CONVEXITY AND EIGENVALUES OF SYMMETRIC MATRICES

Adrian Lewis

Cornell University

April 6, 2005

orie.cornell.edu/∼aslewis

Technion: Lecture 1

1. OUTLINE

- Hyperbolic polynomials and the Lax conjecture

1. OUTLINE

- Hyperbolic polynomials and the Lax conjecture
- Convexity and semidefinite programming

1. OUTLINE

- Hyperbolic polynomials and the Lax conjecture
- Convexity and semidefinite programming
- Spectral functions

1. OUTLINE

- Hyperbolic polynomials and the Lax conjecture
- Convexity and semidefinite programming
- Spectral functions
- Von Neumann and unitarily invariant norms

1. OUTLINE

- Hyperbolic polynomials and the Lax conjecture
- Convexity and semidefinite programming
- Spectral functions
- Von Neumann and unitarily invariant norms
- Duality and subgradients

1. OUTLINE

- Hyperbolic polynomials and the Lax conjecture
- Convexity and semidefinite programming
- Spectral functions
- Von Neumann and unitarily invariant norms
- Duality and subgradients
- Some Lie algebra...

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u, v, w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u,v,w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

For all real v, w, $\qquad p(u,v,w) = 0 \quad \Rightarrow \quad u$ **real**.

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u,v,w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

For all real v, w, $\quad p(u,v,w) = 0 \quad \Rightarrow \quad u$ **real**.

We say p is **hyperbolic relative to** $d = (1,0,0)$:
$t \mapsto p(x - td)$ always has all real roots.

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u,v,w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

For all real v, w, $\quad p(u,v,w) = 0 \quad \Rightarrow \quad u$ **real**.

We say p is **hyperbolic relative to** $d = (1,0,0)$:
$t \mapsto p(x - td)$ always has all real roots.
Why?

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u,v,w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

For all real v, w, $$p(u,v,w) = 0 \quad \Rightarrow \quad u \textbf{ real}.$$

We say p is **hyperbolic relative to** $d = (1,0,0)$:
$t \mapsto p(x - td)$ always has all real roots.
Why?

$$p(u,v,w) = \det \begin{bmatrix} u & v & w \\ v & u & v \\ w & v & u \end{bmatrix}.$$

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u,v,w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

For all real v, w, $\quad p(u,v,w) = 0 \quad \Rightarrow \quad u$ **real**.

We say p is **hyperbolic relative to** $d = (1,0,0)$:
$t \mapsto p(x - td)$ always has all real roots.
Why?

$$p(u,v,w) = \det \begin{bmatrix} u & v & w \\ v & u & v \\ w & v & u \end{bmatrix}.$$

Lax Conjecture (1958) Hyperbolic polynomials on $\mathbf{R}^3$ relative to $(1,0,0)$ are all of the form

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u,v,w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

For all real v, w, $\quad p(u,v,w) = 0 \quad \Rightarrow \quad u$ **real**.

We say p is **hyperbolic relative to** $d = (1,0,0)$:
$t \mapsto p(x - td)$ always has all real roots.
Why?

$$p(u,v,w) = \det \begin{bmatrix} u & v & w \\ v & u & v \\ w & v & u \end{bmatrix}.$$

Lax Conjecture (1958) Hyperbolic polynomials on $\mathbf{R}^3$ relative to $(1,0,0)$ are all of the form

$$p(u,v,w) = \det(uI + vA + wB) \quad \text{with } A, B \text{ symmetric.}$$

2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

$$p(u,v,w) = u^3 - 2uv^2 - uw^2 + 2v^2w.$$

For all real v, w, $\quad p(u,v,w) = 0 \quad \Rightarrow \quad u$ **real**.

We say p is **hyperbolic relative to** $d = (1,0,0)$:
$t \mapsto p(x - td)$ always has all real roots.
Why?

$$p(u,v,w) = \det \begin{bmatrix} u & v & w \\ v & u & v \\ w & v & u \end{bmatrix}.$$

Lax Conjecture (1958) Hyperbolic polynomials on $\mathbf{R}^3$ relative to $(1,0,0)$ are all of the form

$$p(u,v,w) = \det(uI + vA + wB) \quad \text{with } A, B \text{ symmetric.}$$

True: Helton/Vinnikov 2002, Lewis/Parrilo/Ramana 2004.

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;
- are common (there are open sets of such polynomials);

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;
- are common (there are open sets of such polynomials);
- have surprising convexity properties.

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;
- are common (there are open sets of such polynomials);
- have surprising convexity properties.

Theorem (Gårding 1951) If p is hyperbolic relative to d, then the component of d in $\{x : p(x) > 0\}$ is **convex**.

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;
- are common (there are open sets of such polynomials);
- have surprising convexity properties.

Theorem (Gårding 1951) If p is hyperbolic relative to d, then the component of d in $\{x : p(x) > 0\}$ is **convex**.

We call this component, H, the **hyperbolicity cone**.

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;
- are common (there are open sets of such polynomials);
- have surprising convexity properties.

Theorem (Gårding 1951) If p is hyperbolic relative to d, then the component of d in $\{x : p(x) > 0\}$ is **convex**.

We call this component, H, the **hyperbolicity cone**.

Furthermore (Güler 1997), $-\log p$ is **self-concordant** on H.

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;
- are common (there are open sets of such polynomials);
- have surprising convexity properties.

Theorem (Gårding 1951) If p is hyperbolic relative to d, then the component of d in $\{x : p(x) > 0\}$ is **convex**.

We call this component, H, the **hyperbolicity cone**.

Furthermore (Güler 1997), $-\log p$ is **self-concordant** on H. Hence (Nesterov/Nemirovski, 1994) theoretically efficient interior point methods for **hyperbolic programs**

$$\min\{\langle c, x\rangle : Ax = b,\ x \in H\}.$$

3. CONVEXITY

Hyperbolic polynomials

- are simply defined;
- are common (there are open sets of such polynomials);
- have surprising convexity properties.

Theorem (Gårding 1951) If p is hyperbolic relative to d, then the component of d in $\{x : p(x) > 0\}$ is **convex**.

We call this component, H, the **hyperbolicity cone**.

Furthermore (Güler 1997), $-\log p$ is **self-concordant** on H. Hence (Nesterov/Nemirovski, 1994) theoretically efficient interior point methods for **hyperbolic programs**

$$\min\{\langle c, x\rangle : Ax = b,\ x \in H\}.$$

(Damped Newton's method for penalized version

$$\min\{\langle c, x\rangle - \mu \log p(x) : Ax = b\}, \qquad \text{as } \mu \downarrow 0.)$$

4. SEMIDEFINITE PROGRAMMING

Example: The **determinant** is hyperbolic on

$$\mathbf{S}^n = \{n \times n \text{ real symmetric matrices}\}$$

relative to the identity I,

4. SEMIDEFINITE PROGRAMMING

Example: The **determinant** is hyperbolic on

$$\mathbf{S}^n = \{n \times n \text{ real symmetric matrices}\}$$

relative to the identity I, since each $X \in \mathbf{S}^n$ has **all real eigenvalues** $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

4. SEMIDEFINITE PROGRAMMING

Example: The **determinant** is hyperbolic on

$$\mathbf{S}^n = \{n \times n \text{ real symmetric matrices}\}$$

relative to the identity I, since each $X \in \mathbf{S}^n$ has **all real eigenvalues** $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

The hyperbolicity cone is

$$\mathbf{S}^n_{++} = \{\text{positive definites}\}$$

4. SEMIDEFINITE PROGRAMMING

Example: The **determinant** is hyperbolic on

$$\mathbf{S}^n \ = \ \{n \times n \text{ real symmetric matrices}\}$$

relative to the identity I, since each $X \in \mathbf{S}^n$ has **all real eigenvalues** $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

The hyperbolicity cone is

$$\mathbf{S}^n_{++} = \{\text{positive definites}\}$$

and $-\log\det$ is self-concordant on $\mathbf{S}^n_{++}$.

4. SEMIDEFINITE PROGRAMMING

Example: The **determinant** is hyperbolic on

$$\mathbf{S}^n = \{n \times n \text{ real symmetric matrices}\}$$

relative to the identity I, since each $X \in \mathbf{S}^n$ has **all real eigenvalues** $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

The hyperbolicity cone is

$$\mathbf{S}^n_{++} = \{\text{positive definites}\}$$

and $-\log\det$ is self-concordant on $\mathbf{S}^n_{++}$.

Hence **semidefinite programming**:

$$\begin{array}{lrcl} \text{minimize} & \langle C, X \rangle & & \\ \text{subject to} & \langle A_i, X \rangle & = & b_i \quad (i = 1, \ldots, m) \\ & X & \in & \mathbf{S}^n_{++}. \end{array}$$

4. SEMIDEFINITE PROGRAMMING

Example: The **determinant** is hyperbolic on

$$\mathbf{S}^n = \{n \times n \text{ real symmetric matrices}\}$$

relative to the identity I, since each $X \in \mathbf{S}^n$ has **all real eigenvalues** $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

The hyperbolicity cone is

$$\mathbf{S}^n_{++} = \{\text{positive definites}\}$$

and $-\log\det$ is self-concordant on $\mathbf{S}^n_{++}$.

Hence **semidefinite programming**:

$$\begin{array}{lrcl} \text{minimize} & \langle C, X \rangle & & \\ \text{subject to} & \langle A_i, X \rangle & = & b_i \quad (i = 1, \ldots, m) \\ & X & \in & \mathbf{S}^n_{++}. \end{array}$$

A powerful, tractable generalization of linear programming (eg: Ben-Tal/Nemirovski 2001).

5. HYPERBOLIC PROGRAMMING

Since the Lax conjecture is true, all three-dimensional hyperbolicity cones are **semidefinite slices**:

$$\left\{ x : \sum_i x_i A_i \in \mathbf{S}^n_{++} \right\}$$

for given $A_i \in \mathbf{S}^n$.

5. HYPERBOLIC PROGRAMMING

Since the Lax conjecture is true, all three-dimensional hyperbolicity cones are **semidefinite slices**:

$$\left\{ x : \sum_i x_i A_i \in \mathbf{S}^n_{++} \right\}$$

for given $A_i \in \mathbf{S}^n$.

Same is true for all **homogeneous** cones — open convex pointed cones K such that for every $x, y \in K$ there is an automorphism $\Gamma : K \to K$ such that $\Gamma x = y$ (Chua 2003, Faybusovich 2002).

5. HYPERBOLIC PROGRAMMING

Since the Lax conjecture is true, all three-dimensional hyperbolicity cones are **semidefinite slices**:

$$\left\{ x : \sum_i x_i A_i \in \mathbf{S}^n_{++} \right\}$$

for given $A_i \in \mathbf{S}^n$.

Same is true for all **homogeneous** cones — open convex pointed cones K such that for every $x, y \in K$ there is an automorphism $\Gamma : K \to K$ such that $\Gamma x = y$ (Chua 2003, Faybusovich 2002).

So, is hyperbolic programming genuinely more general than semidefinite programming?

5. HYPERBOLIC PROGRAMMING

Since the Lax conjecture is true, all three-dimensional hyperbolicity cones are **semidefinite slices**:

$$\left\{ x : \sum_i x_i A_i \in \mathbf{S}^n_{++} \right\}$$

for given $A_i \in \mathbf{S}^n$.

Same is true for all **homogeneous** cones —
open convex pointed cones K such that for every $x, y \in K$ there is an automorphism $\Gamma : K \to K$ such that $\Gamma x = y$ (Chua 2003, Faybusovich 2002).

So, is hyperbolic programming genuinely more general than semidefinite programming?

Are all hyperbolicity cones projections of semidefinite slices?

6. CONVEXITY AND SYMMETRY

The convexity of $\mathbf{S}^n_{++}$ and $-\log\det$ are special cases of:

6. CONVEXITY AND SYMMETRY

The convexity of $\mathbf{S}^n_{++}$ and $-\log\det$ are special cases of:

Theorem (Davis 1957) If $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and **permutation-invariant**, then the function

$$X \in \mathbf{S}^n \quad \mapsto \quad f(\lambda_1(X), \ldots, \lambda_n(X))$$

is convex.

6. CONVEXITY AND SYMMETRY

The convexity of $\mathbf{S}^n_{++}$ and $-\log\det$ are special cases of:

Theorem (Davis 1957) If $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and **permutation-invariant**, then the function

$$X \in \mathbf{S}^n \quad \mapsto \quad f(\lambda_1(X), \ldots, \lambda_n(X))$$

is convex.

Consider $f(x) =$

$$\begin{cases} 0 & (x > 0) \\ +\infty & (x \not> 0) \end{cases}$$

6. CONVEXITY AND SYMMETRY

The convexity of $\mathbf{S}^n_{++}$ and $-\log\det$ are special cases of:

Theorem (Davis 1957) If $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and **permutation-invariant**, then the function

$$X \in \mathbf{S}^n \quad \mapsto \quad f(\lambda_1(X), \ldots, \lambda_n(X))$$

is convex.

Consider $f(x) =$

$$\begin{cases} 0 & (x > 0) \\ +\infty & (x \not> 0) \end{cases} \qquad \text{or} \qquad \begin{cases} -\sum_i \log x_i & (x > 0) \\ +\infty & (x \not> 0). \end{cases}$$

6. CONVEXITY AND SYMMETRY

The convexity of $\mathbf{S}^n_{++}$ and $-\log\det$ are special cases of:

Theorem (Davis 1957) If $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and **permutation-invariant**, then the function

$$X \in \mathbf{S}^n \quad \mapsto \quad f(\lambda_1(X), \ldots, \lambda_n(X))$$

is convex.

Consider $f(x) =$

$$\begin{cases} 0 & (x > 0) \\ +\infty & (x \not> 0) \end{cases} \qquad \text{or} \qquad \begin{cases} -\sum_i \log x_i & (x > 0) \\ +\infty & (x \not> 0). \end{cases}$$

This result extends to hyperbolic polynomials p (relative to d), interpreting $\{\lambda_i(x)\}$ as the roots of $t \mapsto p(x - td)$ (Bauschke/Güler/Lewis/Sendov 2001).

7. INVARIANCE

A function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ is **spectral** if

$$F(U^T X U) = F(X) \quad \text{whenever} \quad U^T U = I$$

7. INVARIANCE

A function $F : \mathbf{S}^n \rightarrow \overline{\mathbf{R}}$ is **spectral** if

$$F(U^T X U) = F(X) \text{ whenever } U^T U = I$$

(because then

$$F(X) = F(\text{Diag}(\lambda_i(X))) \qquad (*)$$

by the spectral decomposition.)

7. INVARIANCE

A function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ is **spectral** if

$$F(U^T X U) = F(X) \text{ whenever } U^T U = I$$

(because then

$$F(X) = F(\mathrm{Diag}(\lambda_i(X))) \qquad (*)$$

by the spectral decomposition.)

The Davis result **characterizes** convex spectral functions:

7. INVARIANCE

A function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ is **spectral** if

$$F(U^T X U) = F(X) \text{ whenever } U^T U = I$$

(because then

$$F(X) = F(\mathrm{Diag}(\lambda_i(X))) \qquad (*)$$

by the spectral decomposition.)

The Davis result **characterizes** convex spectral functions:

Theorem A spectral function F is convex $\Leftrightarrow$ F is convex on $\mathbf{D}^n = \{n \times n$ real diagonals$\}$.

7. INVARIANCE

A function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ is **spectral** if

$$F(U^T X U) = F(X) \;\text{ whenever }\; U^T U = I$$

(because then

$$F(X) = F(\text{Diag}(\lambda_i(X))) \qquad (*)$$

by the spectral decomposition.)

The Davis result **characterizes** convex spectral functions:

Theorem A spectral function F is convex $\Leftrightarrow$ F is convex on $\mathbf{D}^n = \{n \times n \text{ real diagonals}\}$.

Proof "$\Rightarrow$" is immediate.

7. INVARIANCE

A function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ is **spectral** if

$$F(U^T X U) = F(X) \;\text{ whenever }\; U^T U = I$$

(because then

$$F(X) = F(\text{Diag}(\lambda_i(X))) \qquad (*)$$

by the spectral decomposition.)

The Davis result **characterizes** convex spectral functions:

Theorem A spectral function F is convex $\Leftrightarrow$ F is convex on

$$\mathbf{D}^n = \{n \times n \text{ real diagonals}\}.$$

Proof "$\Rightarrow$" is immediate. To see "$\Leftarrow$",
use $(*)$ and note $F \circ \text{Diag}$ is permutation-invariant. $\square$

7. INVARIANCE

A function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ is **spectral** if

$$F(U^T X U) = F(X) \ \text{ whenever } \ U^T U = I$$

(because then

$$F(X) = F(\text{Diag}(\lambda_i(X))) \qquad\qquad (*)$$

by the spectral decomposition.)

The Davis result **characterizes** convex spectral functions:

Theorem A spectral function F is convex $\Leftrightarrow$ F is convex on $\mathbf{D}^n = \{n \times n \text{ real diagonals}\}$.

Proof "$\Rightarrow$" is immediate. To see "$\Leftarrow$", use $(*)$ and note $F \circ \text{Diag}$ is permutation-invariant. $\square$

Reminiscent of a famous result of von Neumann...

8. UNITARILY INVARIANT NORMS

A function G on

$$\mathbf{M}^n = \{n \times n \text{ complex matrices}\}$$

is **unitarily invariant**

8. UNITARILY INVARIANT NORMS

A function G on

$$\mathbf{M}^n = \{n \times n \text{ complex matrices}\}$$

is **unitarily invariant** if

$$G(UXV) = G(X) \quad \text{whenever} \quad U^*U = I = V^*V.$$

8. UNITARILY INVARIANT NORMS

A function G on

$$\mathbf{M}^n \;=\; \{n \times n \text{ complex matrices}\}$$

is **unitarily invariant** if

$$G(UXV) = G(X) \;\text{ whenever }\; U^*U = I = V^*V.$$

Theorem (von Neumann 1937) A unitarily invariant function G is a norm $\quad\Leftrightarrow\quad$ the restriction $G|_{\mathbf{D}^n}$ is a norm.

8. UNITARILY INVARIANT NORMS

A function G on

$$\mathbf{M}^n = \{n \times n \text{ complex matrices}\}$$

is **unitarily invariant** if

$$G(UXV) = G(X) \text{ whenever } U^*U = I = V^*V.$$

Theorem (von Neumann 1937) A unitarily invariant function G is a norm $\Leftrightarrow$ the restriction $G|_{\mathbf{D}^n}$ is a norm.

Example: Schatten p-norms $(1 \le p \le \infty)$

$$\|X\| = \left(\sum_j \sigma_j(X)^p\right)^{1/p}$$

8. UNITARILY INVARIANT NORMS

A function G on

$$\mathbf{M}^n = \{n \times n \text{ complex matrices}\}$$

is **unitarily invariant** if

$$G(UXV) = G(X) \text{ whenever } U^*U = I = V^*V.$$

Theorem (von Neumann 1937) A unitarily invariant function G is a norm $\Leftrightarrow$ the restriction $G|_{\mathbf{D}^n}$ is a norm.

Example: **Schatten p-norms** $(1 \le p \le \infty)$

$$\|X\| = \left(\sum_j \sigma_j(X)^p\right)^{1/p}$$

(where **singular value** $\sigma_j(X) = \sqrt{\lambda_j(X^*X)}$).

8. UNITARILY INVARIANT NORMS

A function G on

$$\mathbf{M}^n = \{n \times n \text{ complex matrices}\}$$

is **unitarily invariant** if

$$G(UXV) = G(X) \text{ whenever } U^*U = I = V^*V.$$

Theorem (von Neumann 1937) A unitarily invariant function G is a norm $\Leftrightarrow$ the restriction $G|_{\mathbf{D}^n}$ is a norm.

Example: **Schatten p-norms** $(1 \le p \le \infty)$

$$\|X\| = \left(\sum_j \sigma_j(X)^p \right)^{1/p}$$

(where **singular value** $\sigma_j(X) = \sqrt{\lambda_j(X^*X)}$).

Parallels von Neumann $\leftrightarrow$ Davis run deeper...

9. DUALITY

Von Neumann's proof was duality-based.

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \ \text{bounded},$$

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \ \text{bounded},$$

then the **dual function**

$$G_*(Y) \ = \ \sup\{\langle X, Y \rangle : G(X) \leq 1\}$$

is a norm.

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \ \text{bounded},$$

then the **dual function**

$$G_*(Y) \ = \ \sup\{\langle X, Y \rangle : G(X) \leq 1\}$$

is a norm. Furthermore, G is a norm $\ \Leftrightarrow\ G = G_{**}$.

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \ \text{bounded},$$

then the **dual function**

$$G_*(Y) \ = \ \sup\{\langle X, Y\rangle : G(X) \leq 1\}$$

is a norm. Furthermore, G is a norm $\Leftrightarrow$ $G = G_{**}$.

For invariant G on $\mathbf{M}^n$ (with $\langle X, Y\rangle = \mathrm{Re}\,\mathrm{trace}(X^*Y)$),

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \le 1\} \ \text{bounded},$$

then the **dual function**

$$G_*(Y) \ = \ \sup\{\langle X, Y\rangle : G(X) \le 1\}$$

is a norm. Furthermore, G is a norm $\Leftrightarrow$ $G = G_{**}$.

For invariant G on $\mathbf{M}^n$ (with $\langle X, Y\rangle = \mathrm{Re}\,\mathrm{trace}(X^*Y)$),
if $G|_{\mathbf{D}^n}$ is a norm, $G|_{\mathbf{D}^n} = (G|_{\mathbf{D}^n})_{**}$,

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \ \text{bounded},$$

then the **dual function**

$$G_*(Y) \ = \ \sup\{\langle X, Y \rangle : G(X) \leq 1\}$$

is a norm. Furthermore, G is a norm $\Leftrightarrow$ $G = G_{**}$.

For invariant G on $\mathbf{M}^n$ (with $\langle X, Y \rangle = \mathrm{Re}\,\mathrm{trace}(X^* Y)$),
if $G|_{\mathbf{D}^n}$ is a norm, $G|_{\mathbf{D}^n} = (G|_{\mathbf{D}^n})_{**}$, so

$$G \ = \ (G|_{\mathbf{D}^n})_{**} \circ \mathrm{Diag} \circ \sigma$$

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \ \text{bounded},$$

then the **dual function**

$$G_*(Y) \ = \ \sup\{\langle X, Y\rangle : G(X) \leq 1\}$$

is a norm. Furthermore, G is a norm $\Leftrightarrow$ $G = G_{**}$.

For invariant G on $\mathbf{M}^n$ (with $\langle X, Y\rangle = \mathrm{Re}\,\mathrm{trace}(X^*Y)$),
if $G|_{\mathbf{D}^n}$ is a norm, $G|_{\mathbf{D}^n} = (G|_{\mathbf{D}^n})_{**}$, so

$$G \ = \ (G|_{\mathbf{D}^n})_{**} \circ \mathrm{Diag} \circ \sigma \ = \ ((G|_{\mathbf{D}^n})_* \circ \mathrm{Diag} \circ \sigma)_*$$

(by a variational argument),

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \text{ bounded,}$$

then the **dual function**

$$G_*(Y) \;=\; \sup\{\langle X, Y\rangle : G(X) \leq 1\}$$

is a norm. Furthermore, G is a norm $\Leftrightarrow$ $G = G_{**}$.

For invariant G on $\mathbf{M}^n$ (with $\langle X, Y\rangle = \mathrm{Re}\,\mathrm{trace}(X^*Y)$),
if $G|_{\mathbf{D}^n}$ is a norm, $G|_{\mathbf{D}^n} = (G|_{\mathbf{D}^n})_{**}$, so

$$G \;=\; (G|_{\mathbf{D}^n})_{**} \circ \mathrm{Diag} \circ \sigma \;=\; ((G|_{\mathbf{D}^n})_* \circ \mathrm{Diag} \circ \sigma)_*$$

(by a variational argument), so G is a norm. $\square$

9. DUALITY

Von Neumann's proof was duality-based.

If $\mathbf{E}$ is a Euclidean space and $G : \mathbf{E} \to \mathbf{R}_+$ satisfies

$$G(\alpha X) = |\alpha| G(X) \quad (\alpha \in \mathbf{R},\ X \in \mathbf{E})$$
$$\{X : G(X) \leq 1\} \text{ bounded,}$$

then the **dual function**

$$G_*(Y) \ = \ \sup\{\langle X, Y \rangle : G(X) \leq 1\}$$

is a norm. Furthermore, G is a norm $\Leftrightarrow$ $G = G_{**}$.

For invariant G on $\mathbf{M}^n$ (with $\langle X, Y \rangle = \mathrm{Re}\,\mathrm{trace}(X^* Y)$),
if $G|_{\mathbf{D}^n}$ is a norm, $G|_{\mathbf{D}^n} = (G|_{\mathbf{D}^n})_{**}$, so

$$G \ = \ (G|_{\mathbf{D}^n})_{**} \circ \mathrm{Diag} \circ \sigma \ = \ ((G|_{\mathbf{D}^n})_* \circ \mathrm{Diag} \circ \sigma)_*$$

(by a variational argument), so G is a norm. $\square$

Note also the **duality formula** $\quad G_*|_{\mathbf{D}^n} = (G|_{\mathbf{D}^n})_*.$

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y \rangle - f(X)\}$$

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y \rangle - f(X)\}$$

is lower semicontinuous and convex.

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y\rangle - f(X)\}$$

is lower semicontinuous and convex. Furthermore,
F is lsc and convex $\;\Leftrightarrow\;$ $F = F^{**}$.

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y\rangle - f(X)\}$$

is lower semicontinuous and convex. Furthermore,
F is lsc and convex $\Leftrightarrow$ $F = F^{**}$.

Conjugacy is central for convex optimization duality theory.

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y\rangle - f(X)\}$$

is lower semicontinuous and convex. Furthermore,
F is lsc and convex $\Leftrightarrow$ $F = F^{**}$.

Conjugacy is central for convex optimization duality theory.
Example Typical convex F, G satisfy **Fenchel duality**:

$$\inf_X\{F(X) + G(X)\} \;=\; \sup_Y\{-F^*(Y) - G^*(-Y)\}.$$

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y\rangle - f(X)\}$$

is lower semicontinuous and convex. Furthermore,
F is lsc and convex $\Leftrightarrow$ $F = F^{**}$.

Conjugacy is central for convex optimization duality theory.
Example Typical convex F, G satisfy **Fenchel duality**:

$$\inf_X\{F(X) + G(X)\} \;=\; \sup_Y\{-F^*(Y) - G^*(-Y)\}.$$

Imitating von Neumann's argument now gives the Davis result (for lsc F),

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y\rangle - f(X)\}$$

is lower semicontinuous and convex. Furthermore,
F is lsc and convex $\Leftrightarrow$ $F = F^{**}$.

Conjugacy is central for convex optimization duality theory.
Example Typical convex F, G satisfy **Fenchel duality**:

$$\inf_X\{F(X) + G(X)\} \;=\; \sup_Y\{-F^*(Y) - G^*(-Y)\}.$$

Imitating von Neumann's argument now gives the Davis result (for lsc F), and the duality formula

$$F^*|_{\mathbf{D}^n} = (F|_{\mathbf{D}^n})^*.$$

10. CONJUGATES OF SPECTRAL FUNCTIONS

The **Fenchel conjugate** of a function $F : \mathbf{E} \to (-\infty, +\infty]$,

$$F^*(Y) = \sup\{\langle X, Y\rangle - f(X)\}$$

is lower semicontinuous and convex. Furthermore,
F is lsc and convex $\Leftrightarrow$ $F = F^{**}$.

Conjugacy is central for convex optimization duality theory.
Example Typical convex F, G satisfy **Fenchel duality**:

$$\inf_X\{F(X) + G(X)\} \; = \; \sup_Y\{-F^*(Y) - G^*(-Y)\}.$$

Imitating von Neumann's argument now gives the Davis result (for lsc F), and the duality formula

$$F^*|_{\mathbf{D}^n} = (F|_{\mathbf{D}^n})^*.$$

What is the unifying thread?

11. HORN'S THEOREM (1954)

One route to the Davis result. . .

11. HORN'S THEOREM (1954)

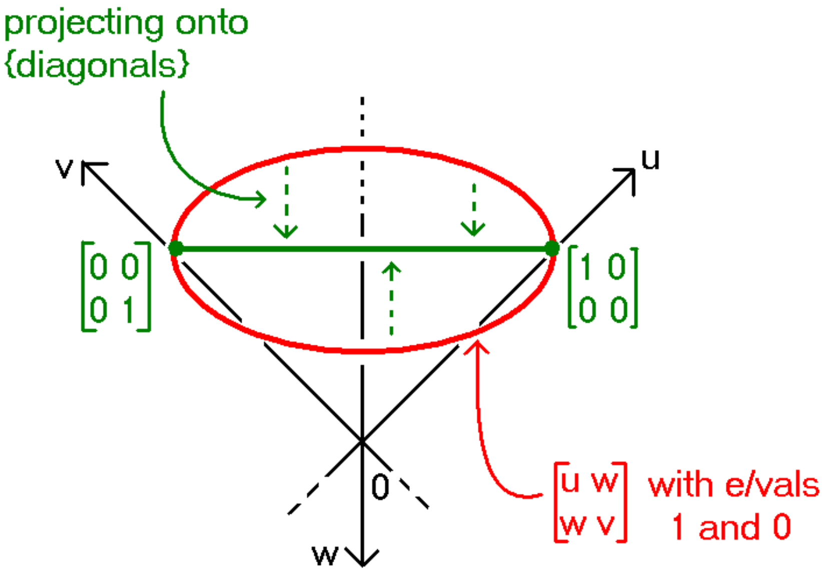
One route to the Davis result...

Theorem Convex combinations of permutations of $y \in \mathbf{R}^n$ give all possible diagonals of $X \in \mathbf{S}^n$ with eigenvalues $\{y_i\}$.

11. HORN'S THEOREM (1954)

One route to the Davis result...

Theorem Convex combinations of permutations of $y \in \mathbf{R}^n$ give all possible diagonals of $X \in \mathbf{S}^n$ with eigenvalues $\{y_i\}$.



12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973)

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G,

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$,

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

Then for $x \in \mathfrak{a}$, $\quad \mathrm{proj}_{\mathfrak{a}}(K \cdot x) = \mathrm{conv}\,(W \cdot x)$,

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

Then for $x \in \mathfrak{a}$, $\qquad \mathrm{proj}_{\mathfrak{a}}(K \cdot x) = \mathrm{conv}\,(W \cdot x)$,

where W is the **Weyl group**.

Example For Horn's theorem,

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

Then for $x \in \mathfrak{a}$, $$\mathrm{proj}_{\mathfrak{a}}(K \cdot x) = \mathrm{conv}\,(W \cdot x),$$
where W is the **Weyl group**.

Example For Horn's theorem, take

$$\{U \in \mathbf{M}^n : U^T U = I, \;\; \det U = 1\}$$

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

Then for $x \in \mathfrak{a}$, $\qquad \mathrm{proj}_{\mathfrak{a}}(K \cdot x) = \mathrm{conv}\,(W \cdot x)$,

where W is the **Weyl group**.

Example For Horn's theorem, take

$$\{U \in \mathbf{M}^n : U^T U = I,\ \det U = 1\} \quad \subset \quad \{U : \det U = 1\},$$

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

Then for $x \in \mathfrak{a}$,
$$\mathrm{proj}_{\mathfrak{a}}(K \cdot x) = \mathrm{conv}\,(W \cdot x),$$
where W is the **Weyl group**.

Example For Horn's theorem, take

$$\{U \in \mathbf{M}^n : U^T U = I,\ \det U = 1\} \quad \subset \quad \{U : \det U = 1\},$$
$$\{\text{skews}\} \oplus \{\text{traceless symmetric}\} \quad = \quad \{\text{traceless}\}$$

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

Then for $x \in \mathfrak{a}$,
$$\mathrm{proj}_{\mathfrak{a}}(K \cdot x) = \mathrm{conv}\,(W \cdot x),$$
where W is the **Weyl group**.

Example For Horn's theorem, take

$$\begin{aligned}
\{U \in \mathbf{M}^n : U^T U = I,\ \det U = 1\} \quad &\subset \quad \{U : \det U = 1\}, \\
\{\text{skews}\} \oplus \{\text{traceless symmetric}\} \quad &= \quad \{\text{traceless}\} \\
\mathfrak{a} = \{\text{diagonals}\} &
\end{aligned}$$

12. AN ALGEBRAIC FRAMEWORK

Theorem (Kostant 1973) Consider

- a real semisimple Lie group G, with Lie algebra $\mathfrak{g}$;
- a maximal compact subgroup $K \subset G$, with Lie algebra $\mathfrak{k}$;
- the corresponding **Cartan decomposition** $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$;
- a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$.

Then for $x \in \mathfrak{a}$, $$\mathrm{proj}_{\mathfrak{a}}(K \cdot x) = \mathrm{conv}\,(W \cdot x),$$
where W is the **Weyl group**.

Example For Horn's theorem, take

$$\{U \in \mathbf{M}^n : U^T U = I,\ \det U = 1\} \quad \subset \quad \{U : \det U = 1\},$$
$$\{\text{skews}\} \oplus \{\text{traceless symmetric}\} \quad = \quad \{\text{traceless}\}$$
$$\mathfrak{a} = \{\text{diagonals}\} \ \text{ and } \ W = \{\text{permutations}\}.$$

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

Theorem Given a maximal compact subgroup K, and a maximal abelian subspace $\mathfrak{a}$ of a corresponding Cartan subspace $\mathfrak{p}$,

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

Theorem Given a maximal compact subgroup K, and a maximal abelian subspace $\mathfrak{a}$ of a corresponding Cartan subspace $\mathfrak{p}$, consider an **invariant** function $F : \mathfrak{p} \to \overline{\mathbf{R}}$:

$$F(k \cdot x) = F(x) \quad (k \in K,\ x \in \mathfrak{p}).$$

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

Theorem Given a maximal compact subgroup K, and a maximal abelian subspace $\mathfrak{a}$ of a corresponding Cartan subspace $\mathfrak{p}$, consider an **invariant** function $F : \mathfrak{p} \to \overline{\mathbf{R}}$:

$$F(k \cdot x) = F(x) \quad (k \in K,\ x \in \mathfrak{p}).$$

Then we have:

Convexity characterization F is convex $\Leftrightarrow$ $F|_{\mathfrak{a}}$ is convex.

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

Theorem Given a maximal compact subgroup K, and a maximal abelian subspace $\mathfrak{a}$ of a corresponding Cartan subspace $\mathfrak{p}$, consider an **invariant** function $F : \mathfrak{p} \to \overline{\mathbf{R}}$:

$$F(k \cdot x) = F(x) \quad (k \in K,\ x \in \mathfrak{p}).$$

Then we have:

Convexity characterization F is convex $\Leftrightarrow$ $F|_{\mathfrak{a}}$ is convex.

Duality formula $F^*|_{\mathfrak{a}} = (F|_{\mathfrak{a}})^*$.

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

Theorem Given a maximal compact subgroup K, and a maximal abelian subspace $\mathfrak{a}$ of a corresponding Cartan subspace $\mathfrak{p}$, consider an **invariant** function $F : \mathfrak{p} \to \overline{\mathbf{R}}$:

$$F(k \cdot x) = F(x) \quad (k \in K,\ x \in \mathfrak{p}).$$

Then we have:

Convexity characterization F is convex $\Leftrightarrow F|_{\mathfrak{a}}$ is convex.

Duality formula $F^*|_{\mathfrak{a}} = (F|_{\mathfrak{a}})^*$.

For invariant functions, convexity **lifts** from $\mathfrak{a}$ to $\mathfrak{p}$,

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

Theorem Given a maximal compact subgroup K, and a maximal abelian subspace $\mathfrak{a}$ of a corresponding Cartan subspace $\mathfrak{p}$, consider an **invariant** function $F : \mathfrak{p} \to \overline{\mathbf{R}}$:

$$F(k \cdot x) = F(x) \quad (k \in K,\ x \in \mathfrak{p}).$$

Then we have:

Convexity characterization F is convex $\Leftrightarrow F|_{\mathfrak{a}}$ is convex.

Duality formula $F^*|_{\mathfrak{a}} = (F|_{\mathfrak{a}})^*$.

For invariant functions, convexity **lifts** from $\mathfrak{a}$ to $\mathfrak{p}$, as do

- polynomiality (**Chevalley restriction theorem**)

13. CONVEXITY AND INVARIANCE

Kostant's theory extends the von Neumann and Davis results.

Theorem Given a maximal compact subgroup K, and a maximal abelian subspace $\mathfrak{a}$ of a corresponding Cartan subspace $\mathfrak{p}$, consider an **invariant** function $F : \mathfrak{p} \to \overline{\mathbf{R}}$:

$$F(k \cdot x) = F(x) \quad (k \in K,\ x \in \mathfrak{p}).$$

Then we have:

Convexity characterization F is convex $\Leftrightarrow F|_{\mathfrak{a}}$ is convex.

Duality formula $F^*|_{\mathfrak{a}} = (F|_{\mathfrak{a}})^*$.

For invariant functions, convexity **lifts** from $\mathfrak{a}$ to $\mathfrak{p}$, as do

- polynomiality (**Chevalley restriction theorem**)
- differentiability.

14. SENSITIVITY

For a nonsmooth convex function $F : \mathbf{E} \rightarrow \overline{\mathbf{R}}$ at $x \in \mathbf{E}$, we replace gradients by **subgradients** $y \in \partial F(x)$:

14. SENSITIVITY

For a nonsmooth convex function $F : \mathbf{E} \to \overline{\mathbf{R}}$ at $x \in \mathbf{E}$, we replace gradients by **subgradients** $y \in \partial F(x)$:

$$F(x) + \langle y, z - x \rangle \;\leq\; F(z) \quad (z \in \mathbf{E}).$$

14. SENSITIVITY

For a nonsmooth convex function $F : \mathbf{E} \to \overline{\mathbf{R}}$ at $x \in \mathbf{E}$, we replace gradients by **subgradients** $y \in \partial F(x)$:

$$F(x) + \langle y, z - x \rangle \;\leq\; F(z) \quad (z \in \mathbf{E}).$$

The set of subgradients encodes the **directional derivative**:

14. **SENSITIVITY**

For a nonsmooth convex function $F : \mathbf{E} \to \overline{\mathbf{R}}$ at $x \in \mathbf{E}$, we replace gradients by **subgradients** $y \in \partial F(x)$:

$$F(x) + \langle y, z - x \rangle \;\leq\; F(z) \quad (z \in \mathbf{E}).$$

The set of subgradients encodes the **directional derivative**:

$$F'(x; w) = \sup\{\langle y, w \rangle : y \in \partial F(x)\}.$$

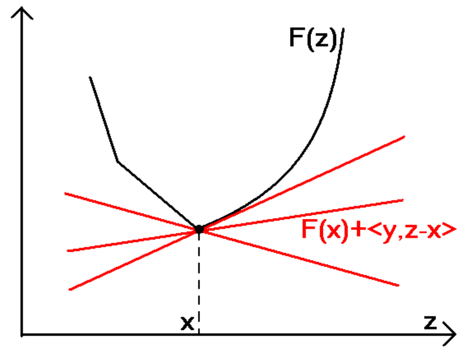
14. SENSITIVITY

For a nonsmooth convex function $F : \mathbf{E} \to \overline{\mathbf{R}}$ at $x \in \mathbf{E}$, we replace gradients by **subgradients** $y \in \partial F(x)$:

$$F(x) + \langle y, z - x \rangle \;\leq\; F(z) \quad (z \in \mathbf{E}).$$

The set of subgradients encodes the **directional derivative**:

$$F'(x; w) = \sup\{\langle y, w \rangle : y \in \partial F(x)\}.$$



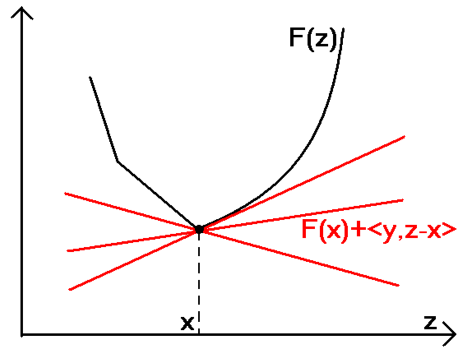
14. SENSITIVITY

For a nonsmooth convex function $F : \mathbf{E} \to \overline{\mathbf{R}}$ at $x \in \mathbf{E}$, we replace gradients by **subgradients** $y \in \partial F(x)$:

$$F(x) + \langle y, z - x \rangle \;\leq\; F(z) \quad (z \in \mathbf{E}).$$

The set of subgradients encodes the **directional derivative**:

$$F'(x; w) = \sup\{\langle y, w \rangle : y \in \partial F(x)\}.$$



Theorem In Kostant's framework, if $F : \mathfrak{p} \to \overline{\mathbf{R}}$ is K-invariant, then $y \in \partial F(x) \quad \Leftrightarrow$

$$y = k \cdot v,\;\; x = k \cdot u,\;\; \text{with } k \in K,\;\; v \in \partial F|_{\mathfrak{a}}(u).$$

15. SUBGRADIENTS OF SPECTRAL FUNCTIONS

Theorem Suppose $f : \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ is convex and permutation-invariant,

15. SUBGRADIENTS OF SPECTRAL FUNCTIONS

Theorem Suppose $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and permutation-invariant, so the spectral function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ defined by

$$F(X) = f(\lambda(X))$$

is convex (by Davis' theorem).

15. SUBGRADIENTS OF SPECTRAL FUNCTIONS

Theorem Suppose $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and permutation-invariant, so the spectral function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ defined by

$$F(X) = f(\lambda(X))$$

is convex (by Davis' theorem). Then $Y \in \partial F(X) \quad \Leftrightarrow$
X and Y have a **simultaneous spectral decomposition**,

$$U^T U = I, \quad U^T (\mathrm{Diag}\, x) U = X, \quad U^T (\mathrm{Diag}\, y) U = Y,$$

for some U and $y \in \partial f(x)$.

15. SUBGRADIENTS OF SPECTRAL FUNCTIONS

Theorem Suppose $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and permutation-invariant, so the spectral function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ defined by

$$F(X) = f(\lambda(X))$$

is convex (by Davis' theorem). Then $Y \in \partial F(X) \quad \Leftrightarrow$ X and Y have a **simultaneous spectral decomposition**,

$$U^T U = I, \quad U^T (\mathrm{Diag}\, x) U = X, \quad U^T (\mathrm{Diag}\, y) U = Y,$$

for some U and $y \in \partial f(x)$.

This subgradient formula also extends to **nonconvex** functions.

15. SUBGRADIENTS OF SPECTRAL FUNCTIONS

Theorem Suppose $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and permutation-invariant, so the spectral function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ defined by

$$F(X) = f(\lambda(X))$$

is convex (by Davis' theorem). Then $Y \in \partial F(X) \quad \Leftrightarrow$ X and Y have a **simultaneous spectral decomposition**,

$$U^T U = I, \quad U^T (\mathrm{Diag}\, x) U = X, \quad U^T (\mathrm{Diag}\, y) U = Y,$$

for some U and $y \in \partial f(x)$.

This subgradient formula also extends to **nonconvex** functions. Eg **Clarke's generalized gradient**:

15. SUBGRADIENTS OF SPECTRAL FUNCTIONS

Theorem Suppose $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ is convex and permutation-invariant, so the spectral function $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ defined by

$$F(X) = f(\lambda(X))$$

is convex (by Davis' theorem). Then $Y \in \partial F(X) \quad \Leftrightarrow$ X and Y have a **simultaneous spectral decomposition**,

$$U^T U = I, \quad U^T (\operatorname{Diag} x) U = X, \quad U^T (\operatorname{Diag} y) U = Y,$$

for some U and $y \in \partial f(x)$.

This subgradient formula also extends to **nonconvex** functions. Eg **Clarke's generalized gradient**:

$$\partial f(x) \;=\; \operatorname{cl} \operatorname{conv} \{ \lim \nabla f(x_r) : x_r \to x \}$$

(for Lipschitz f) etc...

16. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950) If $X, Z \in \mathbf{S}^n$, then $\lambda(Z) - \lambda(X)$ is a convex combination of permutations of $\lambda(Z - X)$.

16. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950) If $X, Z \in \mathbf{S}^n$, then $\lambda(Z) - \lambda(X)$ is a convex combination of permutations of $\lambda(Z - X)$.

A proof via nonsmooth analysis

16. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950) If $X, Z \in \mathbf{S}^n$, then $\lambda(Z) - \lambda(X)$ is a convex combination of permutations of $\lambda(Z - X)$.

A proof via nonsmooth analysis

- Via a separating hyperplane, we need, for any $w \in \mathbf{R}^n$

$$w^T(\lambda(Z) - \lambda(X))$$

16. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950) If $X, Z \in \mathbf{S}^n$, then $\lambda(Z) - \lambda(X)$ is a convex combination of permutations of $\lambda(Z - X)$.

A proof via nonsmooth analysis

- Via a separating hyperplane, we need, for any $w \in \mathbf{R}^n$

$$w^T(\lambda(Z) - \lambda(X)) \;\leq\; [w]^T \lambda(Z - X),$$

where $w \mapsto [w]$ maps components into decreasing order.

16. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950) If $X, Z \in \mathbf{S}^n$, then $\lambda(Z) - \lambda(X)$ is a convex combination of permutations of $\lambda(Z - X)$.

A proof via nonsmooth analysis

- Via a separating hyperplane, we need, for any $w \in \mathbf{R}^n$

$$w^T(\lambda(Z) - \lambda(X)) \;\leq\; [w]^T \lambda(Z - X),$$

where $w \mapsto [w]$ maps components into decreasing order.

- Consider the (nonconvex) spectral function

$$F(X) = w^T \lambda(X).$$

16. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950) If $X, Z \in \mathbf{S}^n$, then $\lambda(Z) - \lambda(X)$ is a convex combination of permutations of $\lambda(Z - X)$.

A proof via nonsmooth analysis

- Via a separating hyperplane, we need, for any $w \in \mathbf{R}^n$

  $$w^T(\lambda(Z) - \lambda(X)) \;\leq\; [w]^T \lambda(Z - X),$$

  where $w \mapsto [w]$ maps components into decreasing order.

- Consider the (nonconvex) spectral function

  $$F(X) = w^T \lambda(X).$$

  A **nonsmooth mean value theorem** shows

  $$F(Z) - F(X) = \langle Y, Z - X \rangle$$

  for some $Y \in \partial F(W)$ where $W \in [X, Z]$.

16. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950) If $X, Z \in \mathbf{S}^n$, then $\lambda(Z) - \lambda(X)$ is a convex combination of permutations of $\lambda(Z - X)$.

A proof via nonsmooth analysis

- Via a separating hyperplane, we need, for any $w \in \mathbf{R}^n$

$$w^T(\lambda(Z) - \lambda(X)) \;\leq\; [w]^T\lambda(Z - X),$$

  where $w \mapsto [w]$ maps components into decreasing order.

- Consider the (nonconvex) spectral function

$$F(X) = w^T\lambda(X).$$

  A **nonsmooth mean value theorem** shows

$$F(Z) - F(X) = \langle Y, Z - X\rangle$$

  for some $Y \in \partial F(W)$ where $W \in [X, Z]$.

- Now apply the subgradient formula. $\square$

17. SUMMARY

- **Hyperbolic polynomials** give a simple, general framework for studying **primal** convex optimization.

17. SUMMARY

- **Hyperbolic polynomials** give a simple, general framework for studying **primal** convex optimization. But:
  - No apparent duality theory;

17. SUMMARY

- **Hyperbolic polynomials** give a simple, general framework for studying **primal** convex optimization. But:
  - No apparent duality theory;
  - Is it really more general than semidefinite programming?

17. SUMMARY

- **Hyperbolic polynomials** give a simple, general framework for studying **primal** convex optimization. But:
  - No apparent duality theory;
  - Is it really more general than semidefinite programming?
- Semisimple **Lie theory** gives a broad framework for studying **duality**:

17. SUMMARY

- **Hyperbolic polynomials** give a simple, general framework for studying **primal** convex optimization. But:
  - No apparent duality theory;
  - Is it really more general than semidefinite programming?
- Semisimple **Lie theory** gives a broad framework for studying **duality**:
  - Fenchel conjugates;

17. SUMMARY

- **Hyperbolic polynomials** give a simple, general framework for studying **primal** convex optimization. But:
  - No apparent duality theory;
  - Is it really more general than semidefinite programming?
- Semisimple **Lie theory** gives a broad framework for studying **duality**:
  - Fenchel conjugates;
  - convex and nonconvex subdifferentials.