



ELSEVIER

Available online at www.sciencedirect.comLINEAR ALGEBRA
AND ITS
APPLICATIONS

Linear Algebra and its Applications 368 (2003) 343–357

www.elsevier.com/locate/laa

The continuous-time Rayleigh quotient flow on the sphere[☆]

R. Mahony^{a,*}, P.-A. Absil^{b,1}^aDepartment of Engineering, Australian National University, Clayton, Vic. 3800, Australia^bDepartment of Electrical Engineering and Computer Science, Université de Liège,
Bât. B28 Systèmes, Grande Traverse 10, B-4000 Liège, Belgium

Received 22 February 2001; accepted 22 November 2002

Submitted by V. Mehrmann

Abstract

A continuous-time differential equation inspired by 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 time-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 the Rayleigh quotient flow. It is shown that a complete (non-degenerate) solution of the Rayleigh quotient flow visits each of the eigenvectors of A in ascending order.

© 2003 Elsevier Science Inc. All rights reserved.

Keywords: Rayleigh quotient; Inverse iteration; Newton's method; Symmetric eigenvalue problem

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*. The Rayleigh quotient iteration emerged from a more general shifted inverse iteration [37] (cf. [18] for a historical overview). In the early fifties a number of authors [9,33] proposed using the Rayleigh quotient (first introduced by Lord Rayleigh in

[☆] This paper presents research partially supported by the Belgian Program on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture.

* Corresponding author.

E-mail addresses: mahony@ieee.org (R. Mahony), absil@montefiore.ulg.ac.be (P.-A. Absil).

¹ Research Fellow of the Belgian National Fund for Scientific Research (FNRS).

his study of vibrating systems [27]) as an adaptive shift for the inverse iteration. In the late fifties Ostrowski [24] made a rigorous study of its numerical properties and the iteration quickly became a fundamental part of the working toolbox of numerical mathematicians. The Rayleigh quotient iteration is 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. [35] 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. World-wide interest in such flows was ignited when iterates of the unshifted QR-algorithm were shown to be unit time samples of a particular Lax-pair equation [10,13,22,32]. This work sparked extensive research on using dynamical systems to solve linear algebraic problems [5,7,8,17,36]. Despite a strong understanding of the QR-algorithm and power method in terms of continuous-time flows [2,6,30] an analogous interpretation of the shifted QR-algorithm and the Rayleigh quotient iteration has proved elusive. Chu [6] studied flows on the sphere that compute spectral components of a symmetric matrix. As a remark at the end of his 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 [30] 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 [23,26,28,31]. 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 analysed. The major contribution of the paper is to provide a comprehensive study of 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 exploited. A number of prior works have noted the existence of a strong link between the Rayleigh quotient vector field and the Newton vector [23,26,28,30,31]. The two vector fields differ only by a scaling factor. In this paper we link the scaling factor to the property of constant increase in the Rayleigh quotient along solutions to the Rayleigh quotient flow first noted by Chu [6]. 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. Graphical illustrations of the phase portrait are given for the 2×2 and 3×3 cases.

The paper is divided into five sections including the introduction. In Section 2, a brief review of the power method, its associated flow and the Rayleigh quotient

iteration is given before the singular differential equation whose solution we term the Rayleigh quotient flow is defined. Section 3 proves two lemmas that make the link between the Newton and the Rayleigh quotient flows precise. 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 briefly discussed 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 $r_A : \mathbb{R}^n / \{0\} \rightarrow \mathbb{R}$,

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

The Rayleigh quotient provides 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 [15, p. 440]. The Rayleigh quotient is invariant with respect to scaling, $r_A(sx) = r_A(x)$ 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

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

Definition 2.1. Let $\pi_x : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the projection of \mathbb{R}^n onto the tangent space $T_x S^{n-1}$ 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 $T_x^\perp S^{n-1}$ at $x \in S^{n-1}$ is the span of x . Define the normal projection to be $\pi_x^\perp : T_x \mathbb{R}^n \rightarrow \text{sp}\{x\}$

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

The continuous-time flow associated with the power method is

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

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

$$\dot{x} = \frac{1}{2}\pi_x(2Ax) = \frac{1}{2}\pi_x\left(\frac{\partial}{\partial x}r_A\right) = \frac{1}{2}\text{grad } r_A(x), \quad (4)$$

where the gradient is taken with respect to the natural Riemannian metric on S^{n-1} induced by its embedding in \mathbb{R}^n [16, Theorem 2.2].

The Rayleigh quotient iteration comes from a class of algorithms known as shifted inverse iterations. For a given estimate $x_k \in S^{n-1}$ the Rayleigh quotient iterate is generated by solving a linear system for $Z_{k+1} \in \mathbb{R}^n$,

$$(A - r_A(x_k)I_n)Z_{k+1} = x_k. \quad (5)$$

Since the linear systems obtained in this manner tend to be numerically ill-conditioned it is necessary to renormalise after each iteration,

$$x_{k+1} = \frac{Z_{k+1}}{|Z_{k+1}|}. \quad (6)$$

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 eigendirection \hat{v} . After normalisation x_{k+1} will be a better approximation of \hat{v} than x_k and $r_A(x_{k+1})$ will be a better estimate of $\hat{\lambda}$. The Rayleigh quotient scheme exhibits local cubic convergence to an eigenvector of A [25]. 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) [26]. The boundaries of the attractive domains have quite complicated shapes [3].

The continuous-time flow that is the topic of this paper is generated by the singular vector field obtained by projecting the solution of Eq. (5) onto the tangent space of the sphere S^{n-1}

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

The flow is defined on all S^{n-1} except those vectors such that $r_A(x)$ is an eigenvalue of A . The flow generated by integrating the singular ODE Eq. (7) is termed the *Rayleigh quotient flow*. Although Eq. (7) is derived in an analogous manner to the power method flow (cf. Eq. (3)) 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

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

$$x = (A - r_A(x)I_n)Z. \quad (9)$$

The differential/algebraic structure emphasises the connection to the classical form (Eqs. (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 Eqs. (8) and (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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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

$$\begin{aligned} \dot{\xi} &= Z \\ (Hf)_\xi Z &= -Df(\xi). \end{aligned}$$

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

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

A number of authors have proposed generalisations of the Newton algorithm to a Riemannian manifold [14; 34, p. 279], the theses of Smith [30] and Mahony [20] and related papers [21,29], and the recent paper [12]. The usual definition of the Hessian on a Riemannian manifold is $\text{Hess } f = \nabla \text{grad } f$ [11, p. 141]. The Newton vector field is defined as

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

and corresponds to minimising a local quadratic model of the cost f using normal coordinates on the Riemannian manifold as local coordinates [19]. The continuous-time Newton flow on an arbitrary Riemannian manifold is given by

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

where N is the Newton vector field. It is straightforward to show that the linearisation 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 [4, 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

$$(A - r_A(x)I_n)Z = x. \quad (11)$$

Let $Z = Z_T + Z_N$ be decomposed into tangential and normal components $Z_T := \pi_x(Z) \in T_x S^{n-1}$ and $Z_N := \pi_x^\perp(Z) = \langle Z, x \rangle x \in T_x^\perp S^{n-1}$ (cf. Definition 2.1). Then

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

where

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

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

$$(A - r_A(x)I_n)Z_T + (A - r_A(x)I_n)Z_N = x. \quad (12)$$

Recalling Eq. (4) note that

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

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

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

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

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

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

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 .

Lemma 3.1 shows that the flow associated with Eqs. (8) and (9) is equivalent to solving a differential equation

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

where for the moment the scaling factor $\mu = \langle Z, x \rangle$ depends on knowing $Z \in \mathbb{R}^n$ (cf. Eq. (11)). 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

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

Using the properties of the covariant derivative, the Rayleigh quotient flow Eqs. (8) and (9) may be written

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

Thus, the solutions of Eqs. (8) and (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 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 $(A - r_A(x)I_n)Z = x$. Chu [6] studied Eq. (7) and noted that $\dot{r}_A(x(t))$ is constant along solutions of the flow. The following lemma revisits this result.

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

$$\frac{d}{dt} r_A(x(t)) = Dr_A[Z_T] = 2 \tag{16}$$

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

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

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

Proof. From the normal component of Eq. (12) one obtains

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

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

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

Thus $Dr_A[Z_T] = Z_T^T Ax + x^T AZ_T = 2$. Replacing $Z_T = \mu N$ into Eq. (18) yields $\mu = 1/(x^T AN)$. \square

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 Section 4.1 the special cases $n = 2$,

$n = 3$ are briefly discussed and the phase portrait sketched. Section 4.2 presents a full analysis of the phase portrait of the general case.

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

Consider the case of a 2×2 matrix $A = A^T$ with eigenvalues $\lambda_1 < \lambda_2$. 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 intermediary point the Hessian is zero and, since the gradient of the Rayleigh quotient is zero only at eigenvectors, the Newton vector field is infinite. By continuity, it follows that this point is the boundary of the basins of attraction for the maximal and minimal eigenvectors. In contrast the Rayleigh quotient vector field is well defined as long as $r_A(x)$ is not an eigenvalue of A . It follows that the scaling factor μ must converge to zero at the singular point and that Eqs. (14) and (15) are not valid at this point. From Eq. (16) the scaling factor must be negative in the domain of attraction for the minimal eigenvector and positive in the domain of attraction for the maximal eigenvector. The qualitative phase portrait of this case is shown in Fig. 1.

The qualitative phase portrait of the 3×3 case $A = A^T$ is shown in Figs. 2 and 3 for the case $\lambda_2 = \frac{1}{2}(\lambda_1 + \lambda_3)$ with $\lambda_1 < \lambda_3$. The Hessian is singular (and the Newton vector field unbounded) along two surfaces that separate the eigenvectors of A and form the boundaries of basins of attraction of the Newton flow. The Rayleigh quotient vector field is unbounded along the set $r_A(x) = \lambda_2$, the intermediary eigenvalue of A .

4.2. General case

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

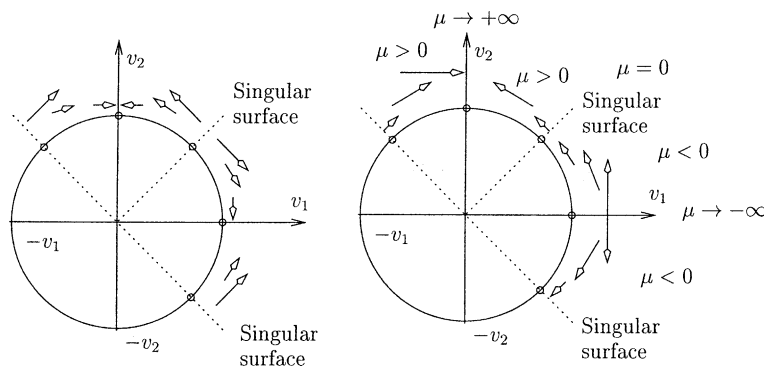


Fig. 1. Phase portrait of Newton flow (left) and Rayleigh quotient flow (right) for $A \in \mathbb{R}^{2 \times 2}$. The phase portrait is only shown on a hemisphere. Due to the symmetry of the Rayleigh quotient the phase portrait on the other hemisphere may be obtained by reflexion.

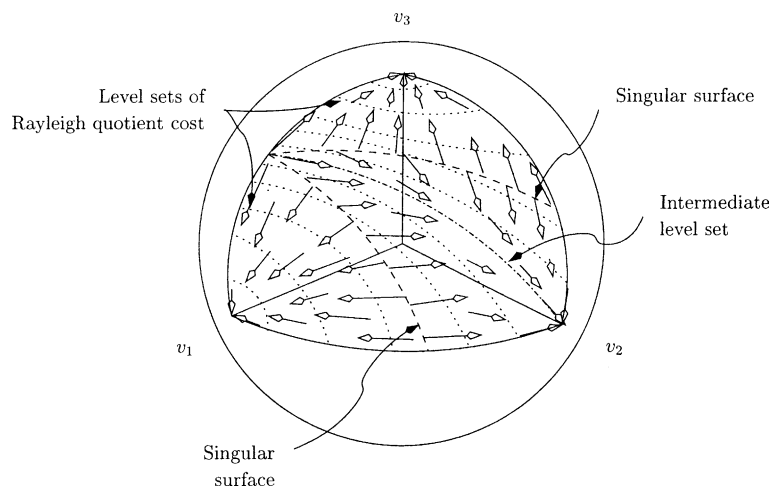


Fig. 2. Phase portrait of Newton flow for $A = A^T \in \mathbb{R}^{3 \times 3}$ and $\lambda_2 = (\lambda_1 + \lambda_3)/2$. The phase portrait on the remaining seven spherical triangles is obtained from symmetry of the Rayleigh quotient.

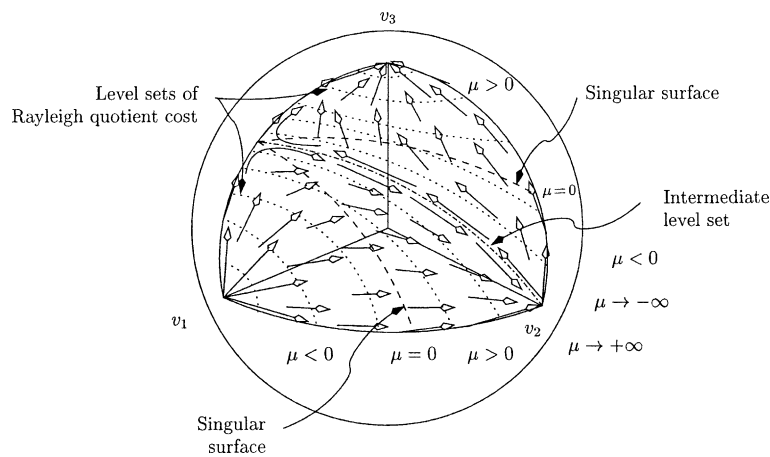


Fig. 3. Phase portrait of the Rayleigh quotient flow for $A = A^T \in \mathbb{R}^{3 \times 3}$ and $\lambda_2 = (\lambda_1 + \lambda_3)/2$. The phase portrait on the remaining seven spherical triangles is obtained from symmetry of the Rayleigh quotient.

Lemma 4.1 (Distinct eigenvalues). *Let A be an $n \times n$ symmetric matrix with distinct eigenvalues $\lambda_1 < \lambda_2 < \dots < \lambda_n$ and associated eigenvectors v_1, \dots, v_n . Let $x_0 \in S^{n-1}$ such that $r_A(x_0)$ 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 Eqs. (8) and (9)*

admits one and only one solution $x(t)$. This solution resides in \mathcal{S} and $\lim_{t \rightarrow t_1} x(t) = v_k$, i.e. the flow converges in finite time to the next larger eigenvector of A inside \mathcal{S} .

Proof. Change coordinates to obtain $A = \text{diag}(\lambda_1, \dots, \lambda_n)$, $v_i = e_i$, $i = 1, \dots, n$ where $\{e_1, \dots, e_n\}$ is the canonical basis of \mathbb{R}^n . Consider the extended flow defined on all $\mathbb{R}^n / \{0\}$

$$\dot{y} = Z, \quad (A - r_A(y)I)Z = y, \quad y(0) = x_0, \tag{19}$$

where $r_A(y) = (y^T A y) / (y^T y)$. Let $y(t)$ be a solution of Eq. (19) and set $x(t) = y(t) / |y(t)|$. It is easily verified that $\dot{x} = \pi_x((A - r_A(x))^{-1}x)$ and hence $x(t)$ is a local solution to the Rayleigh quotient flow Eqs. (8) and (9).

Let \mathcal{J} be the subset of $\{1, \dots, n\}$ such that $y_{i \in \mathcal{J}}(0) = 0$ and $y_{i \notin \mathcal{J}}(0) \neq 0$. The smallest invariant subspace of A containing $y(0)$ is $\mathcal{S} = \{y \in \mathbb{R}^n : y_{i \in \mathcal{J}} = 0\}$. Let m be such that $\lambda_{m-1} < r_A(y(0)) < \lambda_m$. Two cases have to be considered.

First case. If $m \notin \mathcal{J}$ the eigenvalue λ_m of A is an eigenvalue of $A|_{\mathcal{S}}$ and m is equal to k (cf. lemma statement). Eq. (19) may be decomposed into n independent differential equations

$$\dot{y}_i = (\lambda_i - r_A(y))^{-1} y_i, \quad y_i(0) = x_{0_i}, \quad i = 1, \dots, n, \tag{20}$$

where y_i denotes the i th component of the vector y . Recall that $\dot{r}_A(y(t)) = 2$ (Lemma 3.3). Let $t_1 = (\lambda_k - r_A(x_0)) / 2$. For $i \neq k$ then $(\lambda_i - r_A(y)) \neq 0$ is bounded away from zero on $[0, t_1)$ and the solution $y_i(t)$ exists and is unique. By inspection $y_{i \in \mathcal{J}}(t) = 0$ and it follows that the solution resides in \mathcal{S} . The k th element of $y(t)$ satisfies

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

whose solution is $y_k(t) = y_k(0)t_1 / (t_1 - t)$, ($y_k(0) \neq 0$ by assumption). Thus $y_k(t)$ converges to infinity as t tends to t_1 , while the other components remain bounded. Consequently, $y / |y|$ converges to $\text{sgn}(y_k(0))e_k$. That is, the solution x of the Rayleigh quotient flow Eqs. (8) and (9) converges to v_k .

Second case. If $m \in \mathcal{J}$ then λ_m is not an eigenvalue of $A|_{\mathcal{S}}$ and $m < k$. Set $T_1 = (\lambda_m - r_A(y(0))) / 2$ and consider a time-interval $(0, T_1 + \epsilon)$ for $\epsilon > 0$. The m th component of $y(t)$ verifies

$$\dot{y}_m(t) = \frac{1}{2(T_1 - t)} y_m(t), \quad y_m(0) = 0. \tag{21}$$

The unique continuous solution to this equation on $(0, T_1 + \epsilon)$ is $y_m(t) = 0$. Indeed, for $t \in (0, T_1)$, a straightforward integration shows that $y_m(t) = 0$. Now suppose $y_m(\hat{t} > T_1) \neq 0$, then integrating Eq. (21) backward in time gives $\lim_{t \rightarrow (T_1)_+} y_m(t) = \infty$, and the solution $y_m(t)$ would not be continuous at $t = T_1$. For $i \neq k$ then $(\lambda_i - r_A(y)) \neq 0$ is bounded away from zero on $[0, T_1 + \epsilon)$ and the solution $y_i(t)$ exists and is unique. Furthermore, $y_{i \in \mathcal{J}}(t) = 0$ and the solution resides in \mathcal{S} . Thus, the

Rayleigh quotient flow is well defined and remains in \mathcal{S} until at least a time $T_1 + \epsilon$ where $\lambda_m < r_A(y(T_1 + \epsilon)) < \lambda_{m+1}$. Either $m + 1 \notin \mathcal{I}$ and the first case applies, or the second case applies again and the solution may be continued to a second time $T_1 + T_2 + \epsilon$ at which $\lambda_{m+1} < r_A(y(T_1 + T_2 + \epsilon)) < \lambda_{m+2}$. Since the Rayleigh quotient is increasing constantly along the solution it follows that \mathcal{S} must contain an eigenvector v_k of A with $\lambda_k > r_A(y(t))$. Repeating the second case until the first case applies completes the proof. \square

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 η_1, \dots, η_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 Eqs. (8) and (9) reduces to the invariant subspace $\mathcal{S} = \text{span}(x_0^{(1)}, \dots, x_0^{(q)})$ of A where the eigenvalues $\lambda_1, \dots, \lambda_q$ of $A|_{\mathcal{S}}$ have algebraic multiplicity one, and Lemma 4.1 applies in \mathcal{S} .*

Proof. 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. (20) and $y_i(t) = 0, i = q + 1, \dots, n$ (cf. proof of Lemma 4.1). This proves that \mathcal{S} is invariant under the Rayleigh quotient flow and the result follows from Lemma 4.1. \square

The above results lead to the conclusion that the solutions of the Rayleigh quotient flow Eqs. (8) and (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 Eqs. (8) and (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 Eqs. (8) and (9) on $(0, \epsilon)$ such that $\lim_{t \rightarrow 0_+} x(t) = v_k$ and $\lim_{t \rightarrow 0_+} (\dot{x}(t)/|\dot{x}(t)|) = Z_k$.*

Proof. Point (i) is direct, a possible choice is $Z_k = v_n$ where v_n is an eigenvector associated to the maximal eigenvalue of A . To prove point (ii) the local structure of the Rayleigh quotient flow at an eigenvector v_k is considered. Recall that the linearisation of the Newton flow is a homogeneous exponentially stable linear system around each critical point (cf. Eq. (10)). Using the relationships Eqs. (10) and

(15) for $r_A(x) \neq \lambda_k$ it follows that the Rayleigh quotient flow must have a local representation of the form

$$\dot{\xi} = -\mu\xi + \mathbf{O}(|\xi|^2). \quad (22)$$

In a neighbourhood of v_k the Rayleigh quotient may be approximated by

$$r_A(\xi) = \frac{v_k^T A v_k + \xi^T A \xi}{(1 + |\xi|^2)} = \lambda_k + \xi^T (A - \lambda_k I) \xi + \mathbf{O}(\xi^4),$$

since $\xi^T A v_k = \lambda_k \xi^T v_k = 0$ for $\xi \in T_{v_k} S^{n-1}$. Recalling Lemma 3.3 and using Eq. (22) it follows that

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

Choosing Z_k as above and choosing $\xi_0 = \tau Z_k$ for $\tau > 0$ sufficiently small it is clear that $\mu < 0$. Solving the approximation Eq. (22) in reverse time proves point (ii) for the approximate system. By continuity of solutions this result extends to solutions of the full system. \square

Using Lemma 4.3, it is a simple matter to continue a solution of Eqs. (8) and (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 Eqs. (8) and (9) with $\lim_{t \rightarrow t_1} x(t) = v_k$ 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 Eqs. (8) and (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 Eqs. (8) and (9).
- (ii) $\dot{r}_A(x(t))$ is equal to the positive constant 2 for $t \in [0, t_*]$.
- (iii) $x(t_*) = \lim_{t \rightarrow t_*} x(t)$ is the eigenvector of A corresponding to the largest eigenvalue of A .

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

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

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 $\lim_{t \rightarrow t_1} x(t) = v_{k_1}$ is the eigenvector corresponding to the least eigenvalue of $A|_{\mathcal{S}_0}$ strictly greater than $r_A(x_0)$.

Let $\mathcal{S}_i, i = 1, \dots, (p - 1)$, 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 $\lim_{t \rightarrow t_{i+1}} x(t) = v_{k_{i+1}}$ 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 $x(t_*) = \lim_{t \rightarrow t_*} x(t)$ is well defined. Point (ii) was proved in Lemma 3.3 and holds at critical times by continuity of the solution. 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 parameterisation in point (iv) is a consequence of Lemma 4.3, while (v) and (vi) follow from Lemmas 4.1 and 4.2. \square

Remark 4.6. Under the generic condition that x_0 and the Z_k 's are such that the subspaces $\mathcal{S}_0, \dots, \mathcal{S}_{p-1}$ 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 $r_A(x_0)$. We term a complete solution with this property a *non-degenerate* solution.

Remark 4.7. The Rayleigh quotient flow may be solved in reverse time. All results obtained are valid (with appropriate modification of the statements) and a non-degenerate solution will converge to a minimal eigenvector of A , visiting each eigenvector in descending order of magnitude of the associated eigenvalues.

The authors have verified the results summarised in Theorem 4.5 numerically using a simple fixed step Euler integration scheme. Convergence to a neighbourhood of the next highest eigenvector occurs in a finite number of steps. Continuing the iteration leads to instability due to the growth of the vector field close to an eigenvector. Significant numerical ill-conditioning also occurs at the intermediate level sets where $r_A(x) = \lambda_k$. An effective numerical algorithm based on the Rayleigh quotient flow would need to be tailored to deal with these difficulties in an efficient manner. Since the Rayleigh quotient iteration is itself cubically convergent and well understood it is not clear that an algorithm based on the Rayleigh quotient flow would contribute to the toolbox of methods for computing single eigenvector–eigenvalue pairs of symmetric matrices. However, the conceptual framework developed has already helped in the analysis of generalised subspace iterations [1] and the authors believe that the perspective provided by studying continuous-time flows

associated with discrete iterations has the potential to contribute significantly to work in this field.

5. Conclusions

This paper has derived the full phase portrait of the Rayleigh quotient flow. Connections between the Rayleigh quotient iteration and the Newton method have been known for a long time. The significant difference between the topological structures of the Rayleigh quotient flow and the Newton flow highlights the importance of the role of the scaling factor. The presence of the scaling factor in the Rayleigh quotient flow is directly due to its derivation from an inverse shift algorithm. Understanding the structure of inverse shift algorithms provides a key motivation for the results presented in this paper. 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 [35]. The present paper 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

- [1] P.-A. Absil, R. Mahony, R. Sepulchre, P. Van Dooren, A Grassmann–Rayleigh quotient iteration for computing invariant subspaces, *SIAM Rev.* 44 (1) (2002) 57–73.
- [2] G. Ammar, C. Martin, The geometry of matrix eigenvalue methods, *Acta Appl. Math.* 5 (1986) 239–278.
- [3] S. Batterson, J. Smillie, The dynamics of the Rayleigh quotient iteration, *SIAM J. Numer. Anal.* 26 (3) (1989) 624–636.
- [4] W.M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press Inc., London, UK, 1986.
- [5] R.W. Brockett, Least squares matching problems, *Linear Algebra Appl.* 122–124 (1989) 761–777.
- [6] M.T. Chu, Curves on S^{n-1} that lead to eigenvalues or their means of a matrix, *SIAM J. Algebra Discrete Methods* 7 (3) (1986) 425–432.
- [7] M.T. Chu, On the continuous realization of iterative processes, *SIAM Rev.* 30 (3) (1988) 375–387.
- [8] M.T. Chu, K.R. Driessel, The projected gradient method for least squares matrix approximations with spectral constraints, *SIAM J. Numer. Anal.* 27 (4) (1990) 1050–1060.
- [9] S. Crandall, Iterative procedures related to relaxation methods for eigenvalue problems, *Proc. Roy. Soc. London* 207 (1951) 416–423.
- [10] P. Deift, T. Nanda, C. Tomei, Ordinary differential equations for the symmetric eigenvalue problem, *SIAM J. Numer. Anal.* 20 (1) (1983) 1–22.
- [11] M.P. do Carmo, *Riemannian geometry*, in: *Mathematics: Theory and Applications*, Birkhäuser, Boston, MA, 1992.
- [12] A. Edelman, T.A. Arias, S.T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.* 20 (2) (1998) 303–353.
- [13] H. Flashka, The Toda lattice, II. Existence of integrals, *Phys. Rev. B* 9 (4) (1974) 1924–1925.
- [14] D. Gabay, Minimizing a differentiable function over a differentiable manifold, *J. Optim. Theory Appl.* 37 (2) (1982) 177–219.

- [15] G.H. Golub, C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, USA, 1989.
- [16] U. Helmke, Isospectral flows and linear programming, *J. Austral. Math. Soc.* 34 (3) (1992).
- [17] U. Helmke, J.B. Moore, *Optimization and dynamical systems*, in: *Communications and Control Engineering*, Springer-Verlag, London, UK, 1994.
- [18] I.C.F. Ipsen, A history of inverse iteration, in: B. Huppert, H. Schneider (Eds.), Helmut Wielandt, *Mathematische Werke, Mathematical Works*, in: *Matrix Theory and Analysis*, vol. II, Walter de Gruyter, New York, 1995.
- [19] R. Mahony, J. Manton, Geometry of the Newton method on non-compact Lie-groups, *J. Global Optim.* 23 (3) (2002) 309–327.
- [20] R.E. Mahony, *Optimization algorithms on homogeneous spaces: with applications in linear systems theory*, Ph.D. thesis, Department of Systems Engineering, Canberra, Australia, 1994.
- [21] R.E. Mahony, The constrained Newton method on a Lie-group and the symmetric eigenvalue problem, *Linear Algebra Appl.* 248 (1996) 67–89.
- [22] T. Nanda, Differential equations and the QR algorithm, *SIAM J. Numer. Anal.* 22 (2) (1985) 310–321.
- [23] M. Osborne, S. Michaelson, The numerical solution of eigenvalue problems in which the eigenvalue parameter appears nonlinearly, with applications to differential equations, *Comput. J.* 7 (1964) 66–71.
- [24] A.M. Ostrowski, On the convergence of the Rayleigh quotient iteration for the computation of characteristic roots and vectors: Parts I–VI, *Arch. for Ration. Mech. Anal.* 1–4 (1958/1959) 233–241, 423–428, 325–340, 341–347, 472–481, 153–165.
- [25] B.N. Parlett, The Rayleigh quotient iteration and some generalisations for nonnormal matrices, *Math. Comp.* 28 (127) (1974) 679–693.
- [26] G. Peters, J.H. Wilkinson, Inverse iteration, ill-conditioned equations and Newton’s method, *SIAM Rev.* 21 (3) (1979) 339–360.
- [27] L. Rayleigh, *The Theory of Sound*, second ed., MacMillan, New-York, NY, USA, 1932.
- [28] M. Shub, Some remarks on dynamical systems and numerical analysis, in: L. Lara-Carrero, J. Lewowicz (Eds.), *Dynamical Systems and Partial Differential Equations*, Proc. VII ELAM, Equinoccio, Universidad Simón Bolívar, Caracas, Bolíviar, 1986, pp. 69–92.
- [29] S. Smith, Optimization techniques on Riemannian manifolds, *Fields Inst. Commun.* 3 (1994) 113–136 (Proceedings Fields Inst. Workshop on Hamiltonian and Gradient Flows, Algorithms and Control).
- [30] S.T. Smith, *Geometric optimization methods for adaptive filtering*, Ph.D. thesis, Division of Applied Science, 1993.
- [31] Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Rev.* 15 (4) (1973) 727–764.
- [32] W.W. Symes, The QR algorithm and scattering for the finite nonperiodic Toda lattice, *Phys. D* 4 (1982) 275–280.
- [33] G. Temple, The accuracy of Rayleigh’s method of calculating the natural frequencies of vibrating systems, *Proc. Roy. Soc. London* 211 (1952) 204–224.
- [34] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers, 1994.
- [35] D.S. Watkins, Understanding the QR algorithm, *SIAM Rev.* 24 (4) (1982) 427–440.
- [36] D.S. Watkins, L. Elsner, Self similar flows, *Linear Algebra Appl.* 110 (1988) 213–242.
- [37] H. Wielandt, Das Iterationsverfahren bei nicht selbstadjungierten linearen Eigenwertaufgaben, *Math. z.* 50 (1944) 93–143.