Differentiability Properties of the Eigenvalues of Real Symmetric Matrices

(and Related Topics)

Hristo S. Sendov

12 September, 2002

Contents

(i) Notation & Examples

- What are spectral functions and why we are interested in them
- Why statisticians are interested in spectral functions

(ii) When can we differentiate eigenvalues

(iii) Just a few words on singular values

(iv) How hyperbolic polynomials generalize eigenvalues & singular values and make things more complicated

Eigenvalues

S^n - $n \times n$ symmetric matrices

$\langle A, B \rangle :=$ tr AB

det $(A - \lambda I_n) = 0 \rightarrow$ *eigenvalues*

$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$

$\lambda(A) = (\lambda_1(A), ..., \lambda_n(A))$

$A = U^T \Big(\text{Diag}\, \lambda(A)\Big) U$, $\quad U$ - orthogonal

$\lambda_n(A) > 0$, A – positive definite, $A \succ 0$

$\lambda_n(A) \geq 0$, A – positive semidefinite, $A \succeq 0$

Spectral Functions

Spectral Functions:

$$\boxed{\begin{array}{c} F : S^n \to \bar{\mathbb{R}} \ \text{ s.t.} \\ F(U^T X U) = F(X) \\ \forall X \in S^n \ \& \ U - \text{ orthogonal} \end{array}}$$

$$\Updownarrow$$

Symmetric Functions:

$$\boxed{\begin{array}{c} F = f \circ \lambda, \ \text{ where } f : \mathbb{R}^n \mapsto \bar{\mathbb{R}} \ \text{ s.t.} \\ f(Px) = f(x) \\ P - \text{ permutation matrix} \end{array}}$$

Davis (1957) proved:

$$\boxed{\begin{array}{c} f(x) - \text{ convex \& symmetric on } \mathbb{R}^n \\ \Updownarrow \\ (f \circ \lambda)(X) - \text{ convex on } S^n \end{array}}$$

Miscellaneous Examples

- Minkowski Determinant Theorem

$$X \in S^n \mapsto -\sqrt[n]{\det X}, \ \ X \succeq 0, \ \text{sublinear}$$

$$x \in \mathbb{R}^n \mapsto f(x) = -\sqrt[n]{x_1 x_2 \cdots x_n}$$

- spectral analysis

$$X \in S^n \mapsto F(X) = \lambda_1(X)$$

$$x \in \mathbb{R}^n \mapsto f(x) = \max\{x_1, ..., x_n\}$$

- optimal experimental design in statistics

$$X \in S^n \mapsto F(X) = \begin{cases} \operatorname{tr} X^{-1}, & \text{if } X \succ 0 \\ +\infty, & \text{o/w} \end{cases}$$

$$x \in \mathbb{R}^n \mapsto f(x) = \begin{cases} \frac{1}{x_1} + \cdots + \frac{1}{x_n}, & \text{if } x_i > 0 \\ +\infty, & \text{o/w} \end{cases}$$

More Misc. Examples

- interior point methods in optimization
  maximum likelihood estimates

$$X \in S^n \mapsto F(X) = -\log\det(X)$$

$$x \in \mathbb{R}^n \mapsto f(x) = -\sum_{i=1}^{n} \log(x_i)$$

- update for Quasi-Newton algorithms

$$X \in S^n \mapsto F(X) = \begin{cases} \frac{\operatorname{tr}(X)}{n\det(X)}, & \text{if } X \succ 0 \\ +\infty, & \text{o/w} \end{cases}$$

$$x \in \mathbb{R}^n \mapsto f(x) = \begin{cases} \frac{\sum_{i=1}^n x_i}{n\prod_{i=1}^n x_i}, & \text{if } x_i > 0 \\ +\infty, & \text{o/w} \end{cases}$$

Example: Wishart Distribution

$Z \quad n \times m$ random matrix, $n \geq m$

columns of Z independent $\sim N(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} \in S^n$

Let $A := Z^T Z$, then $A \sim W_m(n, \mathbf{\Sigma})$

The density of A is

$$\frac{C}{(\det \mathbf{\Sigma})^{n/2}} e^{\operatorname{tr}(-\frac{1}{2}\mathbf{\Sigma}^{-1} A)} (\det A)^{(n-m-1)/2}$$

$\mathbf{\Sigma} = I_n \Rightarrow$ density is a spectral function

Multivariate Beta Distribution

Ω – probability space

$X : \Omega \rightarrow S^n$

$X(\Omega) = \{A \in S^n | A \succ 0, \, I_n - A \succ 0\}$

X is multivariate Beta(p, q) if p.d.f. is

$$C_1 \cdot (\det A)^{p-(n+1)/2} \det(I_n - A)^{q-(n+1)/2}$$

Then the joint p.d.f of the eigenvalues of A is

$$C_2 \cdot \prod \left(\lambda_i^{p-(n+1)/2} (1 - \lambda_i)^{q-(n+1)/2} \right) \prod_{i<j}^{n} (\lambda_i - \lambda_j)$$

Notice both p.d.f. spectral

Eigenvalues of a Random Matrix

Ω – probability space

$X : \Omega \to S^n$

$X(\Omega) \subseteq \{A \in S^n | A \succeq 0\}$

$f(A)$ – density function of X

Then the joint density of the eigenvalues is

$$C \cdot \prod_{i<j}^{n} (\lambda_i(A) - \lambda_j(A)) \int_{O(n)} f(U^T \Lambda U)(dU)$$

$O(n)$ – orthogonal $n \times n$ matrices

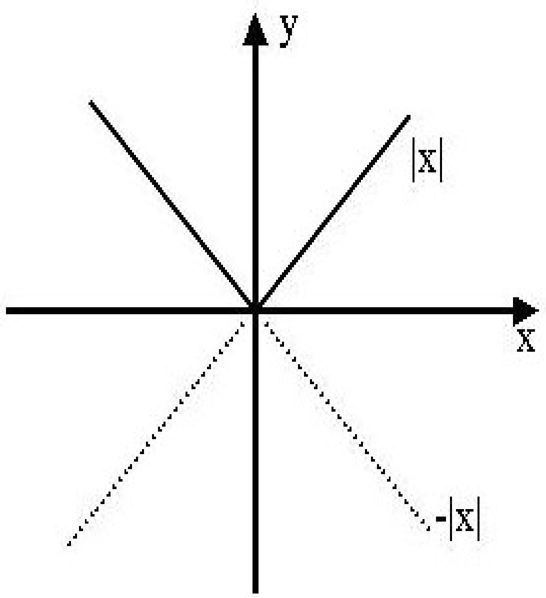
$\Lambda = \mathrm{Diag}\,(\lambda_1(A), ..., \lambda_n(A))$

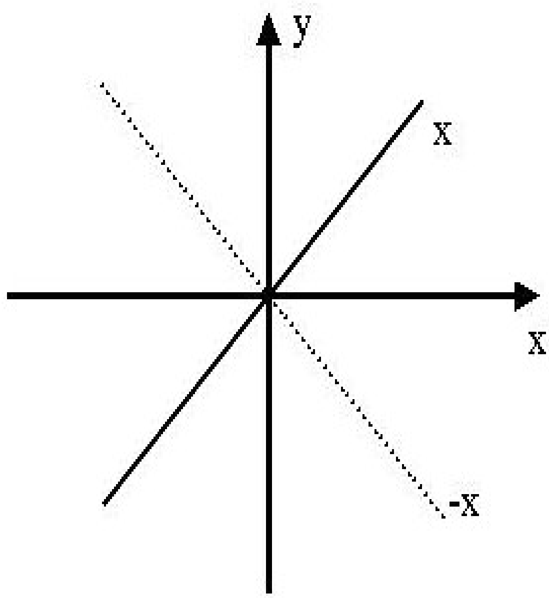
Why Differentiability?

- To approximate a complicated object with a simpler one
- Sensitivity of a function when the argument is perturbed
- Newton's method relies on first and second order information
- Second order optimality conditions
- Interior point methods in optimization

First Example on Differentiability

Consider $A(x) = \begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix} \in S^2, \quad x \in \mathbb{R}$





- $\{\lambda_1(A(x)), \lambda_2(A(x))\} = \{|x|, -|x|\}$
  Not differentiable at $x = 0$

- But $\{\mu_1(A(x)), \mu_2(A(x))\} = \{x, -x\}$
  Differentiable everywhere

Theorem (Rellich, 1953)

$$\boxed{\begin{array}{l} A(x) = (a_{ij}(x)) \in S^n, \quad a_{ij}(x) \in C^1, \quad x \in I \subset \mathbb{R} \\ \\ \text{Then } \exists\ \mu_i(x) \in C^1,\ \forall i = 1, ..., n \text{ and } x \in I \\ \\ \{\mu_1(x), ..., \mu_n(x)\} \text{ are all eigenvalues of } A(x)\ \forall x \end{array}}$$

- But $A(x) \in C^\infty \not\Rightarrow \exists \mu_i(x) \in C^2$
- Twist $A(x)$ - analytic $\Rightarrow \exists \mu_i(x)$ - analytic

Source: Kato, Perturbation Theory for Linear Operators

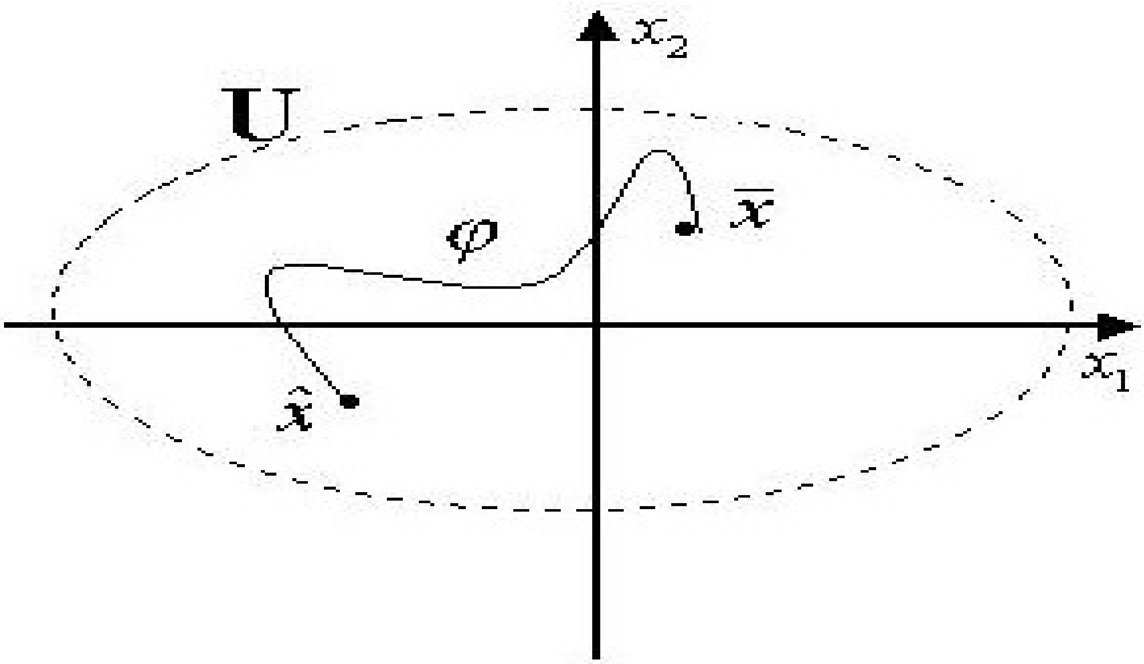
Rellich does not hold in $\mathbb{R}^2$

Let $A(x) = \begin{pmatrix} x_1 & x_2 \\ x_2 & -x_1 \end{pmatrix} \in S^2, \qquad x = (x_1, x_2)$

$\{\|x\|, -\|x\|\}$ – the eigenvalues

Suppose $\exists\ \{\mu_1(x), \mu_2(x)\}$ diff. in $U \ni 0$

Fix $\bar{x} \in U$ w.l.o.g. $\mu_1(\bar{x}) = \|\bar{x}\|$, $\mu_2(\bar{x}) = -\|\bar{x}\|$



$\Rightarrow \mu_1(\hat{x}) = \|\hat{x}\|$, $\mu_2(\hat{x}) = -\|\hat{x}\|$, $\forall \hat{x}$

Differentiability in S^n

$F : S^n \to \mathbb{R}$ differentiable at $A \in S^n$ if

$\exists\ \nabla F(A)[\cdot] : S^n \to \mathbb{R}$ - linear

$$\lim_{H \to 0} \frac{|F(A+H) - F(A) - \nabla F(A)[H]|}{\|H\|} = 0$$

Think of $\nabla F(A)$ as an element of S^n

$\nabla F(A)[H] := \langle \nabla F(A), H \rangle := \mathrm{tr}\,(\nabla F(A) H)$

Differentiability of Spectral Functions

(Lewis, 1995)

$F(A) = (f \circ \lambda)(A)$, $\quad f$ - symmetric

$$\boxed{\begin{array}{c} F = f \circ \lambda \text{ differentiable at } A \\ \Updownarrow \\ f \text{ differentiable at } \lambda(A) \end{array}}$$

$$\nabla (f \circ \lambda)(A) = U^T \Big(\text{Diag}\big(\nabla f(\lambda(A)) \big) \Big) U$$

where $A = U^T \big(\text{Diag}\, \lambda(A) \big) U$, $\quad U$-orthogonal

- Clearly $F = f \circ \lambda \in C^1 \Leftrightarrow f \in C^1$

- $\lambda_i(A)$ diff. $\Leftrightarrow \lambda_{i-1}(A) > \lambda_i(A) > \lambda_{i+1}(A)$

Twice Diff. of Spectral Functions

$F : S^n \to \mathbb{R}$ twice differentiable at $A \in S^n$ if

$\nabla F(\cdot) : S^n \to S^n$ is differentiable at A, i.e.

$\exists\ \nabla^2 F(A)[\cdot,\cdot] : S^n \times S^n \to \mathbb{R}$ – symm., bilinear

$$\lim_{H \to 0} \frac{\|\nabla F(A+H) - \nabla F(A) - \nabla^2 F(A)[H,\cdot]\|}{\|H\|} = 0$$

(Sendov, 2000) When F spectral

$$\boxed{\begin{array}{c} F = f \circ \lambda \text{ twice differentiable at } A \\ \Updownarrow \\ f \text{ twice differentiable at } \lambda(A) \end{array}}$$

Hessian of Spectral Functions

(Sendov 2000)

$$\boxed{\begin{array}{c}\nabla^2(f\circ\lambda)(A)[H,H]\\ \| \\ \nabla^2 f(\lambda(A))[\mathrm{diag}\,\tilde{H},\mathrm{diag}\,\tilde{H}]+\langle\mathcal{A},\tilde{H}\circ\tilde{H}\rangle\end{array}}$$

$\tilde{H}=UHU^T$, $\quad A=U^T\Big(\mathrm{Diag}\,\lambda(A)\Big)U$, $\quad U$-orth.

$\lambda_1=\cdots=\lambda_{k_1}>\lambda_{k_1+1}=\cdots=\lambda_{k_2}>\lambda_{k_2+1}\cdots\lambda_n$

$I_1=\{1,...,k_1\},\ I_2=\{k_1+1,...,k_2\},\ldots$

$$\mathcal{A}_{ij}=\begin{cases}0, & \text{if }\ i=j\\ b_p, & \text{if }\ i\neq j \text{ but } i,j\in I_p\\ c_{pq}, & \text{if } i\in I_p,\ j\in I_q,\ p\neq q\end{cases}$$

$\mathcal{A}$ – symmetric

Additional Properties (Sendov 2000)

- $$\nabla^2(f \circ \lambda)(A)[\cdot,\cdot] \succeq 0 \;\Leftrightarrow\; f \circ \lambda \text{ convex}$$
  $$\Leftrightarrow\; f \text{ convex } \;\Leftrightarrow \nabla^2 f(\lambda(A)) \succeq 0$$

$$\boxed{\begin{array}{c} \nabla^2(f \circ \lambda)(A)[\cdot,\cdot] \succ 0 \\ \Updownarrow \\ \nabla^2 f(\lambda(A)) \succ 0 \end{array}}$$

$$\boxed{\begin{array}{c} F = f \circ \lambda \in C^2 \text{ at } A \\ \Updownarrow \\ f \in C^2 \text{ at } \lambda(A) \end{array}}$$

- As $A_n \to A$, $\lambda(A_n)$ changes block structure

Quadratic Expansions (q.e.)

(Sendov 2000)

- $F = f \circ \lambda$ has q.e. at $A \Leftrightarrow f$ has at $\lambda(A)$

f has q.e. at x if

$$\boxed{\begin{aligned} &f(x+h) = \\ &\quad f(x) + \langle \nabla f(x), h\rangle + \tfrac{1}{2}\langle h, \nabla^2 f(x) h\rangle + O(\|h\|^3) \end{aligned}}$$

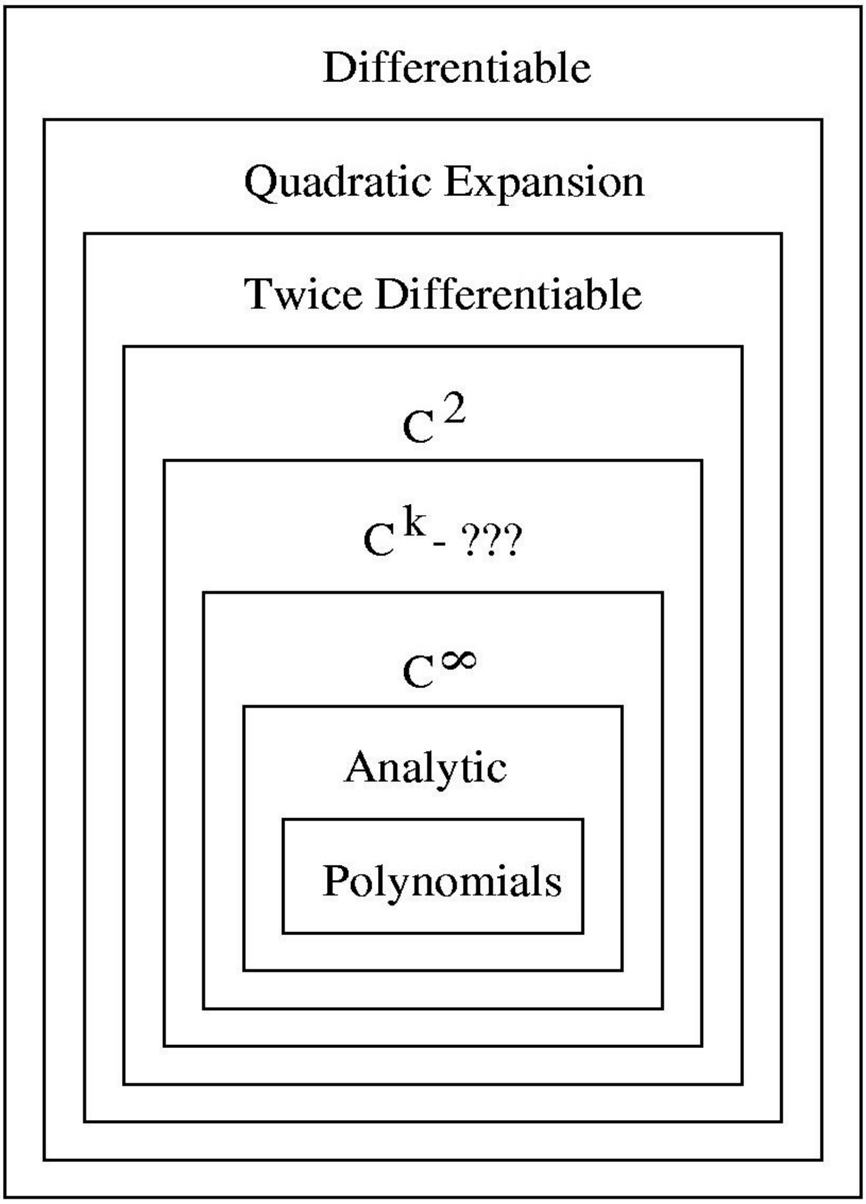
- q.e. $\Rightarrow$ differentiability

- twice diff. $\Rightarrow$ q.e. $\not\Rightarrow$ twice diff.

<u>E.g.</u> $$f(x) = \begin{cases} x^3 \sin(1/x), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Alexandrov (1939): f - convex $\Rightarrow$ q.e. a.e.

A Chain of Preserved Properties



Singular Values

$M_{n,m}$ - $n \times m$ rectangular matrices

Every $A \in M_{n,m}$ decomposes as

$$A = V \Sigma W$$

V, W - orthogonal

$\Sigma =$Diag(σ)

$\sigma_1 \geq ... \geq \sigma_n \geq 0$ - unique

singular values of A

Singular Value Functions

$$\boxed{\begin{array}{c} F : M_{n,m} \to \bar{\mathbb{R}} \ \text{ s.t.} \\ F(VXW) = F(X) \\ \forall X \in M_{n,m} \ \&\ V, W - \text{ orthogonal} \end{array}}$$

$$\Updownarrow$$

$$\boxed{\begin{array}{c} F = g \circ \sigma, \ \text{ where } g : \mathbb{R}^n \to \bar{\mathbb{R}} \ \text{ s.t.} \\ g(Px) = g(x) \\ P - \text{ signed permutation matrix} \end{array}}$$

Such g - *absolutely symmetric*

Such F - *singular value function*

Von Neumann's Result

$\|\cdot\| : M_{n,m} \to \mathbb{R}_+$ - orthog. invariant if

$$\|VXW\| = \|X\|, \ \forall V, W \text{ — orthogonal}$$

$\|\cdot\|$ — singular value function

Von Neumann (1937) proved:

$$\boxed{\begin{array}{c} (g \circ \sigma)(X) - \text{ norm on } M_{n,m} \\ \Updownarrow \\ g(x) - \text{ norm on } \mathbb{R}^n \end{array}}$$

E.g. The spectral norm:

$g(x) = \max\{|x_1|, ..., |x_n|\}$

$(g \circ \sigma)(X) = \sigma_1(X)$

$$\left(= |||X|||_2 = \sqrt{\lambda_1(X^T X)} \right)$$

Hyperbolic Polynomials

(A unifying framework, cf. Güler '97)

$p : \mathbb{R}^m \to \mathbb{R}$ - multivariate polynomial

$p(x)$ *homogeneous, degree* n, if

$$p(tx) = t^n p(x), \ \forall t \in \mathbb{R} \ x \in \mathbb{R}^m$$

Hyperbolic Polynomials :

$p(x)$ is homogeneous of deg. n

$\exists d$ with $p(d) \neq 0$ s.t.

$t \mapsto p(x - td)$ has all *real* roots $\forall x$

Roots of Hyperbolic Polynomials

$p(x)$ - hyperbolic polynomial

$\lambda(x) := (\lambda_1(x), \ldots, \lambda_n(x))$, its roots

E.g.

- $X \in S^n$

$$p(X) = \det X, \quad d = I,$$
$$\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))$$

- $X \in M_{n,m}$

$$p(X, \alpha) = \det\left(\alpha^2 I_n - XX^T\right), \quad d = (\mathbf{0}, 1)$$
$$\lambda(X, \alpha) = \Big(\alpha + \sigma_1(X), \alpha + \sigma_2(X), \ldots$$
$$\ldots, \alpha - \sigma_2(X), \alpha - \sigma_1(X)\Big)$$

Cauchy Problem

$p : \mathbb{R}^m \to \mathbb{R}$ homogeneous of degree n

$x_k \to -\frac{i\partial}{\partial x_k}$, $\quad p \to p(D)$, $\quad$ P.D.Operator

$$x_1^2 - \sum_{k=2}^{n} x_k^2 \longrightarrow -\frac{\partial^2}{\partial x_1^2} + \sum_{k=2}^{n} \frac{\partial^2}{\partial x_k^2}$$

Problem: Given p, $H = \{x | \langle x, d\rangle \geq 0\}$
$p(d) \neq 0$, is there a solution to

$$p(D)u = f$$

with $\operatorname{supp}(u) \subseteq H$, for any $f \in C^\infty(H)$?

Answer: $\Leftrightarrow$ $p(x)$ hyperbolic w.r.t. d

Results on Hyperbolic Polynomials

(Bauschke, Güler, Lewis, Sendov, 1998)

$p(x)$ - hyperbolic polynomial on $\mathbb{R}^m$, degree n

$\lambda(x)$ - its roots vector $\in \mathbb{R}^n$

f - symmetric on $\mathbb{R}^n$

With small restrictions:

$$\boxed{\begin{array}{c} f - \text{ convex on } \mathbb{R}^n \\ \Updownarrow \\ (f \circ \lambda)(x) - \text{ convex on } \mathbb{R}^m \end{array}}$$

$$\boxed{\begin{array}{c} f \text{ differentiable at } \lambda(x) \\ \Updownarrow \\ f \circ \lambda \text{ differentiable at } x \end{array}}$$

Minkowski Det Theorem Revisited

p - hyperbolic on $\mathbb{R}^m$, with roots $\lambda_i(x)$

$C := \{x \in \mathbb{R}^m \,|\, \lambda_i(x) \geq 0 \, \forall i\}$ - convex cone

(Gårding, 1959)

$$x \mapsto -\sqrt[n]{p(x)} \text{ sublinear on } C$$

Example (Minkowski Det. Theorem)

$p(X) = \det(X)$, then $C = \{X \in S^n \,|\, X \succeq 0\}$

$$\sqrt[n]{\det(X+Y)} \geq \sqrt[n]{\det X} + \sqrt[n]{\det Y}$$

Conclusion

- Spectral functions inherit differentiability, convexity, and many other properties from their corresponding symmetric functions

- Singular value functions also inherit many properties from their corresponding absolutely symmetric functions

- Functions of the roots of hyperbolic polynomials in many respects generalize both the spectral functions and the singular value functions.