Proceedings of the VII ELAM.

UNIVERSIDAD SIMON BOLIVAR
Caracas, 16-27 de julio, 1984

DYNAMICAL SYSTEMS AND PARTIAL DIFFERENTIAL EQUATIONS: Proceedings of the VII ELAM.

Editors Lorenzo Lara-Carrero
Jorge Lewowicz

Equinoccio / VII ELAM /

SOME REMARKS ON DYNAMICAL SYSTEMS AND NUMERICAL ANALYSIS

Michael Shub
Queens College and the
Graduate School of CUNY
Flushing, NY 11367/USA

Dynamical systems and numerical analysis are intimately linked. Numerical methods are used to approximate solutions of equations, and are frequently themselves iterative methods - that is, dynamical systems. An understanding of the local and global geometry and dynamics both of the systems to be approximated and/or of the numerical methods themselves can be useful to crucial for the understanding of the functioning of the numerical method, the validity of the numerical results and their applicability to the original problem.

In this lecture, we comment on the Euler and Taylor approximations to the solutions of a differential equation in Section 1, Newton and gradient methods for finding zeros of a vector field in Sections 2 and 3, and the particular case of the Rayleigh quotient iteration in Section 3.

Much of the material we discuss comes from Hirsch-Smale, Smale, and Shub-Smale.

Section 1:

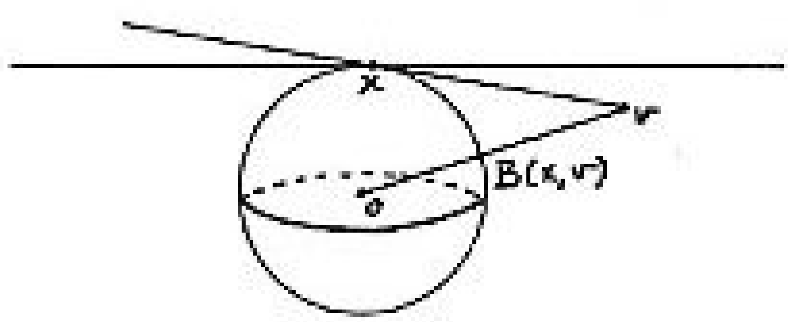
Given a differential equation $\dot{x} = V(x)$ in R^n, Euler's method with step size $h \neq 0$ for the approximation of a solution is given by $x_{i+1} = x_i + hV(x_i)$. Let $\Psi_t(x)$ denote the solution through x at time t. Setting $h = \frac{t}{n}$ and $x_0 = x, x_n$ is a good approximation to $\Psi_t(x)$; x_n converges to $\Psi_t(x)$ as $n \to \infty$ with good estimates on the error for V which are C^1 bounded (See Atkinson [1978]§6.2 for example). This is the traditional numerical analysis approach. It ignores the global nature of Ψ_t and $E_h(x) = x + hV(x)$. In fact, the first result we prove is that E_h and Ψ_h are tangent at $h = 0$ in the C^1 topology for V which are C^2 bounded. It can be useful to have this result on general manifolds, so first we define the Euler approximation for general manifolds. Given a

* Partially supported by NSF Grant #MCS 8201267.

smooth complete Riemannian manifold M which perhaps has boundary and a C^1 vector field $V:M\to TM$, we suppose that V points in along the boundary or is tangent to the boundary. Let $\Psi_t(x)$ denote the solutions of $\dot{x} = V(x)$ with initial condition x. Thus Ψ_t is defined for all $t \geq 0$ and $\Psi_t:M \to M$ is an embedding. Now we wish to describe the displacement of x by $tV(x)$.

For this we suppose a chart $B_x:U_x \to M$ defined on a neighborhood U_x of 0 in T_xM mapping 0 to x. In case x is a boundary point we suppose tangent vectors to the boundary are mapped to the boundary and inward pointing vectors are mapped into M, U_x then is a neighborhood of 0 in the half space. The sum $x + tV(x)$ is now simply replaced by $E_t(x) = B_x[tV(x)]$. This agrees with the usual definition in Euclidean space where $B_x(v) = x + v$ for any tangent vector v at x. In general, we require regularity conditions on B_x, $DB_x(0) = Id_{T_xM}$, $B:U \to M$ defined by $(x,v) \to [B_x(v)]$ is smooth with Lipschitz first and second derivatives which are bounded along M. Moreover, B is supposed to be defined on a uniform neighborhood of the zero section in TM. This last condition can be interpreted as a statement about the natural extension of B to the double of M if M has boundary.

Examples of maps B on manifolds M are provided by the exponential map of a Riemannian metric. For the unit sphere in Euclidean space, $S^{n-1} = \{x \in R^n \mid \|x\| = 1\}$. Let $B(x,v) = \dfrac{x + v}{\|x + v\|}$ where $v \in T_xS^{n-1}$.



This B has the added advantage of being injective on all T_xS^{n-1}.

<u>Theorem 1:</u>

Let V be a C^2 bounded vector field. Then E_t and Ψ_t are tangent at $t = 0$ in the C^1 topology. More precisely, there is a constant C such that in the one norm $||E_t - \Psi_t||_1 \leq Ct^2$

Proof:

It suffices to prove this theorem in Euclidean space for general B either by adapting the proof to charts or embedding the manifold M in Euclidean space and extending all the structures. Thus we assume $V: R^n \to R^n$ and is C^2 bounded, i.e. there exists $K_0, K_1, K_2 \geq 0$ such that $||V(x)|| \leq K_0$, $||DV(x)|| \leq K_1$, and $||D^2V(x)|| \leq K_2$ for all $x \in R^n$. We break the proof into three steps first some elementary estimates, next the case $B(x,v) = x + v$ so $E_t(x) = x + tV$ and finally the case of general B.

Remark:

Given a contraction map f of a complete metric space $d[f(x),f(y)] < \lambda d(x,y)$ and $0 \leq \lambda < 1$, let x_0 be the fixed point of f. Then $d(x,x_0) \leq \frac{d[x,f(x)]}{1 - \lambda}$.

Lemma 1:

For $0 \leq s < \frac{1}{K_1}$

a) $||\Psi_s(x) - x|| \leq \frac{sK_0}{1-sK_1}$

b) $||D_x\Psi_s(x) - Id|| \leq \frac{sK_1}{1-sK_1}$

c) $||D_x\Psi_s(x)|| \leq \frac{1}{1-sK_1}$

d) $||D^2{}_x\Psi_s(x)|| \leq \frac{sK_2}{(1-sK_1)^3}$

e) $||D_t(D_xV[\Psi_s(x)]D_x\Psi_s(x))|| \leq \frac{K_0K_2}{1-sK_1} + \frac{K_0K_1K_2s}{(1-sK_1)^3}$

Proof:

a) $\Psi_s(x) = x + \int_0^s V[\Psi_t(x)]dt$ (I)

The contraction rate to determine Ψ is s Lip $V = sK_1$ and the identity function is moved at most by sK_0.

b) Differentiate (I)

$$D_x\Psi_s(x) = I + \int_0^s D_xV[\Psi_t(x)]D_x\Psi_t(x)dt. \tag{II}$$

The contraction rate is sK_1 and the constant function the identity is moved at most sK_1.

c) Add one to b).

d) Differentiate (II)

$$D_x^2\Psi_s(x) = \int_0^s D_x^2V(\Psi_t(x))D_x\Psi_t(x)D_x\Psi_t(x) + D_xV(\Psi_t(x))D_x^2\Psi_t(x)dt.$$

The contraction rate is sK_1 and the zero function is moved at most $K_2 \cdot \frac{1}{(1-sK_1)^2}$.

e) $D_t(D_xV(\Psi_s(x)) \circ D_x\Psi_s(x)) =$

$D_x^2V(\Psi_x(x)) \circ D_t\Psi_s(x) \circ D_x(\Psi_x(x) + D_xV(\Psi_s(x)) \circ D_x^2(\Psi_s(x)) \circ D_t\Psi_s(x)$.

Multiply the bounds and add.

Now we return to the proof of the theorem and consider the case that $B(x,v) = x + v$ so $E_t(x) = x + tV(x)$, then for $0 < t < \frac{1}{K_1}$

$$||\Psi_t(x) - E_t(x)|| = ||x + \int_0^t V(\Psi_s(x))dx - (x + tV(x))||$$

$$= ||\int_0^t V(\Psi_s(x)) - V(x)dx|| \leq t \sup_{0\leq s\leq t} ||V(\Psi_s(x)) - V(x)||$$

$$\leq K_1 t \sup_{0\leq s\leq t} ||\Psi_s(x) - x|| \leq K_1 \frac{K_0}{1-tK_1} t^2 \tag{III}$$

by Lemma 1, and

$$||D_x\Psi_t(x) - DE_t(x)|| = || I + \int_0^t D_xV(\Psi_s(x))D_x\Psi_s(x)ds - (I + tD_xV(x)||$$

$$= \left\| \int_0^t D_x V(\Psi_s(x)) D_x \Psi_s(x) - D_x V(x)\,ds \right\|$$

$$\leq t \sup_{0 \leq s \leq t} \left\| D_x V(\Psi_s(x)) D_x \Psi_s(x) - D_x V(x) \right\|$$

$$\leq t^2 \sup_{0 \leq s \leq t} \left\| D_t (D_x V(\Psi_s(x)) D_x \Psi_s(x)) \right\| \qquad \text{which by Lemma 1}$$

$$\leq \left[\frac{K_0 K_2}{1 - tK_1} + \frac{K_0 K_1 K_2 t}{(1-tK_1)^3} \right] t^2 \tag{IV}$$

This finishes the proof of theorem 1 in the case of the usual Euler method in Euclidean space. In fact, formulas III and IV show that there is a constant C such that in the one norm for t sufficiently small $||E_t - \Psi_t||_1 \leq Ct^2$.

Now we turn to the case of general B. We need only show that $E_t(x) = B_x(tV)$ and $x + tV(x)$ are tangent in the C^1 topology. The The hypotheses on B were intended to make this true. By Taylor's formula given $r > 0$, there is a constant $C > 0$ such that for $||w|| < r$,

$B(x,w) = x + \frac{\partial B}{\partial w}(x,0)^w + \frac{1}{2}\frac{\partial^2 B}{\partial w^2}(x,0)(w,w) + R(x,w)$ and $||R(x,w)|| < C_0 ||w||^2$.

Now $\frac{\partial B_x}{\partial w}(0)$ is the identity and $\frac{\partial^2 B_x}{\partial w^2}$ is bounded for $||w|| < r$. So for

$||w|| < r$, there is a constant C_1 such that $||B_x(w) - x + w|| < C_1 ||w||^2$.

Hence $||E_t(x) - (x + tV(x))|| = ||B_x(tV(x)) - (x + tV(x))||$

$$\leq C_1 ||tV(x)||^2 \leq C_1 K_0 t^2.$$

Now for the derivative...

$$||D_x(B(x,tV(x)) - D_x(x + tV(x))||$$

$$= ||D_{(x,w)} B(x,tV(x)) \circ D_x(x,tV(x)) - (I + tD_x V(x))||$$

$$= \left\| \left(\frac{\partial B}{\partial x}(x,tV(x)), \frac{\partial B}{\partial w}(x,tV(x))\right) \circ \begin{pmatrix} I \\ tD_x V(x) \end{pmatrix} - (I,I) \begin{pmatrix} I \\ tD_x V(x) \end{pmatrix} \right\|$$

$$= \left\| \left(\frac{\partial B}{\partial x}(x,tV(x)), \frac{\partial B}{\partial w}(x,tV(x))\right) - \left(\frac{\partial B}{\partial x}(x,0), \frac{\partial B}{\partial w}(x,0)\right) \begin{pmatrix} I \\ tD_x V(x) \end{pmatrix} \right\|$$

Since $\frac{\partial B}{\partial w}(x,0) = Id$, $\left(\frac{\partial^2 B}{\partial x \partial w}\right)(x,0) = 0$

and by the symmetry of the second derivative $\frac{\partial}{\partial w}\left(\frac{\partial B}{\partial x}\right)(x,0) = 0$. Thus there is a constant C_2 such that $\left|\left|\frac{\partial}{\partial w}\frac{\partial B}{\partial x}(x,w)\right|\right| < C_2||w||$ and by the mean value theorem $\left|\left|\frac{\partial B}{\partial x}(x,tV(x)) - \frac{\partial B}{\partial x}(x,0)\right|\right| \leq C_2||tV(x)||^2 \leq C_2K_0t^2$. (V)

There is a C_3 such that

$$\left|\left|\frac{\partial B}{\partial w}(x,w) - \frac{\partial B}{\partial w}(x,0)\right|\right| < C_3||w|| \text{ so } \left|\left|\left(\frac{\partial B}{\partial w}(x,tV(x)) - \frac{\partial B}{\partial w}(x,0)\right)tD_xV(x)\right|\right|$$

$$\left|\left|\left(\frac{\partial B}{\partial w}(x,tV(x)) - \frac{\partial B}{\partial w}(x,0)\right)tD_xV(x)\right|\right| \leq C_3||tV(x)||\ ||tD_xV(x)||$$

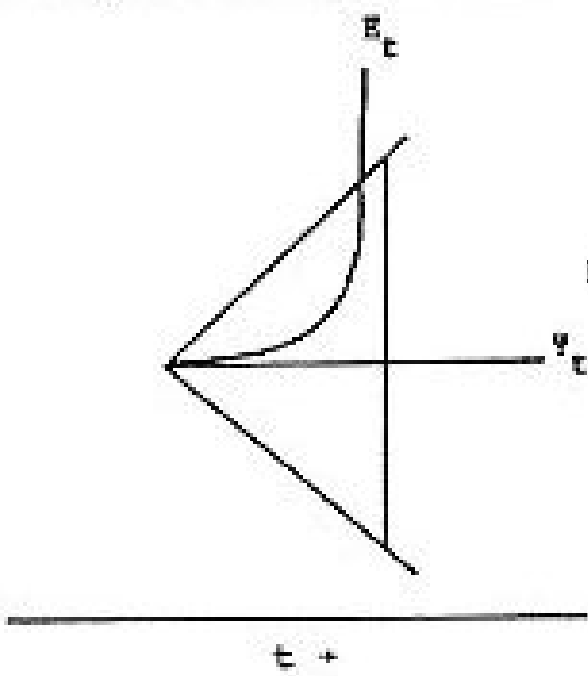
$$\leq C_3K_0K_1t^2. \qquad \text{(VI)}$$

Altogether we have shown that in the one norm there is a constant C_4 such that

$$||E_t(x) - (x + tV(x)||_1 \leq C_4t^2$$

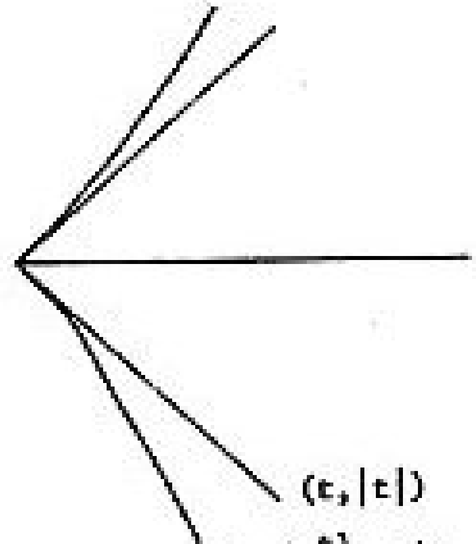
and we are done.

For application it is best to keep the explicit formulas III, IV, V, VI. Generally speaking many stability theorems for differential equations actually give stability for any map sufficiently C^1 close to Ψ_t, the neighborhood of validity decays linearly as t tends to zero, so E_t enters this neighborhood as t tends to zero.



Neighborhood of validity in C^1.

The easiest example of this phenomenon is the following. For $\lambda>0$, $\Psi_t(x) = e^{t\lambda}(x)$ taking R to R is a flow of expanding linear maps for $t>0$ and for each $t>0$ any linear map A such that $||A - \Psi_t|| < e^{t\lambda} - 1$ is also expanding. The size of this neighborhood decays slower than any multiple of t^2 since the derivative at 0 is 1.



The theorem implies the (in this case trivial) fact that the Euler approximation to the flow $E_t(x) = (1 + t\lambda)x$ enters this neighborhood and is expanding for small t. I am certain that other less trivial examples of the phenomenon are:

1) the neighborhood of persistence of a normally hyperbolic invariant manifold including a hyperbolic fixed point or a hyperbolic closed orbit for the flow Ψ_t (See Hirsch, Pugh, Shub and Braun, Hershenov)

2) the neighborhood of conjugacy for the time t map of a Morse-Smale gradient flow (See Palis-Smale)

3) the neighborhood of persistence of a pseudo-hyperbolic splitting on an overflowing invariant set for the flow Ψ_t (See Hirsch, Pugh, Shub)

but I don't know if the explicit estimates establishing these facts are in the literature so I will stop short of claiming that they are true. When one wants to exploit one of these facts one will have to do it with estimates anyway.

For example, one might exploit 3 to try to prove that the Lorenz attractor has Williams' presumed attracting bundle. The techniques would be to find one for E_t, t small enough such that Ψ_t is in the domain of persistence of E_t. The existence of such a t is guaranteed by the theorem and 3. This is a good project.

It is possible that higher order derivatives or higher order of contact could be useful, such as in the examination of bifurcation phenomena. It is easy to generalize theorem 1, which we now do, without keeping track of the constant.

For a C^{k+1} vector field $\dot{x} = V(x)$, write the solution as $\Psi(x,t)$. The k^{th} order Taylor method for the approximation of $\Psi(x,t)$ is simply

$T_k(x,t) = x + tD_t\Psi(x,0) + \frac{1}{2}t^2D_t^2\Psi(x,0) + \ldots + \frac{1}{k!}t^kD_t^k\Psi$.

The derivatives $D_t^j\Psi(x,0)$ are computed from V. For instance, $T_1(x,t) = x + tD_t\Psi(x,0) = x + tV(x)$ and is just $E_t(x)$.

The second derivative $D_t^2\Psi(x,t) = D_tV(\Psi(x,t)) = D_xV(\Psi(x,t))D_t\Psi(x,t)$ $= D_xV(\Psi(x,t))V(\Psi(x,t))$ and $D_t^2\Psi(x,0) = D_xV(x) \cdot V(x)$ so $T_2(x,t) =$

$$x + tV(x) + \tfrac{1}{2}t^2D_xV(x) \cdot V(x), \text{ etc.}$$

The spatial derivative of $T_k(x,t)$ applied to the tangent vector u

$D_xT_k(x,t)u = u + D_x(tD_t\Psi(x,0))u + D_x(\frac{1}{2}t^2D_t^2\ (x,0))u + \ldots$ $+ D_x(\frac{1}{k!}t^kD_t^k\Psi(x,0))u$. Since Ψ is C^{k+1}, the partials commute and

$D_xT_k(x,t)u = u + tD_t(D_x\Psi(x,0)u) + \frac{1}{2}t^2D_t^2(D_x\Psi(x,0)u) + \ldots + \frac{1}{k!}t^kD_t^k(D_x\Psi(x,0)u)$

which is just the k^{th} order Taylor approximation for the spatial derivative $D_x\Psi(x,t)u$. The flow $\gamma(x,u,t) = (\Psi(x,t), D_x\Psi(x,t)u)$ is the solution of the vector field $F(x,u) = (V(x), DV(x)u)$ and $(T_k(x,t),$ $D_xT_k(x,t)u)$ is the k^{th} order Taylor approximation to $\gamma(x,u,t)$. Thus C^0 estimates for $\gamma(x,u,t)$ and its k^{th} order Taylor approximation are C^0 and C^1 estimates for Ψ and its k^{th} order Taylor approximation. If V is C^{k+j}, this argument can be applied inductively to the first j derivatives of $T_k(x,t)$ and the first derivative of $\Psi(x,t)$.

<u>Theorem 2:</u>

Let M be a complete region in R^n with smooth boundary and let V be a C^{k+j} vector field $0 < k,\ j < \infty$ tangent to or pointing in along the boundary of M, which is bounded and which has bounded derivatives up to order k+j. Then there is a constant $C > 0$ and a $t_0 > 0$ such that

$||T_k(x,t) - \Psi(x,t)||_j < Ct^{k+1}$ for all $0 \leq t \leq t_0$.

Here the j norm means the sup of the norm of the differences of the values and the first j derivatives of $T_k(x,t)$ and $\Psi(x,t)$.

Proof:

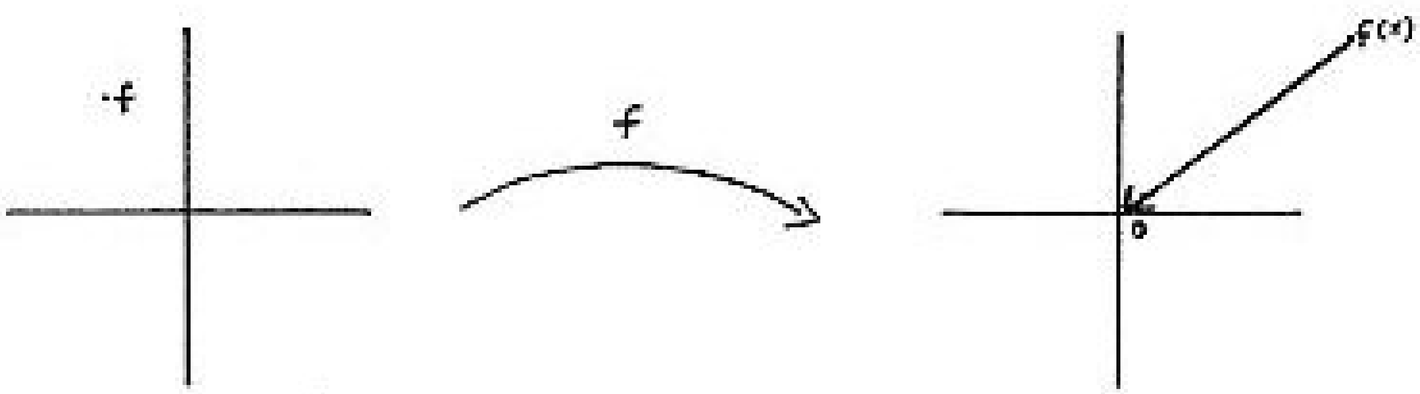
It suffices to prove the theorem for $j = 1$ by induction. It is well known and fairly simple to see from the Taylor series expansion that the estimate is valid in the C^0 norm once the C^k size of V is bounded. Thus we are done as soon as the C^0 size of $D_x\Psi(t,x)u$ is bounded along the orbit of $\Psi(t,x)$ for $0 \leq t \leq 1$ and $||u|| = 1$, but this follows from Lemma 1 c).

Section 2:

Given a vector field (or more generally a section of a vector bundle) f then we can ask to find the zeros of f. One approach would be to integrate f and hope (this is effective if $f = -\text{grad}\ \phi$), but also we may define the new vector field $V = -\frac{1}{2}\ \text{grad} <f,f> = -\frac{1}{2}\ \text{grad}\ ||f||^2$.
Other approaches are Newton's method, which we will define for general manifolds, but in R^n is

$$x \to x - (Df_x)^{-1}f(x), \text{ and the Newton vector field } N(x) = -(Df_x)^{-1}f(x)$$

If $f: R^n \to R^n$



Newton up to first order inverts f at f(x) and evaluates at 0 i.e. pulls the end point of the ray back. The Newton differential equation is $\dot{x} = N(x)$, its solution curves are mapped by f to solution of $\dot{y} = -y$, so they are the inverse images of the rays. Newton's method may be thought of as an approximation to the time one map of this equation by Euler's method. Integration of the Newton differential equation is thus another way to find the zeros of f (this equation has "poles" where the Newton method does).

In R^n, the gradient vector field $V(x) = -\frac{1}{2}\text{grad} \langle f(x), f(x)\rangle$ is determined by the linear map $w \to -\langle f(x), Df(x)w\rangle$ and is therefore $V(x) = -Df^t(x) \cdot f(x)$. By f this vector maps to $-Df(x) \cdot Df^t(x)f(x)$ which gives a negative inner product with $f(x)$;

$\langle -Df(x) \cdot Df^t(x)f(x), f(x)\rangle = -\langle Df^t(x)f(x), Df^t(x)f(x)\rangle \leq 0$ and hence

the image of a solution curve of the gradient equation $\dot{x} = V(x)$ by f is transversal to and inward pointing on spheres. In the case that Df is conformal Df^{-1} and Df^t differ by a real multiple. So the solution curves of $\dot{x} = N(x)$ and $\dot{x} = V(x)$ are the same. In particular, if f is an analytic function of one variable $f: C \to C$, then

$Df^t(z) = \overline{f'(z)}$, $(Df(z))^{-1} = \dfrac{1}{f'(z)}$ so $-\text{grad}||f(z)||^2 = \overline{f'(z)}f'(z)N(z)$

$= ||f(z)||^2 N(z)$ which is a positive real multiple of $N(z)$. Now let f be a complex polynomial $f: C \to C$. For each critical point θ of f, let

$W^s(\theta) = \{z \in C | \Psi_t(x) \to \theta \text{ as } t \to \infty\}$

$W^u(\theta) = \{z \in C | \Psi_t(x) \to \theta \text{ as } t \to \infty\}$

where θ_t is the solution of $\dot{z} = -\text{grad}|f(z)|^2$ and for any $z \in C$ denote by f_z^{-1} the branch of the inverse of f taking $f(z)$ back to z. The discussion above proves the following

<u>Proposition 1:</u> Let $f(z) = \sum_{i=0}^{d} a_i z^i$ be a non-constant complex polynomial.

a) The image by f of a solution curve of $-\text{grad}\,|f(z)|^2$ or $N(z)$ through the point z_0 lies on the half ray through $f(z_0)$ pointing towards the origin. If z_0 is not on the stable or unstable manifold of a critical point the image is the entire half ray. If $z_0 \in W^s(\theta)$ or $z_0 \in W^u(\theta)$ for a saddle point θ, then the image terminates at $f(\theta)$.

b) If $f(z_0) = w$ and $f'(z_0) \neq 0$, then the solution curve of $-\text{grad}|f(z)|^2$ or $N(z)$ through z_0 is the image of the half ray through w by the analytic continuation of the branch of f^{-1} taking w to $z_0, f_{z_0}^{-1}$.

c) If $f(\xi) = 0$ and $f'(\xi) \neq 0$, then the stable manifold $W^s(\xi)$ is the image by the analytic continuation f_ξ^{-1} which is defined on the whole complex plane minus a certain number of half lines from infinity to $f(\theta_i)$ where θ_i, $i = 1, \ldots\ldots, k$ are the critical points of f on the boundary of $W^s(\xi)$. It is tempting to try to find a solution to the polynomial equation $f(z) = 0$ by picking a point z_0 and either:

a) Take the solution curve of $-\text{grad}\,|f(z)|^2$ through z_0. With a finite number of exceptions this curve tends to a root of $f(z)$.

b) Take the solution curve of $N(z)$ through z_0. With a finite number of exceptions this curve tends to a root of $f(z)$.

c) Take $f_{z_0}^{-1}$ of the ray $(1-h)f(z_0)$ for $0 \leq h \leq 1$. With the exception of a finite number of rays substituting $h = 1$ gives a zero of f.

As Smale points out in Smale 1981, one can prove the fundamental theorem of algebra by these methods. Method c) is the easiest. Moreover, many numerical methods for solving polynomial equations are intimately connected to these theoretical methods. Hirsch-Smale, Smale 1981, and Shub-Smale 1982, 1983 discuss these points.

From Shub-Smale 1982, we have the following simple proposition.

Proposition 2:

Let $f: C \to C$ be a non-constant polynomial. Then $V(z) = -\text{grad}|f(z)|^2$ points in along any circle centered at zero which contains all the roots of f.

Proof:

We need only show that the inner product $\langle N(z), z\rangle < 0$ for any z on the circle.

$N(z) = \dfrac{f(z)}{f'(z)} = \dfrac{\frac{1}{f'(z)}}{f(z)} = \dfrac{1}{\Sigma \frac{1}{z-a}}$ where the a runs through the roots of f.

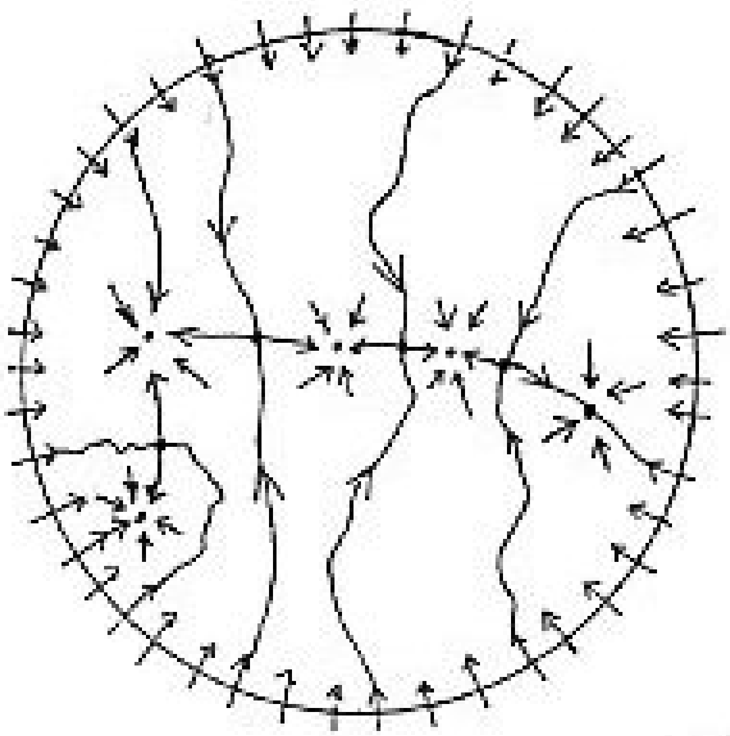
Now $\langle N(z), z\rangle = \text{Re}\left(\dfrac{1}{\frac{1}{z-a}}\,\bar{z}\right) < 0$ iff

$$\mathrm{Re}(\overline{\Sigma \frac{1}{z-a}}\ \bar{z}) < 0 \quad \text{iff}$$

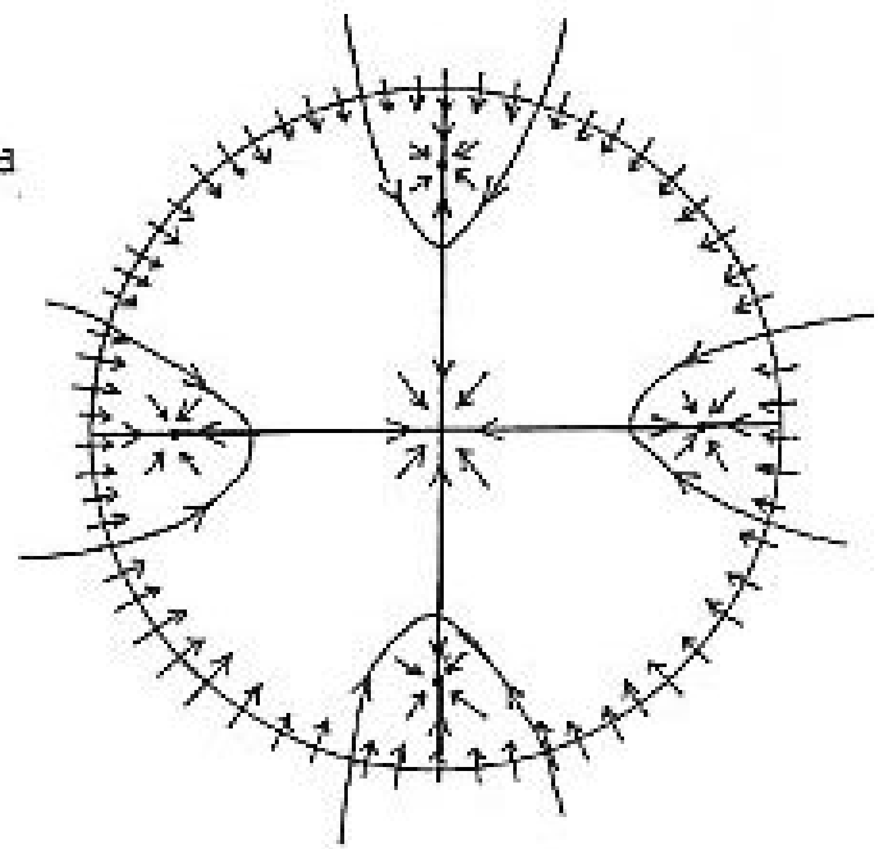
$$\mathrm{Re}(\Sigma \frac{1}{z-a}\ z) < 0$$

But now $\mathrm{Re}(z\ \overline{z-a}) < 0$ since the root a is inside the circle on which z lies. Thus $\mathrm{Re}(z\ \frac{1}{z-a}) < 0$ for each a individually and we are done.

If f has simple roots and f' as well then $|f(z)|^2$ is a Morse-function pointing in along the boundary of the disc and $-\mathrm{grad}\ |f|^2$ is generally a Morse-Smale vector field.



and



are conceivable for 5^{th} degree equations. The latter occurs for $z(z^{4}-1)$. The fact that $-\mathrm{grad}\ |f(z)|^2$ is generally Morse-Smale together with Section 1 can be used to give a preliminary indication of why numerical methods work, but see <u>Smale 1981</u> and <u>Shub-Smale 1982,1983</u> for much more of this.

Section 3:

To define Newton's method for the problem of finding the zeros of a vector field on a manifold, $V: M \to TM$ we employ the function $B_x:U_x \to M$ of Section 1, and assume that $V(x) \in U_x$. Then we let

$N(x) \equiv B_x((V'(x))^{-1}(-V(x)))$ where $V'(x):T_xM \to T_xM$, has to be made sense of via a connection and $(V'(x))^{-1}$ is assumed to exist. As for the usual Newton's method we have :

Proposition 3:

The fixed points of $N(x)$ correspond to the zeros of V. If V has simple zeros, $N'(x) = 0$ at each fixed point.

Proof:

Since B_x defines a chart $B_x(v) = x$ iff $v = 0$ and hence $N(x) = x$ iff $V(x) = 0$. The simplicity of the zero of V is the same as the assertion that $(V'(x))^{-1}$ exists. Now write in a chart

$$x \to (x, (V'(x))^{-1}(-V(x)) \qquad \text{and}$$

differentiate at a point x with $V(x) = 0$ in the direction v obtaining $(v, (V'(x)^{-1})'(v)(-V(x)) + (V'(x))^{-1}(-V'(x)(v)) = (v,-v)$. The derivative of B at $(x,0)$ is (Id,Id) so $D_x(N(x)) = 0$.

If we consider the linear map $A:R^n \to R^n$ then the zeros of the vector field $V(x) = A(x) - \frac{\langle Ax,x \rangle}{\langle x,x \rangle} x \equiv A(x) - \rho(x)x$ are the eigenvectors of A. $V(x) = A(x) - \langle Ax,x \rangle x$ defines a vector field on the sphere S^{n-1}.

Let $B_x:T_xS^{n-1} \to S^{n-1}$ be defined by $v \to \frac{x + v}{||x+v||}$ as in Section 1.

Of the various suggested methods to find the zeros of V we may solve $\dot{x} = V(x)$ on S^{n-1}, $\Psi_t(x) = \dfrac{e^{tA}(x)}{||e^{tA}(x)||}$ which is a Morse-Smale flow for A with eigenvalues with distinct real parts, and has normally hyperbolic invariant circles if the eigenvalues of A have distinct real parts except for complex conjugate pairs. These are generic and the induced dynamics are stable amone linearly induced flows. [See G. Palis] In fact for symmetric A, $V(x) = -\mathrm{grad}(\frac{1}{2}\langle Ax,x\rangle)$ which is a Morse function on S^{n-1}, when A has distinct eigenvalues.

Section 1 then asserts that the Euler approximation

$$E_t(x) = \frac{x + t(A(x) - \langle x,A(x)\rangle x)}{||x + t(A(x) - \langle x,A(x)\rangle x||}$$

has similar properties for small t. The fixed points are once again eigenvectors of A. As far as I know this is an unexplored and probably uninteresting method to find eigenvectors.

The Newton's method is more interesting.

$V_A(x) = V(x) = A(x) - \langle A(x),x\rangle x = A(x) - \rho(x)x$.

In R^n, $V'(x)v = A(v) - \langle A(v),x\rangle\ x - \langle A(x),v\rangle\ x\ - \rho(x)v$.

To project onto the tangent space of S^{n-1} subtract the projection on x (this is the induced connection).

$\langle V'(x)v,x\rangle = \langle A(v),x\rangle - \langle A(v),x\rangle\langle x,x\rangle - \langle A(x),v\rangle - \rho(x)\langle x,v\rangle$.

$\langle x,v\rangle = 0$ for v tangent to S^{n-1} at x so $\langle V'(x)v,x\rangle = -\langle A(x),v\rangle$ and $V'(x)v = A(v) - \langle Av,x\rangle x - \rho(x)v$. Now solving $V'(x)w = V(x)$ gives $A(w) - \langle A(w),x\rangle x - \rho(x)w = A(x) - \rho(x)x$ or $(A - \rho(x)I)w = (A - \rho(x)I)x + \langle A(w),x\rangle x$. Thus if we can solve for w and $(A - \rho(x)I)$ is invertible $w = x + (A - \rho(x)I)^{-1}\langle A(w),x\rangle x$. Newton's iteration then is $N_A(x) =$

$$\frac{x-w}{||x-w||} = \frac{-(A - \rho(x)I)^{-1}\langle A(w),x\rangle x}{||(A - \rho(x)I)^{-1}\langle A(w),x\rangle x||} = \pm\frac{(A - \rho(x)I)^{-1}x}{||(A - \rho(x)I^{-1}x||}$$ where

we have written N_A to stress the dependence on A. $N_A(-x) = -N_A(x)$, so N_A is defined on a subset of projective (n-1) space, RP(n-1). Up to sign N_A is a well known numerical method called the inverse power method with Rayleigh quotient shift. As above, $\rho_A(x) = \rho(x) = \frac{\langle x,Ax\rangle}{\langle x,x\rangle}$.

For $x \in R^n$, let $F_A(x) = (A - \rho(x)I)^{-1}x$.

<u>Lemma 2:</u>

For $x \in R^n$, $x \neq 0$, $\lambda \in R^n$, $\lambda \neq 0$, $A:R^n \to R^n$ linear and $O:R^n \to R^n$ an orthogonal linear map

1) $\rho_A(x) = \rho_A(\lambda x)$

2) $\rho_{\lambda A}(x) = \lambda\rho_A(x)$

3) $\rho_{O^{-1}AO}(x) = \rho_A(Ox)$

4) $F_A(\lambda x) = \lambda F_A(x)$

5) $F_{\lambda A}(x) = \lambda F_A(x)$

6) $F_{O^{-1}AO} = O^{-1}F_A O$

Proof:

1) $\rho_A(x) = \dfrac{\langle A(x),x\rangle}{\langle x,x\rangle} = \dfrac{\langle A\lambda x,\lambda x\rangle}{\langle x,\ x\rangle} = \rho_A(\lambda x)$

2) $\rho_{\lambda A}(x) = \dfrac{\langle \lambda Ax,x\rangle}{\langle\ x,x\rangle} = \lambda\rho A(x)$

3) $\rho_{O^{-1}AO}(x) = \dfrac{\langle O^{-1}AOx,x\rangle}{\langle x,x\rangle} = \dfrac{\langle AOx,(O^{-1})^t x\rangle}{\langle Ox,Ox\rangle} = \dfrac{\langle AOx,Ox\rangle}{\langle Ox,Ox\rangle} = \rho_A(Ox)$

4) $F_A(\lambda x) = (A - \rho_A(\lambda x)I)^{-1}\lambda x = (A - \rho_A(x)I)^{-1}\lambda x = \lambda F_A(x)$

5) $F_{\lambda A}(x) = (\lambda A - \rho_{\lambda A}(x)I)^{-1}x = (\lambda A - \lambda\rho_A(x)I)^{-1}x = \lambda F_A(x)$

6) $O^{-1}F_A O(x) = O^{-1}(A - \rho_A(Ox)I)^{-1}Ox = (O^{-1}(A - \rho_{O^{-1}AO}(x)I)O)^{-1}x$

$= (O^{-1}AO - \rho_{O^{-1}AO}(x)I)^{-1}x = F_{O^{-1}AO}(x)$.

Lemma 2.4 implies that F_A may be defined on a subset U of the projective space of R^n, $RP(n-1)$ where $(A - \rho_A(x)I)^{-1}$ exists. Denote this map by R_A, it agrees with the map which N_A induces on projective space. R_A has, for this context, another rather unnatural lift from projective space to the sphere S^{n-1} by the formula

$$\tilde{R}_A(x) = \frac{(A - \rho(x)I)^{-1}(x)}{||(A - \rho(x)I)^{-1}(x)||}$$

for x a unit vector such that $A - \rho(x)I$ is invertible. Parlett calls this map the Rayleigh quotient iteration. It seems more appropriate to refer to R_A or N_A by this name. N_A is a different lift of R_A to S^{n-1}.

Given the orthogonal transformation O, it acts on projective space by acting on lines.

Proposition 3:

Let A be linear, O orthogonal and λ real non-zero. Then $R_{\lambda A} = R_A$ and $O^{-1}R_A O = R_{O^{-1}AO}$.

Proof:

This follows immediately from Lemmas 2.5 and 6.

Thus the dynamics of R_A on projective space only depend on the orthogonal canonical form of the matrix which can be adjusted by a scalar multiple. This can be seen directly for $N_A(x)$ as well.

Proposition 4:

Let A be linear, O orthogonal and λ real non-zero. Then $N_{\lambda A} = N_A$ and $N_{O^{-1}AO} = O^{-1}N_A O$.

Proof:

$V_{\lambda A} = \lambda V_A$ and $(V'_{\lambda A})^{-1} = \frac{1}{\lambda}(V'_{\lambda A})^{-1}$ so this proves that $N_{\lambda A} = N_A$.

$$\begin{aligned} O^{-1}V_A O(x) &= O^{-1}AO(x) - O^{-1} \langle AOx, Ox\rangle\, Ox \\ &= O^{-1}AO(x) - \langle O^{-1}AOx, x\rangle\, x \\ &= V_{O^{-1}AO}(x). \quad \text{Thus} \end{aligned}$$

$O^{-1}V'_A(Ox)O = V'_{O^{-1}AO}(x)$ and $N_{O^{-1}AO}(x) = B_x((V'_{O^{-1}AO}(x))^{-1}(V_{O^{-1}AO}(x))$

$$\begin{aligned} &= B_x(O^{-1}(V'_A(O(x)))^{-1}(V_A(O(x))) \\ &= O^{-1}B_{O(x)}(V'_A(O(x))^{-1}V_A(O(x)) = O^{-1}N_A(O(x)). \end{aligned}$$

For a non-zero vector x in R^n, let [x] denote the equivalence class of x in RP(n-1). So $R_A[x] = [F_A(x)]$.

Now we follow Parlett for an analysis of R_A for symmetric A.

Proposition 5:

Let A be a symmetric linear map, $A:R^n \to R^n$. Then there is a neighborhood W of the eigenspaces of A in RP(n-1) such that any $w \in W$ either converges cubically to an eigenline of A under iterations of R_A or is itself already an eigenline of A.

Proof:

Let z be a unit eigenvector, $A(z) = \lambda z$. The perpendicular space to the λ eigenspace is left invariant by A. Call this map $A^{\perp}$ and $\rho^{\perp}$, ρ on the perpendicular space. Let U be orthogonal to z. Then

$$[(A - \rho(z + u)I)^{-1}(z + u)] = [\frac{z}{\lambda - \rho(z+u)} + (A^{\perp} - \rho(z + u)I)^{-1}u]$$

$$= [z + (\frac{1}{1+||u||^2})(\lambda - \rho^{\perp}(u))||u||^2(A^{\perp} - \rho(z + u)I)^{-1}u]$$

Now if the norm of u is small $\rho(z + u)$ is close to λ. As $A^{\perp}$ does not have λ as an eigenvalue there is a constant $c > 0$ such that for small u, $||(A^{\perp} - \rho(z + u)I)^{-1}u|| \leq C||u||$ and thus there is a constant $k \geq 0$ such that for $||u||$ small enough and u orthogonal to the eigenspace of λ, $[(A - \rho(z + u)I)^{-1}(z + u)] = [z + v]$ where v is orthogonal to the λ eigenspace and $||v|| \leq k||u||^3$. This is cubic convergence to the eigenvectors.

Now we do a particularly simple case, A is 2 x 2 symmetric. To study the dynamics we suppose that A has been diagonalized $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ with $\lambda_1 < \lambda_2$(If $\lambda_1 = \lambda_2$, R_A is the identity.)

$$\rho(x,y) = \frac{\lambda_1 x^2+\lambda_2 y^2}{x^2+y^2} \quad \text{so}$$

$$R_A[(x,y)] = \left[\frac{x}{\lambda_1-\rho(x,y)}, \frac{y}{\lambda_2-\rho(x,y)}\right] \quad \text{and}$$

$$R_A[(1,y)] = \left(\frac{1}{\lambda_1 - \left(\frac{\lambda_1+\lambda_2 y^2}{1+y^2}\right)}, \frac{y}{\lambda_2 - \left(\frac{\lambda_1+\lambda_2 y^2}{1+y^2}\right)}\right)$$

$$= \left(\frac{1}{(\lambda_1-\lambda_2)y^2}, \frac{y}{\lambda_2-\lambda_1}\right) = [1,-y^3].$$

Similarly

$R_A[(x,1)] = [(-x^3,1)]$.

Thus the dynamics are independent of A. Every point tends precisely cubically to one of the two fixed points with the exception of the orbit of the point $[(1,1)]$ which has period two. The derivative of the square at $[(1,1)]$ is 9. So $[(1,1)]$ is a source. $\rho_A(1,1)=\frac{1}{2}(\lambda_1+\lambda_2)$ as it does at the other point in the orbit. The vector $(1,1)$ is an eigenvector of $(A - \rho_A(1,1)I)^2$.

For general symmetric A, and $[x_0] \in RP(n-1)$ define $[x_i]$ inductively by $R_A[x_{i-1}] = [x_i]$ as long as R_A is defined at $[x_{i-1}]$. The Rayleigh quotient iteration converges for almost every starting point. In fact, we have the next theorem which is taken from Parlett. Let $\rho_k = \rho(x_k)$.

Theorem 3:

Let A be a symmetric, real n x n matrix. Let $[x_0] \in RP(n-1)$, and let R_A be the Rayleigh quotient iteration. Then either

(1) $\rho[x_k]$ is an eigenvalue of A for some $k > 0$

or

(2) $[X_k]$ is defined for all $k \geq 0$ and in this case ρ_k converges and either

(i) $\rho_k \to \lambda$ and $[x_k] \to [z]$ cubically where $A(z) = \lambda z$

or

(ii) $[x_{2k}] \to [x_+]$, $[x_{2k+1}] \to [x_-]$

linearly where x_+ and x_- are the bisectors of a pair of eigenvectors whose eigenvalues have mean $\rho = \lim \rho_k$.

Proof:

The proof of this theorem is rather clever. First consider N_A acting on S_1^{n-1}, x_0 given and $x_{i+1} = N_A(x_i)$ where defined, so x_i is a unit vector. Let $r_k = (A - \rho(x_k)I)x_k$, $||r_k||$ is monotonically decreasing for:

$$||r_{k+1}|| = ||(A - \rho(x_{k+1})I)x_{k+1}||$$

$$\leq ||(A - \rho(x_k)I)x_{k+1}||$$

$= |\langle x_k, (A - \rho(x_k)I)x_{k+1}\rangle|$ since the second vector is a multiple of x_k which is a unit vector

$\leq ||A - \rho_k(x)I)x_k|| \; ||x_{k+1}||$ by the symmetry of A and Cauchy-Schwartz $= ||r_k||$. Moreover, $||r_k|| = ||r_{k+1}||$ iff $\rho_{k+1} = \rho_k$ from the first inequality and $(A - \rho_k I)x_k = \alpha(A - \rho_k I)^{-1}(x_k)$ for some $\alpha \in R$ by the second inequality, so $(A - \rho_k I)^2 x_k = \alpha x_k$.

Now if $||r_k|| \to 0$, it is not too difficult to see that $[x_k]$ is ultimately in the domain of cubic convergence of an eigenvector.

If $||r_k|| \to \tau > 0$, then if $[z]$ is an accumulation point of $[x_k]$, we see that $||r_{k+1}(z)|| = ||r_k(z)||$ by continuity so z is an eigenvector of $(A - \rho(z)I)^2$ but not of $A - \rho(z)I$ (because $\tau > 0$) and this only happens when z bisects a pair of eigenvectors for A. Using continuity one can now essentially finish the argument by the analysis we have done above of the two dimensional case.

The convergence properties of the Rayleigh quotient iteration for general real matrices are unknown. We say that the eigenvector x of A is simple if $A(x) = \lambda x$ for real λ and the dimension of the generalized λ eigenspace of A is one. Convergence to simple eigenvectors is quadratic by:

Proposition 6:

Let x be a unit vector and a simple eigenvector of $A: R^n \to R^n$. Then $(N_A(x))' = 0$.

Proof:

We need only show that $V'_A(x)$ has no kernel. But

$$V'_A(x)v = Av - \lambda v - \langle Av, x\rangle x \qquad \text{and}$$

$(A - \lambda I)v$ cannot be a multiple of x for any v independent of x.

The case of a complex eigenvalue is quite different. Indeed for a rotation of the plane $A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ with $\theta \neq 0 \bmod(2\Pi)$,

$(A - \rho(x)I)^{-1} = \begin{pmatrix} 0 & -(\sin\theta)^{-1} \\ (\sin\theta)^{-1} & 0 \end{pmatrix}$ independent of x and on projective space induces the same map as $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ independent of A. Thus every point has period two and the orbit gives an orthogonal basis of the space. This dynamics is analagous to Theorem 3.2ii. For two by two matrices the dynamics of the Rayleigh quotient iteration can be completely described.

Proposition 7:

Let $A:R^2 \to R^2$ be linear and $R_A:RP(1) \to RP(1)$ the Rayleigh quotient iteration. Then either

1) A is a multiple of the identity and R_A is not defined

2) R_A has period two, A has complex eigenvalues

3) R_A has one orbit of period two, two fixed points and every other point has quadratic convergence to one of the fixed points; A has two distinct real eigenvalues

4) R_A has one fixed point, every other point converges linearly to the fixed point; A has two equal real eigenvalues but is not a multiple of the identity.

The case of complex eigenvalues not conformally conjugate to a rotation such as $\begin{pmatrix} 0 & -1 \\ c & 0 \end{pmatrix}$ has been neglected in this analysis as $c \to 0$ the dynamics of R_A go through period doubling all the way to period 3. (Batterson)

Proof:

By Proposition 3 and the discussion above we have only to consider the case that with respect to the standard basis of R^2, $A = \begin{pmatrix}\lambda_1 & \varepsilon \\ 0 & \lambda_2\end{pmatrix}$.

Computing in the coordinate chart $\{(1,y)\}$ for RP(1)

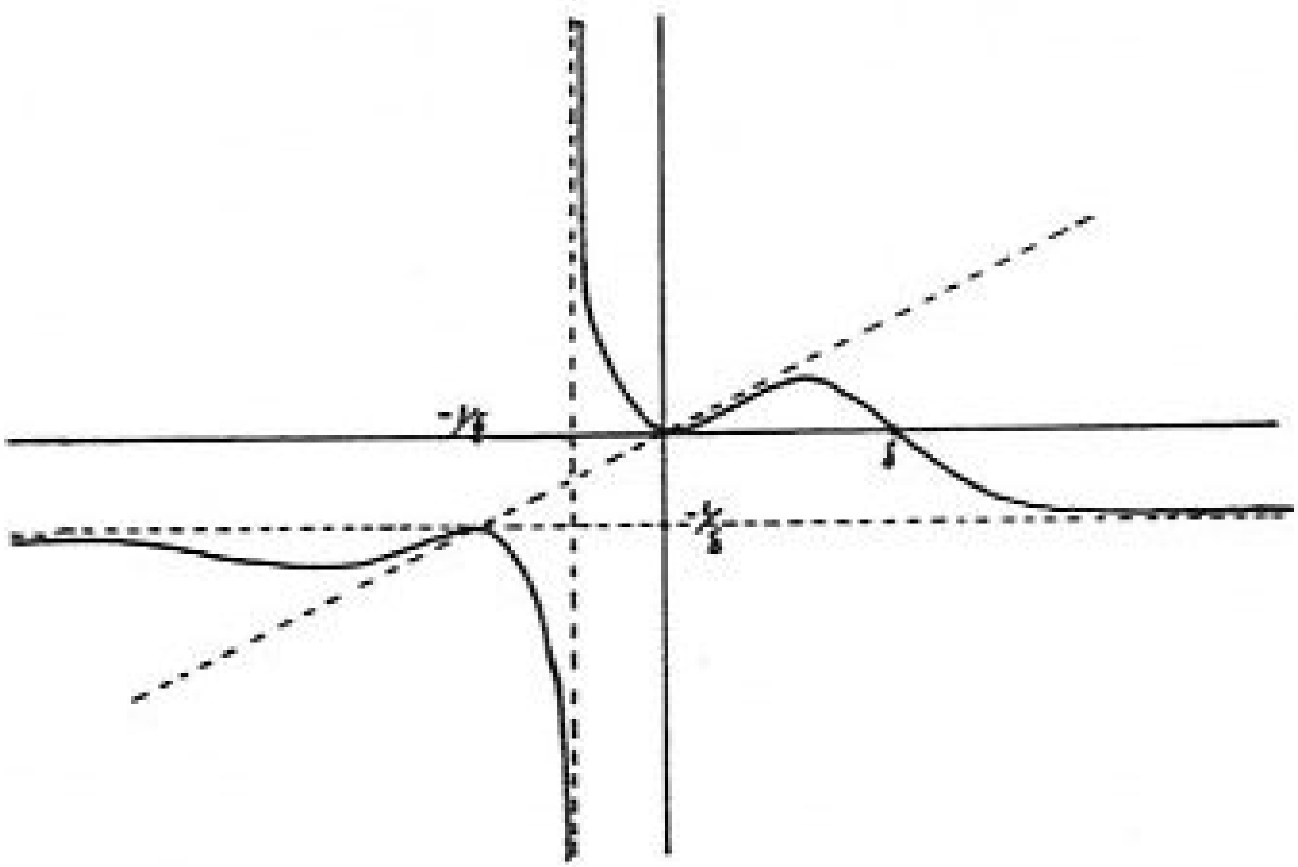
$R_A(y) = \dfrac{(\lambda_1 - \lambda_2)y^3 - \varepsilon y^2}{(\lambda_2 - \lambda_1) - 2\varepsilon y - \varepsilon y^3}$. If $\lambda_1 = \lambda_2$ and $\varepsilon \neq 0$, then

$R_A(y) = \dfrac{y}{y^2+2}$. This is case 4). The dynamics are independent of λ, ε

and converge globally linearly to $y = 0$. If $\lambda_1 \neq \lambda_2$ and $\varepsilon = 0$, this is the symmetric case discussed above. It is a special case of 3). R_A is a homeomorphism and the convergence to the fixed points is cubic. The last case to consider is $\lambda_1 \neq \lambda_2$ and $\varepsilon \neq 0$.

Let $\delta = \varepsilon(\lambda_1 - \lambda_2)^{-1}$.

So $R_\delta(y) = R_A(y) = \dfrac{\delta y^2 - y^3}{\delta y^3 + 2\delta y + 1}$. The fixed points of this transformation are $y = 0$ and $y = -\frac{1}{\delta}$ and these are also critical points by Proposition 6. The other two critical points of R_δ are real and are at $y = \dfrac{-3 \pm \sqrt{9+8\delta^2}}{2\delta}$. We assume that δ is positive since R_δ and $R_{-\delta}$ are conjugate by $y \to -y$. For $y < -\frac{1}{\delta}$, $R_\delta(y) < -\frac{1}{\delta}$. Therefore it is simple to see that the graph looks like



The critical points tend to the super-attractive fixed points. Therefore, it follows from the work of Fatou (See Blanchard) that even as a rational map of the Riemann sphere it is hyperbolic. On the left every point tends to $-\frac{1}{\delta}$. On the right, the basin of attraction of 0 is delimited by a point of period two and outside this basin the map is monotonic on the relevant intervals. Since there can be no other sink (a critical point would have to tend to it, again by the work of Fatou) every point on the right outside the basin of 0 besides the repelling point of period two tends to the fixed point $-\frac{1}{\delta}$. q.e.d.

It is simple to see that for $\delta > 0$, all the R_δ are topologically conjugate, and interesting to note by Sullivan and Mañé, Sad and Sullivan, it follows that the entire family of rational maps of the sphere R_δ, $\delta > 0$ are quasi-conformally conjugate. Hence the Julia sets are all quasi-circles with dynamics topologically conjugate to $y \to -y^3$. No two are conformally conjugate because in the complex plane the other fixed points are just $-i$ and i.

Problem:

Fix $n > 0$. For almost every (A,v) where $A:R^n \to R^n$ is linear and $v \in RP(n-1)$, does $R_A^k(v)$ tend to a super-attractive fixed point or a point of period two?

References

1. K. Atkinson, 1978, An Introduction to Numerical Analysis, John Wiley and Sons, New York.
2. P. Blanchard, 1984, Complex Analytic Dynamics, Bulletin of the A.M.S. (New Series) Vol. 11, pp. 85-141.
3. M. Braun and J. Hershenov, 1977, Periodic Solutions of Finite Difference Equations, Quarterly of Applied Mathematics (April), pp. 139 - 147.

4. M.W. Hirsch, C.C. Pugh and M. Shub, 1977, Invariant Manifolds, Springer Lecture Notes in Mathematics Number 583, Springer-Verlag, Berlin.

5. M.W. Hirsch and S. Smale, 1979, On Algorithms for Solving f(x)=0, Communications on Pure and Applied Mathematics, Vol. 32, pp. 281 - 312.

6. R. Mañé, P. Sad and D. Sullivan, 1983, On the Dynamics of Rational Maps, Annales Scientifiques de l'Ecole Normale Superieure, pp. 193 - 217.

7. G. Palis, 1978, Linearly Induced Vector Fields and R^2 Actions on Spheres, Journal of Differential Geometry, Vol. 13, pp. 163 - 190.

8. J. Palis and S. Smale, 1970, Structural Stability Theorems in Proceedings of Symposium in Pure Mathematics, Vol. 14 (Edited by S.S. Chern and S. Smale), A.M.S., Providence, R.I., pp. 223 - 232.

9. B. Parlett, 1980, The Symmetric Eigenvalue Problem, Prentice Hall, Englewood Cliffs, New Jersey.

10. M. Shub and S. Smale, 1982, 1983, Computational Complexity: On the Geometry of Polynomials and a Theory of Cost: Part I (Preprint to appear in the Annales Scientifiques de l'Ecole Normale Superieure, Part II (Preprint submitted to SIAM Journal on Computing.

11. S. Smale, 1981, The Fundamental Theorem of Algebra and Complexity Theory, Bulletin of the A.M.S. (New Series), Vol. 4, pp. 1 - 36.

12. D. Sullivan, 1982, 1983, Quasi-Conformal Homeomorphisms and Dynamics I and II, Preprints to appear in Annals of Mathematics.