Derivatives of eigenvalue functions

Brendan Ames

Institute for Mathematics and its Applications
University of Minnesota

IMA Postdoc seminar
Tuesday, December 13, 2011

Motivation

- Optimization problems involving eigenvalues are ubiquitous.
  - Semidefinite programming.
  - Graph partitioning and spectral clustering.
  - Sparse optimization (especially rank minimization and its nuclear norm relaxation).
  - Choosing preconditioners in numerical analysis.
  - Quantum mechanics.
- Most algorithms need derivatives (at least (sub)gradient, maybe Hessian).
- Want formulas for derivatives of functions of eigenvalues.

Big Question

- Given a symmetric map $A(t)$ that depends smoothly on parameter t.
- How much of this smoothness is transferred to $\lambda_m(A(t))$?

Big Question

- Given a symmetric map $A(t)$ that depends smoothly on parameter t.
- How much of this smoothness is transferred to $\lambda_m(A(t))$?

- Given a map $F(X)$ that depends smoothly on $\lambda(X)$.
  - i.e. $F(X) = G(\lambda(X))$ for some smooth function G.
- How much of the smoothness of G is inherited by F?

Outline

Outline

Eigenvalues recap

- Given square matrix $A \in \mathbf{R}^{n \times n}$, the scalar $\lambda \in \mathbf{C}$ is an eigenvalue if there exists $v \neq 0 \in \mathbf{R}^n$ such that

$$Av = \lambda v$$

- The eigenvalues of A are the roots of the characteristic polynomial $det(A - \lambda I)$.

- The map $A \mapsto \lambda(A)$ is continuous.

Symmetric matrices

- A matrix A is symmetric if $A = A^T$.
- Symmetric matrices have real eigenvalues.
- There exists orthogonal $U \in \mathbf{R}^{n \times n}$ such that

$$A = U(Diag\,\lambda(A))U^T.$$

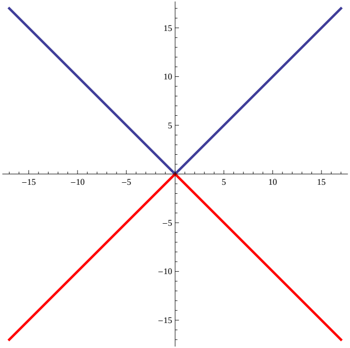
- Can assume that the entries of $\lambda(A)$ are in nonincreasing order:

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{k_1} > \lambda_{k_1+1} = \cdots = \lambda_{k_2} > \cdots = \lambda_{k_r},$$

i.e. A has r blocks of repeated eigenvalues, indexed by $\mathcal{I}_\ell = \{k_{\ell-1} + 1, \ldots, k_\ell\}$, $\ell = 1, \ldots, r$.

(Non) smoothness of $\lambda(A(t))$

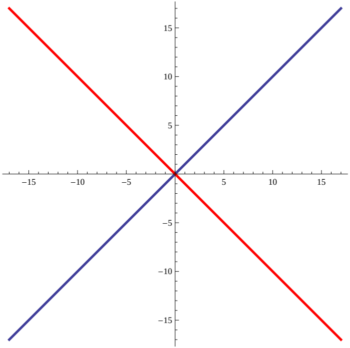
- Consider the symmetric map $A(t) = \begin{pmatrix} 0 & t \\ -t & 0 \end{pmatrix}$.
- $A(t)$ has eigenvalues $\lambda_1(A(t)) = |t|, \lambda_2(A(t)) = -|t|$.



- The map $t \mapsto \lambda(A(t))$ is not differentiable at $t = 0$.

(Non) smoothness of $\lambda(A(t))$

- Consider the symmetric map $A(t) = \begin{pmatrix} 0 & t \\ -t & 0 \end{pmatrix}$.
- $A(t)$ has eigenvalues $\lambda_1(A(t)) = |t|, \lambda_2(A(t)) = -|t|$.



- The map $t \mapsto \lambda(A(t))$ is not differentiable at $t = 0$.

What goes wrong?

- At $t = 0$, the multiplicity of the eigenvalues jump from 1 to 2.
- Rellich 1953: if $A(t)$ is differentiable, then there are differentiable functions $\{\mu_1(t), \ldots, \mu_n(t)\}$ containing all eigenvalues of $A(t)$.
- Kato 1976: if $A(t)$ is analytic, then the $\mu_i(t)$ are analytic.
- In a neighbourhood of t where $A(t)$ has repeated eigenvalues, the roles of the $\mu_i(t)$ as eigenvalues of $A(t)$ are not well-defined.

Distinct Eigenvalues

- If $\lambda_m(A(t^*)) = \lambda_m(A)$ is simple then $\lambda_m(A(t))$ is smooth in a neighbourhood of t^* if $A(t)$ is smooth:

$$\lambda_m(A + tE) = \lambda_m(A) + \sum_{k=1}^{\infty} \lambda_m^{(k)}(A, E) t^k.$$

- Kato 1976 / Ames & Sendov 2011: if $\lambda_m(A)$ is simple then

$$\lambda_m^{(k+1)}(A, E) = -\frac{1}{k+1} \sum_{\substack{v_1 + \cdots + v_{k+1} = k \\ v_1, \ldots, v_{k+1} \geq 0}} Tr(\tilde{E}\tilde{A}^{v_1} \cdots \tilde{E}\tilde{A}^{v_{k+1}})$$

where $\tilde{E} = U^T E U$ and

$$\tilde{A}^k = \begin{cases} -\pi(A) & \text{if } k = 0, \\ ((\lambda_m(A) I - A)^\dagger)^k & \text{if } k \geq 1. \end{cases}$$

How to proceed?

- At best, we can hope for one-sided expansions of $\lambda_m(A(t))$:

$$\lambda_m(A(t)) = a_0 + a_1 t + a_2 t^2 + \cdots$$

for all $t \in [0, \epsilon)$.

- To get this expansion, will have to use generalized derivatives.

Directional derivatives

- A function f is differentiable at x in direction d if the limit

$$\nabla f(x, d) := \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

  exists.

- If $\nabla f(x, d)$ exists for all x, d say that f is directionally differentiable.

- Positively homogeneous in d: $\nabla f(x, \alpha d) = \alpha \nabla f(x, d)$.

- Convex functions are directionally differentiable on the interior of their domains.

“Convexity” of λ_m

- $\lambda_1(A)$ is a convex function of A.
- $\lambda_n(A)$ is a concave function of A.
- Why? by Rayleigh's formula

$$\lambda_1(A) = \max_{\|x\|=1} \langle Ax, x\rangle = \max_{\|x\|=1} Tr(Axx^T)$$

So λ_1 is the support function of $conv\{xx^T : \|x\| = 1\}$

- Support of a convex set is a convex function.
- So λ_1 is convex.

“Convexity” of λ_m

- Ky Fan: can extend Rayleigh’s formula to sum of m largest eigenvalues:

$$\sigma_m(A) = \sum_{i=1}^{m} \lambda_i(A) = \max\{Tr(AXX^T) : X \in \mathbf{R}^{n\times m}, X^TX = I_m\}$$

- σ_m is the support of $conv\{XX^T : X \in \mathbf{R}^{n\times m}, X^TX = I_m\}$
- So σ_m is convex and positively homogeneous.
- So σ_m is directionally differentiable!

Directional derivatives of λ_m

- Hiriart-Urruty & Ye 1993: λ_m is directionally differentiable since $\nabla\lambda_m(A,E) = \nabla\sigma_m(A,E) - \nabla\sigma_{m-1}(A,E)$ for all $A, E \in \Sigma^n$.
- Suppose that $A = U(Diag\,\lambda(A))U^T$ has r blocks of repeated eigenvalues indexed by $\mathcal{I}_1, \ldots, \mathcal{I}_r$. Suppose that $m \in \mathcal{I}_\ell$. Then

$$\nabla\lambda_m(A,E) = \lambda_{m_1}(U_\ell^T E U_\ell),$$

where $m_1 = m - k_{\ell-1}$, U_ℓ is the matrix of columns of U indexed by $\mathcal{I}_\ell$.

- Have first order expansion

$$\lambda_m(A + tE) = \lambda_m(A) + t\lambda_{m_1}(U_\ell^T E U_\ell) + o(t).$$

Second order expansion of λ and beyond

- Torki 2001: can write $U_\ell^T E U_\ell + o(1)$ as

$$\underbrace{U_\ell^T E U_\ell}_{A_1} + tE_1 + o(t)$$

for all sufficiently small $t > 0$.

- Can take directional derivative of $\lambda_{m_1}(A_1 + tE_1)$ to get the second order directional derivative

$$\begin{aligned}\nabla^2 \lambda_m(A, E) &= \lim_{t \downarrow 0} \frac{\lambda_m(A + tE) - \lambda_m(A) - t\nabla\lambda_m(A, E)}{t^2} \\ &= \lambda_{m_2}\Big((U_1)_{\ell_1}^T \Big(\underbrace{U_\ell^T E(\lambda_m(A)I - A)^\dagger E U_\ell}_{=E_1}\Big)(U_1)_{\ell_1}\Big).\end{aligned}$$

Second order expansion of λ and beyond

- Ames & Sendov 2010: can continue this process to get an inductive formula for each coefficient of the (directional) series expansion:

$$\lambda_m(A+tE) = \lambda_m(A) + \sum_{k=1}^{\infty} t^k \nabla^k \lambda_m(A,E)$$

$$= \lambda_m(A) + t\lambda_{m_1}(A_1) + t^2\lambda_{m_2}(A_2) + \cdots + t^K \lambda_{m_K}(A_K) + \cdots$$

Outline

Eigenvalue functions

- We are interested in maps from Σ^n that depend only on the eigenvalues/eigenvectors of the input matrix.

- Two main examples:

  (1) spectral functions

  (2) primary matrix functions.

Spectral functions

- A function $f : \mathbf{R}^n \to \mathbf{R}$ is *symmetric* if $f(x) = f(Px)$ for all x and permutation matrix P.

- A function $F : \Sigma^n \to \mathbf{R}$ is a spectral function if there exists symmetric function g such that

$$F(A) = g(\lambda(A)) = g \circ \lambda(A).$$

- Examples:

$$det(A) = \prod_i \lambda_i(A), \quad \|A\|_2 = \max_i |\lambda_i(A)|, \quad Tr(A) = \sum \lambda_i(A),$$

$$logdet(A) = log \prod_i \lambda_i(A) = \sum_i log(\lambda_i(A))$$

Differentiability of spectral functions

- Lewis 1994: The spectral function $F = g \circ \lambda$ is differentiable at the matrix A if and only if g is differentiable at the vector $\lambda(A)$.

- In this case,

$$\nabla F(A) = U(Diag\, \nabla g(\lambda(A)))U^T$$

for any orthogonal U such that $A = U(Diag\, \lambda(A))U^T$.

Examples

- For any $A \in \Sigma^n$

  $$\nabla Tr(A) = U(Diag\, \nabla(\lambda_1 + \cdots + \lambda_n))U^T = U(Diag\, e)U^T = I.$$

Examples

- For any $A \in \Sigma^n$

$$\nabla Tr(A) = U(Diag\, \nabla(\lambda_1 + \cdots + \lambda_n))U^T = U(Diag\, e)U^T = I.$$

- For any positive definite matrix $A \in \Sigma^n_{++}$

$$\nabla logdet(A) = U \begin{pmatrix} 1/\lambda_1 & & \\ & \ddots & \\ & & 1/\lambda_n \end{pmatrix} U^T = A^{-1}.$$

Proof idea

- If g is differentiable at $\lambda(A)$, can use the first order expansion of g and λ to get

$$\begin{aligned} g(\lambda(A+E)) &= g\Big(\lambda(A) + \nabla\lambda(A,E) + o(\|E\|)\Big) \\ &= g(\lambda(A)) + \nabla g(\lambda(A))^T \nabla\lambda(A,E) + o(\|E\|) \end{aligned}$$

  for all E close to 0.
- Then can use special structure of $\nabla g(\lambda(A))$ to show that

$$\nabla g(\lambda(A))^T \nabla\lambda(A,E) = Tr(\nabla F(A)E) + o(\|E\|)$$

  for all E close to 0.

Higher order derivatives of spectral functions

- Lewis & Sendov 2000: $F = g \circ \lambda$ is twice differentiable A if and only if g is twice differentiable at $\lambda(A)$.
- Sendov 2007: if A has distinct eigenvalues or g is *separable*, the spectral function $F = g \circ \lambda$ is k times differentiable at A if and only if g is k times differentiable at $\lambda(A)$.

Primary matrix functions

- $F : \Sigma^n \to \Sigma^n$ is a primary matrix function if there exists $g : \mathbf{R} \to \mathbf{R}$ such that

$$F(A) = U \begin{pmatrix} g(\lambda_1) & & & \\ & g(\lambda_2) & & \\ & & \ddots & \\ & & & g(\lambda_n) \end{pmatrix} U^T$$

- Examples:

$$\sqrt{A} = U\Big(Diag\Big(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\Big)\Big) U^T,$$

$$exp(A) = U\Big(Diag\Big(e^{\lambda_1}, \dots, e^{\lambda_n}\Big)\Big) U^T$$

Differentiability of PMFs

- Daleckii & Krein 1951: F is differentiable at A if g is differentiable at $\lambda_i(A)$ for all $i = 1, \ldots, n$.
- In this case,

$$\nabla F(A, E) = U\Big(D(\lambda) \circ (U^T E U)\Big) U^T$$

where $\circ$ is the matrix Hadamard product, and $D(\lambda)$ is the matrix of divided differences

$$[D(\lambda)]_{ij} = \begin{cases} \frac{g(\lambda_i) - g(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ \nabla g(\lambda_i) & \text{if } \lambda_i = \lambda_j. \end{cases}$$

Compound matrix

- The k-th compound matrix of $A \in \mathbf{R}^{m \times n}$ is the matrix $A^{(k)} \in \mathbf{R}^{\binom{m}{k} \times \binom{n}{k}}$ defined by

$$[A^{(k)}]_{\rho,\tau} = det(A\rho\tau)$$

for all $\rho \in \mathbf{N}_{m,k}, \tau \in \mathbf{N}_{n,k}$.

- Sometimes called k-th Grassman power or k-th exterior power.

Compound matrix facts

- $(AB)^{(k)} = A^{(k)}B^{(k)}$
- $A^{(k)} = (U(Diag\,\lambda)U^T)^{(k)} = U^{(k)}(Diag\,\lambda)^{(k)}(U^{(k)})^T$
- $\lambda_\rho(A^{(k)}) = \prod_{i\in\rho} \lambda_i(A)$
- If A is nonsingular / positive semidefinite / orthogonal / etc, then so is $A^{(k)}$.

Compound matrix functions

- $F : \Sigma^n \to \Sigma^{\binom{n}{k}}$ is a compound matrix function if there exists symmetric function $g : \mathbf{R}^k \to \mathbf{R}$ such that

$$F(A) = U^{(k)}(Diag\,(g([\lambda(A)_\rho]))(U^{(k)})^T$$

  for all $A \in \Sigma^n$.

Special cases

- $k = n$: have spectral function

$$F(A) = g(\lambda(A))$$

- $k = 1$: have primary matrix function

$$F(A) = U(Diag\, g(\lambda_i(A)))U^T.$$

Open Problem

- Know that F is differentiable at A if g is differentiable at $\lambda(A)$ for $k = 1$ and $k = n$.
- Is this true for all choices of k?