1. OUTLINE

PART I: SYMMETRIC MATRICES

- Hyperbolic polynomials, Lax, and convex optimization
- Convex spectral functions and unitarily invariant norms
- Duality, subgradients, and some Lie algebra

PART II: NONSYMMETRIC MATRICES

- Spectral radius, transient dynamics, and Kreiss’s theorem
- Pseudospectral analysis and optimization
- Distance to uncontrollability
PART I:
SYMMETRIC MATRICES
2. HYPERBOLIC POLYNOMIALS

Example: Consider the homogeneous polynomial

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

For all real \( v, w \), \( p(u, v, w) = 0 \) \( \Rightarrow \) \( u \) 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 \( \mathbb{R}^3 \) relative to \( (1, 0, 0) \) are all of the form

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

3. HYPERBOLICITY CONES

Hyperbolic polynomials

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

**Theorem (Gårding 1951)** Hyperbolic $p$ relative to $d$ has convex **hyperbolicity cone**—component of $d$ in $\{x : p(x) > 0\}$.

Linear optimization over hyperbolicity cones is **tractable**!

**Example:** The **determinant** is hyperbolic relative to $I$ on $S^n = \{n \times n$ real symmetric matrices} : 

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

each $X \in S^n$ has all real eigenvalues $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

The hyperbolicity cone is $S^n_{++} = \{\text{positive definites}\}$.

Hence **semidefinite programming** (generalizing LP).
4. CONVEXITY AND SYMMETRY

Convexity of $S^{n}_{++}$ and $-\log \det$ are special cases of:

**Theorem (Davis 1957)** Convexity and permutation-invariance of $f : \mathbb{R}^{n} \rightarrow \mathbb{R}$ $\Rightarrow$ convexity of

$$X \in S^{n} \mapsto f(\lambda_{1}(X), \ldots, \lambda_{n}(X)).$$

(Eg: $f(x) =$

$$\begin{cases} 
0 & (x > 0) \\
+\infty & (x \nless 0)
\end{cases} \quad \text{or} \quad \begin{cases} 
-\sum_{i} \log x_{i} & (x > 0) \\
+\infty & (x \nless 0)
\end{cases}$$

Extends to hyperbolic $p$ (relative to $d$): interpret $\{\lambda_{i}(x)\}$ as the roots of $t \mapsto p(x - td)$ (Bauschke... 2001). Similarly:

**Lidskii’s Theorem (extended)** $\lambda(z) - \lambda(x)$ is a convex combination of permutations of $\lambda(z - x)$.

Lax conjecture reduces both results to $S^{n}$ case (Gurvits 2004).
5. INVARIANCE

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

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

because then the spectral decomposition $\Rightarrow$

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

The Davis result characterizes convex spectral functions:

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

Reminiscent of . . .

**Theorem (von Neumann 1937)** Unitarily invariant $G : M^n \rightarrow \mathbb{R}$ is a norm $\iff$ the restriction $G|_{D^n}$ is a norm.
6. DUALITY

Von Neumann’s proof depended on a duality formula:

\[ G^*|_{D^n} = (G|_{D^n})^* , \]

for the dual function

\[ G^*(Y) = \sup\{\langle X, Y \rangle : G(X) \leq 1 \}. \]

We can prove Davis’s result by analogy. The Fenchel conjugate of a function \( F : E \to (-\infty, +\infty] \) is

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

and the duality formula

\[ F^*|_{D^n} = (F|_{D^n})^*. \]

holds for spectral \( F \).

What’s the unifying thread?
7. **HORN’S THEOREM (1954)**

One route to the Davis result...

**Theorem** Convex combinations of permutations of $y \in \mathbb{R}^n$ give all possible diagonals of $X \in S^n$ with eigenvalues $\{y_i\}$.
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 $x \in \mathfrak{a}$, $\text{proj}_\mathfrak{a}(K \cdot x) = \text{conv} (W \cdot x)$, where $W$ is the **Weyl group**.

**Example** For Horn’s theorem, take

\[
\begin{align*}
\{ U \in \mathbb{M}^n : U^TU &= I, \ \det U = 1 \} & \subset \{ U : \det U = 1 \}, \\
\{ \text{skews} \} \oplus \{ \text{traceless symmetric} \} & = \{ \text{traceless} \} \\
\mathfrak{a} = \{ \text{diagonals} \} & \text{ and } W = \{ \text{permutations} \}.
\end{align*}
\]
9. CONVEXITY AND INVARIANCE

The von Neumann and Davis results extend to this setting:

**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 $F : \mathfrak{p} \to \mathbb{R}$:

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

Then we have:

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

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

**Convex subdifferential** $y \in \partial F(x)$,

i.e. $F(x) + \langle y, z - x \rangle \leq F(z) \ \forall \ z \in \mathbb{E}$

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

In the $S^n$ case, this becomes...
10. SPECTRAL SUBDIFFERENTIALS

**Theorem**  If \( f : \mathbb{R}^n \to \overline{\mathbb{R}} \) is convex, permutation-invariant, so \( X \in S^n \mapsto F(X) = f(\lambda(X)) \) is convex (by Davis), then: \( Y \in \partial F(X) \iff \text{simultaneous spectral decomposition}, \)

\[
U^TU = I, \quad U^T(\text{Diag } x)U = X, \quad U^T(\text{Diag } y)U = Y,
\]

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

The subdifferential extends to nonconvex, Lipschitz \( f \):

\[
\partial f(x) = \text{conv} \left\{ \lim \nabla f(x_r) : x_r \to x \right\}
\]

(Clarke, 1974). The same characterization holds.

An illustration of nonsmooth analysis in linear algebra...
11. EIGENVALUE PERTURBATION THEORY

Theorem (Lidskii 1950)  If $X, Z \in \mathbb{S}^n$, 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 $w \in \mathbb{R}^n$
  \[ w^T(\lambda(Z) - \lambda(X)) \leq [w]^T \lambda(Z - X), \]
  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 subdifferential formula.  \qed
PART II:

NONSYMMETRIC MATRICES
THE SPECTRAL RADIUS

Question  How should we design a (parametrized) square matrix $A$ to force $A^n \to 0$ quickly as $n \to \infty$?

Theorem  Rate of decay

$$\inf\{\mu : A^n = O(\mu^n) \text{ as } n \to \infty\}$$

equals spectral radius

$$\rho(A) = \max\{|\lambda| : \lambda \in \Lambda(A)\},$$

where $\Lambda(A) = \{\text{eigenvalues of } A\}$ is the spectrum.

Example

The spectral radius of

$$A(t) = \begin{bmatrix} k & 1 \\ t & k - t \end{bmatrix}$$

(with $k$ slightly less than 1)

is minimized at $t = 0$. 

\[0\]
13. **ROBUSTNESS AND TRANSIENT PEAKS**

But

$$A(0) = \begin{bmatrix} k & 1 \\ 0 & k \end{bmatrix}$$

may be unsatisfactory.

**Difficulty I:** $\rho(A(t))$ is highly sensitive to perturbation at $t = 0$ (nonlipschitz).

**Difficulty II:** The trajectory $\{A(0)^n\}$ has a big transient peak:

$$\begin{bmatrix} \frac{n}{n+1} & 1 \\ 0 & \frac{n}{n+1} \end{bmatrix}^n \sim e^{-1} \begin{bmatrix} 1 & n + 1 \\ 0 & 1 \end{bmatrix}$$

for large $n$.

One difficulty is the **multiple eigenvalue**. But this is **typical** at optimal solutions of spectral radius minimization problems. (Burke/Lewis/Overton 2001)
14. PSEUDOSPECTRA

(Following Trefethen...)

A powerful tool to visualize robust properties of eigenvalues:

\[ \Lambda_\epsilon(A) = \bigcup_{\|X - A\| \leq \epsilon} \Lambda(X) = \{ z \in \mathbb{C} : \sigma_{\text{min}}(A - zI) \leq \epsilon \}, \]

where \( \sigma_{\text{min}} \) is the smallest singular value.

Pseudospectra resolve Difficulty II (transient peaks)...

**Kreiss Matrix Theorem (1962)**

\( A^n < K \rho^n \) for all \( n \), with \( K \) not too large \iff \[
\max\{|\lambda| : \lambda \in \Lambda_\epsilon(A)\} < \rho, \] with \( \epsilon \) not too small.

Analogously, in continuous time, \( e^{At} \to 0 \) with peaks not too large when \( \Lambda_\epsilon(A) \) lies in the left halfplane for \( \epsilon \) not too small.
15. **EXAMPLES**

Pseudospectra for a random $5 \times 5$ triangular complex matrix, plotted by T. Wright’s EigTool:

Demmel’s example: \[ A = \begin{bmatrix} 1 & 5 & 5^2 & 5^3 & 5^4 \\ 0 & 1 & 5 & 5^2 & 5^3 \\ 0 & 0 & 1 & 5 & 5^2 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \ldots \]
Notice \( \Lambda_{0.01}(A) \) extends outside the left halfplane: some “unstable” \( X \) satisfies \( \| X - A \| \leq 0.01 \).
We can also compute pseudospectral quantities like

\[ \alpha_\epsilon(A) = \max\{\Re \lambda : \lambda \in \Lambda_\epsilon(A)\} \]

Key pseudospectral properties:

- Components all contain eigenvalues (by maximum modulus principle)
- Intersections with lines (or circles) easy to compute by a Hamiltonian eigensolver.

Hence a fast, accurate, robust criss-cross algorithm for \( \alpha_\epsilon \):

globally and quadratically convergent, available in eigtool.

Returns \( \nabla \alpha_\epsilon \) (when it exists) \( \rightarrow \) nonsmooth gradient sampling for optimizing \( \alpha_\epsilon \) (Burke/Lewis/Overton 2003).
18. **Lipschitz Behavior**

**Difficulty I?** $A \mapsto \Lambda(A)$ isn’t locally **Lipschitz**: no $k$ satisfies

$$d(\Lambda(X), \Lambda(Y)) \leq k\|X - Y\| \quad \text{for all } X, Y \text{ near } A,$$

where $d$ is Hausdorff distance between $U, V \subset \mathbb{C}$.

The pseudospectral map $A \mapsto \Lambda_\epsilon(A)$ can also be nonlipschitz:

\[
A = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}, \quad r = \frac{\sqrt{5} - 1}{2}.
\]

**But...**

**Theorem** Typically (Arnold 1971) $A$ has all eigenspaces one-dimensional. Then, $\Lambda_\epsilon$ is Lipschitz around $A$ for all small $\epsilon > 0$. 
19. CONTROLLABILITY

A system with state $x \in \mathbb{C}^m$ and control $u$,

$$\frac{dx}{dt} = Ax + Bu,$$

controllable if any endpoints can be interpolated by a control.

**Lemma (Hautus 1969)** A matrix-pair $(A, B)$ is controllable

$$\iff \delta = \min\{\sigma_{\min}[A - zI, B] : z \in \mathbb{C}\} > 0.$$  

Eising (1984) showed $\delta$ is the distance to uncontrollability:

$$\delta = \min\{\| (X, Y) \| : (A + X, B + Y) \text{ is uncontrollable}\}$$

Polynomially computable (Gu 2000, 2005) but tough...

**Question (Trefethen)** How many components has the rectangular pseudospectrum

$$\{z \in \mathbb{C} : \sigma_{\min}[A - zI, B] \leq \epsilon\}?$$
20. CONNECTED COMPONENTS

Problem: For \( m \)-by-\( n \) \( P, Q \), bound number of components

\[
c = \# \{ z : \sigma_{\min}(P + zQ) \leq \epsilon \}.
\]

Conjecture: \( c \leq m \). (Easy if \( m = 1 \) or \( m = n \).)

Theorem \( c \leq 2m^2 - m + 1 \). (Burke/Lewis/Overton 2004)

Proof Von Neumann-Wigner (1929) showed \( \{ m \text{-by-} m \text{ Hermitians with multiple eigenvalues} \} \) has codimension 3. Assume \((P, Q) \) “typical”, so

\[
\lambda_{\min}((P + zQ)(P + zQ)^*) \text{ simple } \forall z \in \mathbb{C}.
\]

Milnor (1964) showed degree-\( d \) polynomials \( p \) on \( \mathbb{R}^2 \) satisfy

\[
\# \{(x, y) : p(x, y) = 0\} \leq d(2d - 1),
\]

Result follows by applying this to \( p : \mathbb{C} \cong \mathbb{R}^2 \rightarrow \mathbb{R} \) given by

\[
p(z) = \det((P + zQ)(P + zQ)^* - \epsilon^2 I).
\]

General case follows by perturbation.
21. **SUMMARY**

- Hyperbolicity is elegant for primal convex optimization.
  - But no apparent duality theory;
  - Is it really more general than semidefinite programming?
- Semisimple Lie theory gives a broad duality framework:
  - Fenchel conjugates;
  - convex and nonconvex subdifferentials.
- Spectral radius minimization leads to
  - multiple eigenvalues,
  - nonrobust solutions,
  - transient peaks.
- Pseudospectral optimization circumvents these (by Kreiss).
- Distance to uncontrollability is tractable.