xS^{n-1}$. Projecting Eq. 12 onto T_xS^{n-1} yields

$$\pi_x\left((A - r_A(x)I_n)Z_T\right) + \frac{1}{2}\langle Z, x\rangle \mathrm{grad} r_A(x) = 0.$$

Now since $Z_T \in T_xS^{n-1}$ it is possible to consider the covariant derivative in the direction of Z_T of the smooth tangent vector field $\mathrm{grad} r_A(x) = 2(A - r_A(x)I_n)x$

$$\begin{aligned}\nabla_{Z_T} \mathrm{grad} r_A &= \pi_x\left[D\mathrm{grad} r_A[Z_T]\right] \\ &= \pi_x\left[2(A - r_A(x)I_n)Z_T - 2(Dr_A[Z_T])x\right] \\ &= 2\pi_x\left[(A - r_A(x)I_n)Z_T\right],\end{aligned}$$

since the projection operator annihilates the vector x. Substituting and cancelling the factor of a half completes the proof. $\triangle$

Remark 3.2 An important aspect of the proof of Lemma 3.1 is that only the tangential component of Eq. 11 is used. The normal component is used in the sequel to obtain information on the scaling factor μ. $\square$

Lemma 3.1 shows that the flow associated with Eqn's 8-9 is equivalent to solving a differential equation

$$\dot{x} = Z_T, \quad \nabla_{Z_T} \mathrm{grad} r_A(x) = -\mu \mathrm{grad} r_A(x). \tag{13}$$

where for the moment the scaling factor $\mu = \langle Z, x \rangle$ depends on knowing $Z \in \mathbb{R}^n$. Let N be the Newton vector for the Rayleigh quotient cost functional r_A with respect to the natural geometry on S^{n-1} induced by its embedding in $\mathbb{R}^n$

$$\nabla_N \mathrm{grad} r_A(x) = -\mathrm{grad} r_A(x), \quad N \in T_x S^{n-1}. \tag{14}$$

Comparing Eqn's 13 and 14 and using the properties of the covariant derivative, the Rayleigh quotient flow Eqn's 8-9 may be written

$$\dot{x} = Z_T = \mu N. \tag{15}$$

Thus, the solutions of Eqn's 8-9 correspond to a scaled version of the Newton flow. Although the scaling factor $\mu \in \mathbb{R}$ is a simple scalar value it need not be positive and may well change sign depending on the value of $x \in S^{n-1}$. As a consequence, the topology of the flow induced by Eq. 13 may be quite different to that of the true Newton flow.

It is desirable to obtain an intrinsic interpretation of the scaling factor $\mu = \langle Z, x \rangle$ that does not depend on solving the linear system $(A - r_A(x) I_n) Z = x$ explicitly.

Lemma 3.3 *With the hypotheses and notations of Lemma 3.1, the Rayleigh Quotient flow verifies*

$$\dot{r}_A(x(t)) = D r_A[Z_T] = 2 \tag{16}$$

whenever Z is well defined by Eq. 11. Furthermore, the scaling factor $\mu = \langle Z, x \rangle$ may be written

$$\mu = 1/(x^T A N), \tag{17}$$

where N is the Newton vector, solution of Eq. 14.

Proof Recall that in the proof of Lemma 3.1 only the tangential component of Eq. 12 was used. Considering the normal component of Eq. 12 one obtains

$$\pi_x^\perp \left[(A - r_A(x) I_n) Z_T \right] = \pi_x^\perp(x) = x$$

for $|x| = 1$. Multiplying by x^T and observing that $x^T \pi_x^\perp(v) = x^T v$ for any vector v and $x^T Z_T = 0$ since $Z_T \in T_x S^{n-1}$, one obtains

$$x^T A Z_T = 1. \tag{18}$$

Thus $D r_A[Z_T] = Z_T^T A x + x^T A Z_T = 2$. Replacing $Z_T = \mu N$ into Eq. 18 yields $\mu = 1/(x^T A N)$. $\triangle$

4 Phase Portrait of Rayleigh Quotient and Newton Flows

The phase portrait of the Rayleigh quotient flow and the Newton flow may now be deduced from the information available. In Subsection 4.1, a discussion of the special cases $n = 2$, $n = 3$ is given and provides a qualitative analysis of the Rayleigh quotient flow. In Subsection 4.2, a precise derivation of the full phase portrait is given for the general case.

4.1 Cases $n = 2$ and $n = 3$

In this subsection a qualitative discussion of the phase portrait of the Rayleigh quotient flow is illustrated by the special cases $n = 2$ and $n = 3$.

Recall firstly that the linearization of the Newton flow is a homogeneous exponentially stable linear system around each critical point (cf. Eq. 10). In contrast, Eq. 16 along with a simple Lyapunov argument shows that only the eigenvector corresponding to the maximal eigenvalue of the matrix A remains as a stable equilibrium of the Rayleigh quotient flow. Consider a linear approximation of the Rayleigh quotient flow around an eigenvector v_i of A with eigenvalue λ_i. Recalling the form of Eqn's 10 and 15 it follows that the Rayleigh quotient flow must have a local representation of the form

$$\dot{\xi} = -\mu\xi + \mathbf{O}(|\xi|^2). \tag{19}$$

To compute the scaling factor μ it is sufficient to apply the condition $\dot{r}_A = 2$. In a neighbourhood an eigenvector v_i of A the Rayleigh quotient may be approximated by

$$\begin{aligned} \tilde{r}_A(\xi) &= \frac{v_i^T A v_i + \xi^T A \xi}{(1 + |\xi|^2)} \\ &= \lambda_i + \xi^T (A - \lambda_i I) A \xi + \mathbf{O}(\xi^4), \end{aligned}$$

since $\xi^T A v_i = \lambda_i \xi^T v_i = 0$ for $\xi \in T_{v_i} S^{n-1}$. Combining this with Eq. 19 it is easily verified that

$$\mu = \frac{-1}{\xi^T (A - \lambda_i I)\xi}.$$

It is immediate that the Rayleigh quotient vector field becomes unbounded close to the critical point v_i. This is a necessary consequence of the constant rate of change of the Rayleigh quotient. In a neighbourhood of the maximal eigenvector v_i, $\mu > 0$ and Eq. 19 is locally asymptotically stable. At the minimal eigenvector the scaling factor is negative and Eq. 19 is unstable. At eigenvectors corresponding to intermediate eigenvalues the situation is more complicated and is discussed in the sequel.

Consider the case of a 2×2 matrix A with eigenvalues $\lambda_1 < \lambda_2$. In this case there is no intermediate eigenvalue between λ_1 and λ_2. The Hessian of the Rayleigh quotient is a scalar since the tangent space of S^1 is one-dimensional. At the maximal eigenvector the Hessian is negative while at the minimal eigenvector it is positive. Thus, at some point between the two limiting eigenvectors the Hessian must be degenerate and, since the gradient of the Rayleigh quotient is zero only at eigenvectors, the Newton vector field is infinite. This point acts as a boundary for the Newton flow between the separate domains of attraction for the maximal and minimal eigenvectors. In contrast the Rayleigh quotient vector field is well defined at the singular point since the Rayleigh quotient must lie in the interval between the maximal and minimal eigenvalue (and there are no intermediate eigenvalues) (cf. Eq. 11). Thus, the scaling factor must compensate for the tendency of the Newton vector field to become unbounded in a neighbourhood of the singular point. It follows that the scaling factor μ must converge to zero at the singular point, and that at this point the interpretation given by Eqn's 14 and 15 is no longer valid. Equation 16 shows that the scaling factor must be negative in the domain of attraction to v_1 for the Newton flow. Similarly, the scaling factor will be positive in the domain of attraction of v_2 for the Newton flow. The above analysis leads to a phase portrait shown in Figure 1. It can be seen that even though the mathematical forms of the Newton flow and the Rayleigh quotient flow are closely related the phase portraits of the two flows are quite different.

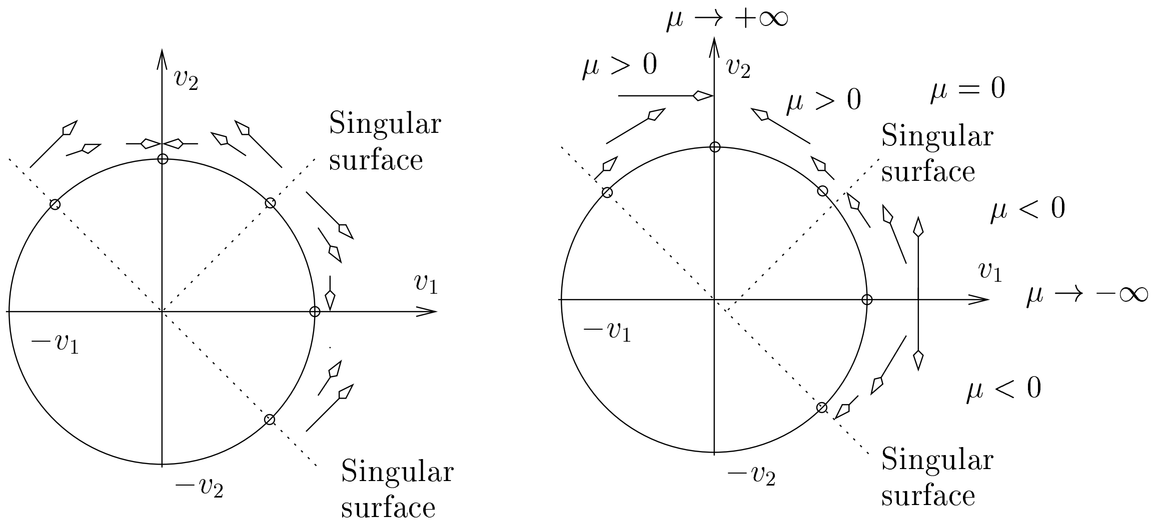

Figure 1: Phase portrait of Newton flow (left) and Rayleigh quotient flow (right) for $A \in \mathbb{R}^{2\times 2}$. The phase portrait is only shown on a hemisphere. Due to the symmetry of the Rayleigh quotient the phase portrait on the other hemisphere may be obtained by reflexion.

The situation where there are intermediate eigenvalues between the maximal and minimal eigenvalues is somewhat more complex. Consider the case of a 3×3 matrix $A = A^T$ with distinct eigenvalues $\lambda_1 < \lambda_2 < \lambda_3$. The Hessian of r_A has two negative eigenvalues at v_3, a positive and negative eigenvalue at v_2 and two positive eigenvalues at v_1. Clearly the set on which the Hessian is singular must intersect any continuous curve drawn between any two of the eigenvectors. The union of this set with the set of points at which the Newton vector field is zero will form the boundary between the domains of attraction of the Newton flow. The stationary points of the Newton flow are exactly the isolated critical points of r_A. Indeed, since the Hessian is bounded, one cannot have both $N = 0$ and $df \neq 0$.

A further set of importance is the level set of r_A corresponding to the intermediate eigenvalue λ_2. On this set the matrix $(A - r_A(x)I_3)$ is singular and Eq. 11 may not have well defined solutions. Indeed, it is clear from inspection that the only case where Eq. 11 may admit a solution is when x is in the column space of $(A - \lambda_2 I_3)$. In the non-degenerate 3×3 case this occurs when x is a linear combination of $\{v_1, v_3\}$. At all other points where $r_A(x) = \lambda_2$ the scaling factor μ is 'infinite'. Recalling again the interpretation Eq. 16 it follows that the scaling factor will tend to infinity in exactly those situations where the Newton flow is tangential to a level set of r_A. It follows that the Newton flow is tangential to the level set $r_A^{-1}(\lambda_2)$. This is a fact which is quite difficult to prove by analysing only the Newton vector field. The local form of the Newton vector field and the Rayleigh quotient vector field in a neighbourhood of v_2 admits an analysis analogous to that given above in Eq. 19 except that the scaling factor converges to infinity along the level set $r_A^{-1}(\lambda_2)$. Near to the level set itself the Rayleigh quotient flow is strongly attractive to v_2 for $r_A(x) < \lambda_2$ and strongly divergent from v_2 if $r_A(x) > \lambda_2$. This introduces a non-uniqueness phenomenon of the Rayleigh quotient flow which is quite common in flows displaying finite time convergence properties. For any point $x \in S^{n-1}$ close to v_2 such that $r_A(x) < \lambda_2$ then $x \to v_2$ in finite time (see Lemma 4.1) and may be extended to a solution beyond this time in an infinite number of ways corresponding to the reverse-time convergence of any point x' with $r_A(x') > \lambda_2$ to v_2. In particular, the solutions of Eqn's 8-9 are not globally unique. To complete the phase portrait the only remaining points of interest are those where the singular set intersects the intermediate level set. In the case of a 3×3 matrix there is only one such point, the intersection of the intermediary level set with $\mathrm{sp}\{v_1, v_3\}$. The phase portrait of the 3×3 case is shown in Figure 2 for the Newton flow and Figure 3 for the Rayleigh quotient flow.

4.2 General case

In this subsection the full phase portrait of the Rayleigh quotient flow is derived.

Lemma 4.1 (distinct eigenvalues) *Let A be an $n \times n$ symmetric matrix with distinct eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ and associated eigenvectors $v_1, \ldots, v_n$. Let $x_0 \in S^{n-1}$ be such that the Rayleigh quotient $r_A(x_0)$ is not an eigenvalue of A. Denote by $\mathcal{S}$ the smallest invariant subspace of A containing x_0, by λ_k the smallest eigenvalue of $A|_{\mathcal{S}}$ greater than $r_A(x_0)$, and by v_k the corresponding eigenvector of $A|_{\mathcal{S}}$. Let $t_1 = (\lambda_k - r_A(x_0))/2$. Then, over the time interval $[0, t_1)$, the Rayleigh quotient flow Eqn's 8-9 admits one and only one solution $x(t)$. This solution resides in $\mathcal{S}$ and $\lim_{t \to t_1} = v_k$, i.e. the flow converges in finite time to the next larger eigenvector of A inside $\mathcal{S}$.*

Proof To simplify the development, express A in the coordinate system achieving $v_1 = e_1, \ldots, v_n = e_n$, where $\{e_1, \ldots, e_n\}$ is the canonical basis of $\mathbb{R}^n$. Thus $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Moreover, let us consider the flow

$$\dot{y} = Z, \quad (A - r_A(y)I)Z = y, \quad y(0) = x_0 \tag{20}$$

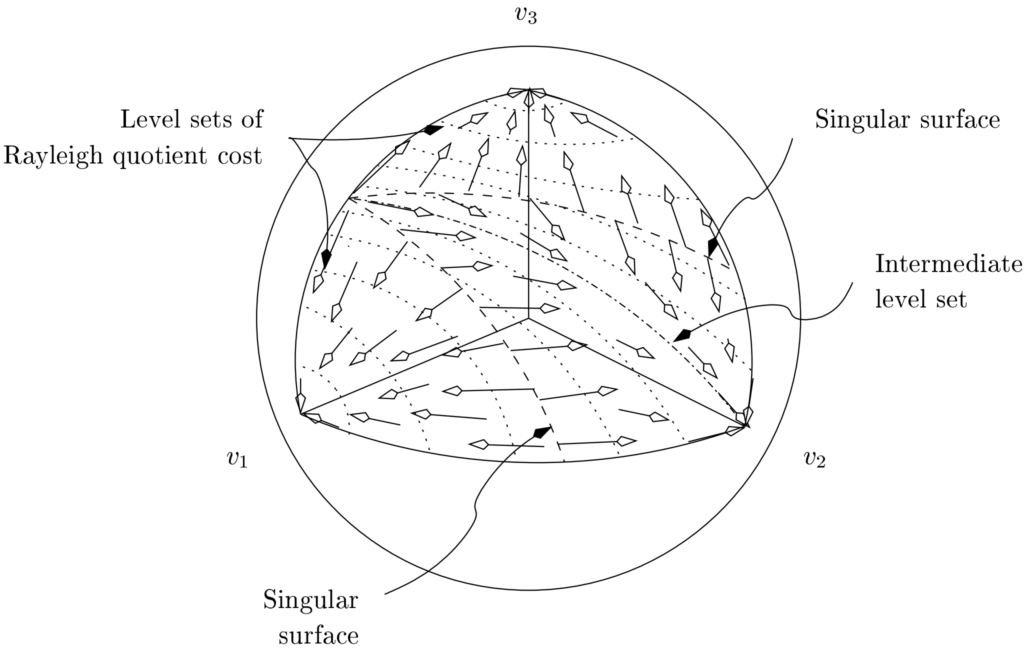

Figure 2: Phase portrait of Newton flow for $A \in \mathbb{R}^{3\times 3}$. The phase portrait on the remaining seven spherical triangles is obtained from symmetry of the Rayleigh quotient.

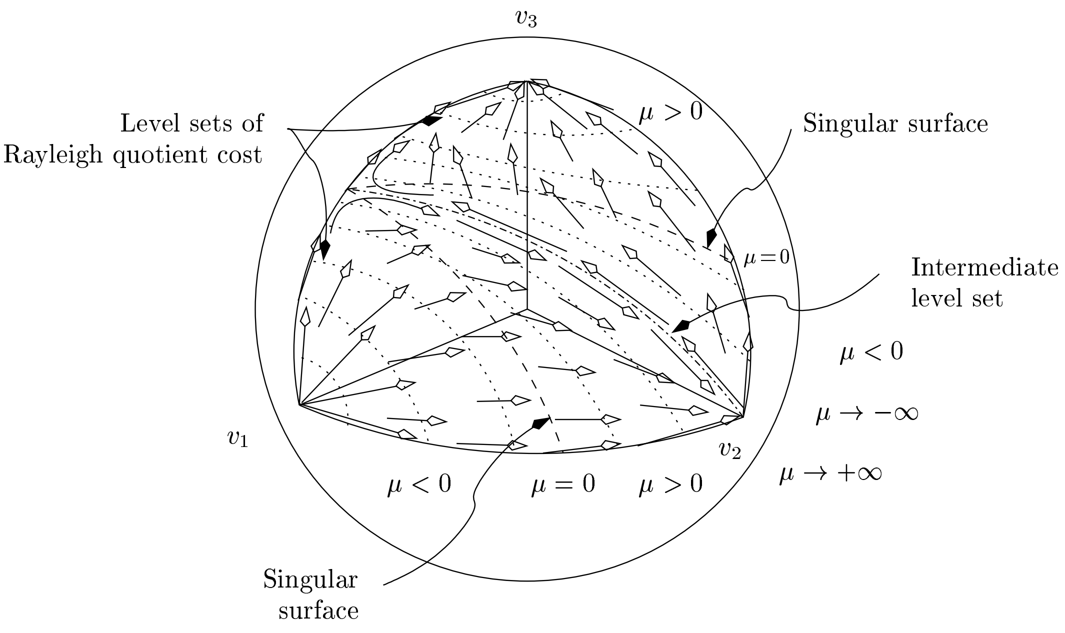

Figure 3: Phase portrait of the Rayleigh quotient flow for $A \in \mathbb{R}^{3\times 3}$. The phase portrait on the remaining seven spherical triangles is obtained from symmetry of the Rayleigh quotient.

where y belongs to $\mathbb{R}^n/0$ and $r_A(y) = (y^TAy)/(y^Ty)$. This flow has the property that the projection of $y(t)$ on S^{n-1}, $x = y/|y|$, verifies the Rayleigh Quotient flow Eqn's 8-9. Indeed, it is easily checked that $\dot{x} = \pi_x((A - r_A(x))^{-1}x)$.

Let $\mathcal{I}$ be the subset of $\{1, \dots, n\}$ such that $y_{i\in\mathcal{I}}(0) = 0$ and $y_{i\notin\mathcal{I}}(0) \neq 0$. Then the smallest invariant subspace of A containing $y(0)$ is $\mathcal{S} = \{y \in \mathbb{R} : y_{i\in\mathcal{I}} = 0\}$.

Let m be such that $\lambda_{m-1} < r_A(y(0)) < \lambda_m$. Two cases have to be considered.

First case: $m \notin \mathcal{I}$, in which case the eigenvalue λ_m of A is also an eigenvalue of $A|_\mathcal{S}$ and m is equal to the k defined in the statement of the Lemma. The equations for the flow of y are

$$\dot{y}_i = (\lambda_i - r_A(y))^{-1}y_i, \; y_i(0) = x_{0_i}, \; i = 1, \dots, n \tag{21}$$

where y_i denotes the ith component of the vector y. Recall that $\dot{r}_A(y(t)) = 2$ (Lemma 3.3). Let t_1 be the time needed for $r_A(y(t))$ to reach the value λ_k, i.e. $t_1 = (\lambda_k - r_A(x_0))/2$. On the time interval $(0, t_1)$, the solution of Eq. 21 is well defined since $(\lambda_i - r_A(y)) \neq 0$ for each $i = 1, \dots, n$. Moreover, on $(0, t_1)$, $(\lambda_{i\neq k} - \rho_A(y))^{-1}$ remain bounded, so $y_{i\neq k}(t)$ are bounded. Also, $y_{i\in\mathcal{I}}(t) = 0$, which means that the solution resides in $\mathcal{S}$. As for y_k, it verifies

$$\dot{y}_k = \frac{1}{2(t_1 - t)}y_k$$

whose solution is $y_k(t) = y_k(0)t_1/(t_1 - t)$. Note that $y_k(0) \neq 0$ by the assumption that $m \notin \mathcal{I}$. Thus $y_k(t)$ converges to infinity as t tends to t_1, while the other components remain bounded. Consequently, $y/|y|$ converges to $\pm e_k$, eigenvector of A relative to λ_k. That is, the solution x of the Rayleigh Quotient flow Eqn's 8-9 converges to v_k.

Second case: $m \in \mathcal{I}$, in which case λ_m is not an eigenvalue of $A|_\mathcal{S}$ and $m < k$. Let T be the time needed for $r_A(y(t))$ to reach λ_m, namely $T = (\lambda_m - r_A(y(0)))/2$. For all $t \in (0, T + \epsilon)$, from Eq. 21, $y_{i\in\mathcal{I},\neq m}(t) = 0$ and well defined, $y_{i\notin\mathcal{I}}(t)$ are well defined and bounded. As for $y_m(t)$, it verifies

$$\dot{y}_m(t) = \frac{1}{2(t_1 - t)}y_m(t), \quad y_m(0) = 0. \tag{22}$$

The only solution to this equation on $(0, T + \epsilon)$ is $y_m(t) = 0$. Indeed, for $t \in (0, T)$, a straightforward integration shows that $y_m(t) = 0$. Now suppose $y_m(\hat{t} > T) \neq 0$, then integrating Eq. 22 backward in time gives $\lim_{t\to T_+} y_m(t) = \infty$, and the solution $y_m(t)$ would not be continuous at $t = T$.
So, in this "second case", the Rayleigh Quotient flow is well defined and remains in $\mathcal{S}$ until at least a time $T+\epsilon$ where $\lambda_m < r_A(y(T+\epsilon)) < \lambda_{m+1}$. Now, define $m := m+1$ and consider again the first case or the second case depending on whether m belongs or not to $\mathcal{I}$.

Eventually, m will equal k. This concludes the proof the Lemma. △

The next lemma concerns the case where the matrix A has repeated eigenvalues.

Lemma 4.2 (repeated eigenvalues) *Suppose that the eigenvalues of A are $\lambda_1, \dots, \lambda_q$ with multiplicities $\eta_1, \dots, \eta_q$ and eigenspaces $\mathcal{V}_1, \dots, \mathcal{V}_q$. Decompose x_0 as $x_0 = x_0^{(1)} + \dots + x_0^{(q)}$ where $x_0^{(i)} \in \mathcal{V}_i$, $i = 1, \dots, q$. Then the Rayleigh quotient flow Eqn's 8-9 reduces to the invariant subspace $\mathcal{S} = \mathrm{sp}(x_0^{(1)}, \dots, x_0^{(q)})$ of A where the eigenvalues $\lambda_1, \dots, \lambda_q$ of A are single, and Lemma 4.1 applies in $\mathcal{S}$.*

Proof We have to prove that $\mathcal{S}$ is invariant under the Rayleigh Quotient flow. Choose an orthonormal coordinate system such that e_j is collinear with $x_0^{(j)}$ for $j = 1, \dots, q$. In this system, A is diagonal and $y_i(0) = 0$, $i = q+1, \dots, n$. Now $y_i(t)$ verifies Eq. 21, and a similar argument to that above shows that $y_i(t) = 0$. △

Lemma 4.1 leads to the conclusion that the solutions of the Rayleigh quotient flow Eqn's 8-9 converge in finite time t_1 to an eigenvector v_k of A where the flow becomes singular. The following lemma shows how the solution of Eqn's 8-9 can be extended beyond t_1.

Lemma 4.3 (continued solutions) *Let A be a symmetric $n \times n$ matrix. Let v_k be an eigenvector of A such that its associated eigenvalue λ_k is not the largest eigenvalue of A. Then there exists $\epsilon > 0$ such that*

i) there exists $Z_k \in T_{v_k} S^{n-1}$ of unitary norm such that $r_A(v_k + \tau Z_k)$ is a monotonically increasing function of $\tau \in (0, \epsilon)$,

ii) there exists a unique solution $x(t)$ of Eqn's 8-9 on $(0, \epsilon)$ such that $\lim_{t \to 0_+} x(t) = v_k$ and $\lim_{t \to 0_+} (\dot{x}(t)/|\dot{x}(t)|) = Z_k$.

Proof Point i) is obvious: a possible choice of Z_k is $Z_k = v_n$ where v_n is an eigenvector associated to the maximal eigenvalue of A. Point ii) is verified by looking at the linear approximation Eq. 19 and matching vector directions at v_k to solutions of Eqn's 8-9 converging to v_k in reverse-time from a neighbourhood of v_k such that $r_A(x) > \lambda_k$. The matching is one to one from the solutions to the directions such that $r_A(v_k + \tau Z_k)$ is a monotonically increasing function of τ. △

Using Lemma 4.3, it is a simple matter to continue a solution of Eqn's 8-9 beyond a critical point by choosing a suitable direction Z_k.

Definition 4.4 (complete solution) *Let A be a symmetric $n \times n$ matrix. Let $x(t)$, $t \in [0, t_1)$, be a solution of Eqn's 8-9 with $\lim_{t \to t_1} x(t) = v_k$ an eigenvector of A. Then t_1 is termed* a critical time. *If $Z_k \in T_{v_k} S^{n-1}$ is suitably chosen then the unique local extension of $x(t)$ on $[0, t_1 + \epsilon)$ defined in Lemma 4.3 is termed* a continued solution. *The direction Z_k is termed an* escape direction. *A* complete solution *is one that is locally continued for all critical times t_{k_i} such that $r_A(x(t_{k_i}))$ is not the largest eigenvalue of A.*

The following theorem describes the phase space of all complete solutions of Eqn's 8-9.

Theorem 4.5 (complete phase portrait) *Let A be an $n \times n$ symmetric matrix. Let $x_0 \in S^{n-1}$ be such that the Rayleigh quotient $r_A(x_0)$ is not an eigenvalue of A. Let $t_* = (\lambda_n - r_A(x_0))/2$ where λ_n is the largest eigenvalue of A.*
Then:

i) There exists a continuous function $x(t)$ on $[0, t_*]$ into S^{n-1} such that $x(t)$ is a complete solution (Definition 4.4) of the Rayleigh Quotient flow Eqn's 8-9.
ii) $\dot{r}_A(x(t))$ is equal to the positive constant 2 for $t \in [0, t_*]$.
iii) $x(t_*) = \lim_{t \to t_*} x(t)$ is the eigenvector of A corresponding to the largest eigenvalue of A.

Let $t_1, \ldots, t_p$ be the critical times of $x(t)$ and $Z_1, \ldots, Z_p$ be the associated escape directions.
Then:

iv) The solution $x(t)$ is uniquely parameterised by $\{x_0, Z_1, \ldots .Z_p\}$.

Let $\mathcal{S}_0$ be the smallest invariant subspace of A containing x_0. Then:

v) $x(t) \in \mathcal{S}_0$ for all $t \in [0, t_1]$ and $\lim_{t \to t_1} x(t) = v_{k_1}$ is the eigenvector corresponding to the least eigenvalue of $A|_{\mathcal{S}_0}$ strictly greater than $r_A(x_0)$.

Let $\mathcal{S}_i$, $i = 1, \ldots, (p-1)$, be the smallest invariant subspace of A containing $x(t_i)$ and Z_i.
Then:

vi) $x(t) \in \mathcal{S}_i$ for all $t \in [t_i, t_{i+1}]$ and $\lim_{t \to t_{i+1}} x(t) = v_{k_{i+1}}$ is the eigenvector corresponding to the least eigenvalue of $A|_{\mathcal{S}_i}$ strictly greater than $r_A(x(t_i))$.

Proof For point i), the existence property stated in Lemma 4.1 and the extension method of Lemma 4.3 ensure that a complete solution exists up to time t_*. Lemma 4.2 ensures that $x(t_*) = \lim_{t \to t_*} x(t)$ is well defined. Point ii) was proved in Lemma 3.3. Point iii) follows from Lemma 4.1 and the fact that $r_A(t_*)$ is equal to the largest eigenvalue of A. The uniqueness of the parameterization in point iv) is a consequence of Lemma 4.3, while v) and vi) follow from Lemmas 4.1 and 4.2. △

Remark 4.6 Under the generic condition that x_0 and the Z_k's are such that the subspaces $\mathcal{S}_0, \ldots, \mathcal{S}_{p-1}$ defined in Theorem 4.5 are of maximal dimension, the complete solution of the Rayleigh quotient flow visits in ascending order all the eigenspaces of A corresponding to the eigenvalues greater than $r_A(x_0)$. □

5 Conclusions

This paper has derived the full phase portrait of the Rayleigh Quotient flow. While the connection between the Rayleigh iteration and the Newton method has been known for a long time, studying the continuous-time flows shows the importance of the scaling factor in modifying the topological structure of the flow. A motivation for considering the Rayleigh quotient flow lies in analysing the convergence and robustness properties of inverse shift algorithms in a rigourous and simple manner. An important application would be to understand the global behaviour of the shifted QR-algorithm, a problem which classical linear algebra analysis techniques have not as yet resolved (Watkins, 1982). This work is part of an ongoing investigation into understanding the geometry of shifted inverse iterations. In further work, the generalised case of a Rayleigh quotient flow evolving on the Grassmann manifold will be considered.

References

Ammar, G., & Martin, C. 1986. The geometry of matrix eigenvalue methods. *Acta applicandae mathematicae*, **5**, 239–278.

Batterson, S., & Smillie, J. 1989. The dynamics of the Rayleigh quotient iteration. *SIAM journal of numerical analysis*, **26**(3), 624–636.

Boothby, W. M. 1986. *An introduction to differentiable manifolds and Riemannian geometry.* London, Britain: Academic Press Inc.

Botsaris, C. A. 1981. 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.

Chu, M. T. 1986. Curves on S^{n-1} that lead to eigenvalues or their means of a matrix. *SIAM journal of algebra and discrete methods*, **7**(3), 425–432.

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

Chu, M. T., & 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.

Deift, P., Nanda, T., & Tomei, C. 1983. Ordinary differential equations for the symmetric eigenvalue problem. *SIAM journal of numerical analysis*, **20**(1), 1–22.

do Carmo, M. P. 1992. *Riemannian geometry.* Mathematics: Theory and Applications. Boston, USA: Birkhäuser.

Edelman, A., Arias, T. A., & 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.

Gabay, D. 1982. Minimizing a differentiable function over a differentiable manifold. *Journal of optimization theory and applications*, **37**(2), 177–219.

Golub, G. H., & F.Van Loan, C. 1989. *Matrix computations.* Baltimore, Maryland U.S.A.: The Johns Hopkins University Press.

Helmke, U. 1992. Isospectral flows and linear programming. *Journal of Australian mathematical society*, **34**(3).

Helmke, U., & Moore, J. B. 1994. *Optimization and dynamical systems.* Communications and Control Engineering. London, UK: Springer-Verlag.

Mahony, R., & Manton, J. 2001. *Geometry of the Newton method on a non-compact Lie-group.* Submitted to J. on Global Optimization, January 2001.

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 eigenvalue problem. *Linear algebra and its applications*, **248**, 67–89.

Nanda, T. 1985. Differential equations and the QR algorithm. *SIAM journal numerical analaysis*, **22**(2), 310–321.

Osborne, M., & Michaelson, S. 1964. The numerical solution of eigenvalue problems in which the eignenvalue parameter appears nonlinearly, with applications to differential equations. *Computer journal*, **7**, 66–71.

Parlett, B. N. 1974. The Rayleigh quotient iteration and some generalisations for nonnormal matrices. *Mathematics of computation*, **28**(127), 679–693.

Peters, G., & Wilkinson, J.H. 1979. Inverse iteration, ill-conditioned equations and Newton's method. *SIAM review*, **21**(3), 339–360.

Smith, S. 1994. Optimization techniques on Riemannian manifolds. *Fields institute commun-ciations*, **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.

Stewart. 1973. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM review*, **15**(4), 727–764.

Symes, W. W. 1982. The QR algorithm and scattering for the finite nonperiodic Toda lattice. *Physica 4d*, 275–280.

Udriste, C. 1994. *Convex functions and optimization methods on Riemannian manifolds.* Kluwer Academic Publishers.

Watkins, D. S. 1982. Understanding the QR algorithm. *SIAM review*, **24**(4), 427–440.

Watkins, D. S., & Elsner, L. 1988. Self similar flows. *Linear algebra and its applications*, **110**, 213–242.

Wielandt, H. 1944. Das Iterationsverfahren bei nicht selbstadjungierten linearen Eigenwertaufgaben. *Math. z.*, **50**, 93–143.