0
0
0

A geometric proof of the spectral theorem for real symmetric matrices

Robert Sachs

Department of Mathematical Sciences
George Mason University
Fairfax, Virginia 22030

rsachs@gmu.edu

January 6, 2011

Introduction

1. Many students find this topic difficult

Introduction

1. Many students find this topic difficult
2. Teaching history of math for first time: Euler, Cauchy, Sturm, Weierstrass, Fischer, Weyl, Courant

Introduction

1. Many students find this topic difficult
2. Teaching history of math for first time: Euler, Cauchy, Sturm, Weierstrass, Fischer, Weyl, Courant
3. In usual proof orthogonality is “accidental” via symmetric matrix and inner product

Introduction

1. Many students find this topic difficult
2. Teaching history of math for first time: Euler, Cauchy, Sturm, Weierstrass, Fischer, Weyl, Courant
3. In usual proof orthogonality is “accidental” via symmetric matrix and inner product
4. Working on multivariable calculus book and want to do Lagrange multiplier idea without assuming linear algebra

REPRISE OF USUAL ARGUMENTS

- **Three main strategies: algebraic, analytic, computational**
- **Algebraic works from Invariant Subspaces, Minimal Polynomial, Show Orthogonality, Geometric and Algebraic Dimensions Equal.**
- **Analytic uses Lagrange Multipliers, Orthogonality constraints (later seen inactive),**
- **Numerical uses Givens Rotations (Euler for principal axes in 3-D), Orthogonality leads to symmetric diagonalization**

KEY BACKGROUND

- Quadratic equations tied to matrix form: $\mathbf{Q}(\mathbf{v}) = \mathbf{v}^{\mathbf{T}}\,\mathbf{H}\,\mathbf{v}$ with $\mathbf{H}$ symmetric

KEY BACKGROUND

- Quadratic equations tied to matrix form: $\mathbf{Q}(\mathbf{v}) = \mathbf{v}^{\mathbf{T}}\,\mathbf{H}\,\mathbf{v}$ with $\mathbf{H}$ symmetric
- $\mathbf{n} = \mathbf{2}$ case is cute use of trig (more below) and the key step

KEY BACKGROUND

- Quadratic equations tied to matrix form: $\mathbf{Q}(\mathbf{v}) = \mathbf{v}^{\mathbf{T}}\,\mathbf{H}\,\mathbf{v}$ with $\mathbf{H}$ symmetric
- $\mathbf{n} = \mathbf{2}$ case is cute use of trig (more below) and the key step
- Scaling quadratically suggests looking on unit sphere

KEY BACKGROUND

- Quadratic equations tied to matrix form: $\mathbf{Q}(\mathbf{v}) = \mathbf{v}^\mathbf{T}\,\mathbf{H}\,\mathbf{v}$ with $\mathbf{H}$ symmetric
- $\mathbf{n} = \mathbf{2}$ case is cute use of trig (more below) and the key step
- Scaling quadratically suggests looking on unit sphere
- Min and max on sphere are eigenvectors (Lagrange multipliers for unit vector constraint)

KEY BACKGROUND

- Quadratic equations tied to matrix form: $\mathbf{Q}(\mathbf{v}) = \mathbf{v}^{\mathbf{T}}\,\mathbf{H}\,\mathbf{v}$ with $\mathbf{H}$ symmetric
- $\mathbf{n} = \mathbf{2}$ case is cute use of trig (more below) and the key step
- Scaling quadratically suggests looking on unit sphere
- Min and max on sphere are eigenvectors (Lagrange multipliers for unit vector constraint)
- Restrictions to subspaces are also quadratic forms

THE CASE OF $\mathbf{n = 2}$

- $\mathbf{Q(x, y) = a\,x^2 + 2b\,x\,y + c\,y^2}$
- Matrix form:

$$\mathbf{r}^{\mathbf{T}} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mathbf{r}$$

  where **r** is position vector.
- On unit circle, $\mathbf{x} = \cos \mathbf{t}$ and $\mathbf{y} = \sin \mathbf{t}$
- Restricted form is $\mathbf{a}\ \cos^2 \mathbf{t} + \mathbf{2\,b}\ \cos \mathbf{t}\ \sin \mathbf{t} + \mathbf{c}\ \sin^2 \mathbf{t}$
- Reexpressed as $\frac{\mathbf{a+c}}{\mathbf{2}} + \frac{\mathbf{a-c}}{\mathbf{2}} \cos \mathbf{2t} + \mathbf{b} \sin \mathbf{2t}$ and also as $\frac{\mathbf{a+c}}{\mathbf{2}} + \mathbf{A} \cos(\mathbf{2t} + \phi)$
- Amplitude $\mathbf{A}$ satisfies $\mathbf{A^2} = (\frac{\mathbf{a-c}}{\mathbf{2}})^{\mathbf{2}} + \mathbf{b^2} = (\frac{\mathbf{a+c}}{\mathbf{2}})^{\mathbf{2}} + (\mathbf{b^2} - \mathbf{ac})$ which leads to description of max and min values as well as average value over circle.
- Orthogonality of min and max vectors is basic trig!
- Role of discriminant / determinant in definiteness

MOVING UP A DIMENSION VIA MIN-MAX

- Consider three variable case
- Max, min values are usual Lagrange multiplier rule (multivariable calculus)
- Third orthogonal direction as eigenvector less clear
- Use previous step and restriction to **any** plane through origin ... have orthogonality of **restricted** max, min directions

MOVING UP A DIMENSION AND THE MIN-MAX ISSUE – cont.

- Restricted form is a nice linear algebra calculation: if we look at vectors $s\mathbf{v_1} + t\mathbf{v_2}$ in quadratic form, it becomes:

$$\begin{pmatrix} s & t \end{pmatrix} \mathbf{V^T} \mathbf{H} \mathbf{V} \begin{pmatrix} s \\ t \end{pmatrix}$$

  where V has vectors $\mathbf{v_1}$ and $\mathbf{v_2}$ as columns. More on this at end.
- Each plane containing origin has minimizing direction orthogonal to maximizing direction, so **if we max over mins on 2-D subspaces** with unit vector restriction, can restrict to a great circle orthogonal to maximizing direction.
- Key point: Claim that extreme vector is an eigenvector also ... i.e. gradient is aligned in the direction of the vector.

EIGENVECTOR CLAIM / ORTHOGONALITY

- In 3-D scenario, at minimax location, gradient of quadratic vanishes in admissible variation direction (angular along great circle)
- Why can't gradient have component in direction of maximal eigenvector? Which direction? (both vector and its negative are critical points, same value!)
- Conclude: zero component BY REFLECTION SYMMETRY / EVENNESS OF QUADRATIC!!
- This holds for all 3-planes in n dimensions – each comes algebraically as having a symmetric matrix hence quadratic form on restriction. Details later as time permits.

MIN-MAX VIEW OF EIGENVALUES

- For all 2-D subspaces, can take min-max or max-min which in 3-D happen at the same place.
- For higher dimensions, the min-max and max-min are typically different.
- Fischer seems to be the first to do this; Courant exploited it more fully (a Wikipedia discussion on this is useless).
- Continue inductively, building on higher dimensional subspaces with orthogonality going up with dimension.

VISUALIZATIONS

VISUALIZATIONS – continued

VISUALIZATIONS – continued

VISUALIZATIONS – continued

SOME DETAILS OF USUAL PROOFS

- STEP 1: Eigenvalues must be real.
- Suppose not, then there is a complex conjugate pair of roots of characteristic polynomial since matrix is real.
- Complex eigenvalue implies complex eigenvector
- Use complex conjugate and transpose together, i.e. Hermitian conjugate, to get a contradiction, as follows:
- $\bar{\mathbf{v}}^T A\mathbf{v} = \lambda \bar{\mathbf{v}}^T \mathbf{v}$ from original equation $A\mathbf{v} = \lambda \mathbf{v}$ after left multiplication by $\bar{\mathbf{v}}^T$. But taking complex conjugate transpose of $A\mathbf{v} = \lambda \mathbf{v}$ and then right multiplying by $\mathbf{v}$ we get (using $A^T = A$ and A real) the same left hand side but on the right $\bar{\lambda} \bar{\mathbf{v}}^T \mathbf{v}$ so we conclude, since $\mathbf{v} \neq$, that $\lambda = \bar{\lambda}$

SOME DETAILS OF USUAL PROOFS – continued

Now there is a fork in the road – algebra proof vs. analysis proof. First one of the algebra versions:

- STEP 2a: Strip off rank one piece, look on orthogonal complement of span of first eigenvector.
- Show that if $v^T \mathbf{e}_1 =$, then $(Av)^T \mathbf{e}_1 =$
- Done with our favorite algebraic lemma.
- Create new orthonormal basis starting with $\mathbf{e_1}$ then write matrix in that basis, find $(n-1) \times (n-1)$ block and λ_1 in upper corner with rest of first row/column zeroed out.
- STEP 3a: Continue in dimension $n - 1$, adding in first entry to get back to original dimension. Find second eigenvector, repeat STEP 2a.

SOME DETAILS OF USUAL PROOFS – continued

- STEP 2b: Find maximum with two constraints: unit vectors, also orthogonal to first eigenvector.
- Two Lagrange multipliers – one for unit vector (λ) and a second one for orthogonality (μ).
- Equation: $Av = \lambda\, v + \mu\, \mathbf{e_1}$
- And then a miracle happens: $\mu =$ by our favorite lemma.
- Find second eigenvector, then add that constraint, which is also inactive – repeat until done.

SUBSPACE AND RESTRICTION

- In subspace the vectors are linear combinations of some basis elements – columns of a rectangular matrix

- View it as matrix product – algebra leads to restricted matrix of form: $C^T A C$ where C has columns given by basis vector – new matrix is lower rank, symmetric.

EXTENSION / INTEGRATION

- Student question: What is gradient of $v^T A v$ when A is square but not symmetric?
- Representative of equivalence class of A under similarity – issue of transpose vs. inverse
- Length of vectors in subspace squared (case of $H = I$) is useful in thinking about surfaces, differential geometry (First Fundamental Form)
- Spherical and ultraspherical coordinates on unit n-sphere

EXTENSION / INTEGRATION – cont.

- Attempt to develop theory for constrained max/min and Hessian matrix
- Eigenvalues of $A^T A$ when A is rectangular
- Complex eigenvalues for non-symmetric real A – what do they mean geometrically
- For complex vector spaces, how is symmetric matrix extended?

EXTENSION / INTEGRATION (continued)

- Non-symmetric matrices and Jordan form
- Schur Factorization
- Ratios of quadratic expressions (non-zero denominator!) ties to Curvature calculations on surfaces.
- Classical view near max/min in 2-D: values taken on twice leads to discriminant – classic principal curvatures computation.
- Fun in number theory: quadratic forms – normal form using integer lattice transformations

CONCLUDING REMARKS

- Orthogonality now comes from 2-D geometry exploited ruthlessly
- Rich area for visualization, experiment, conjecture in high school
- Hate to banish really slick lemmas – love the algebraic fun