Matrix Monotone Functions

Sai Subramanyam Thoota

SPC Lab, Department of ECE
Indian Institute of Science

December 14, 2019

Motivation

- To solve communication and information theoretic problems in wireless communications.
- Unified framework for the performance analysis of multiple antenna systems based on matrix-monotone functions.

Matrix monotonicity

- Set of all positive semidefinite matrices of size $n \times n$ defined as $\mathcal{H}_n$
- Löwner ordering: $\mathbf{A} \leq \mathbf{B}$ means $\mathbf{B} - \mathbf{A}$ is positive semidefinite.
- Let $\mathcal{A} \subset \mathcal{H}_n$. A function $\phi : \mathcal{A} \rightarrow \mathcal{H}_n$ is **matrix-increasing of order n** on $\mathcal{A}$ if

$$\mathbf{A} \leq \mathbf{B} \implies \phi(\mathbf{A}) \leq \phi(\mathbf{B})$$

  for $\mathbf{A}$, $\mathbf{B} \in \mathcal{A}$. The function is strictly matrix-increasing of order n on $\mathcal{A}$ if

$$\mathbf{A} < \mathbf{B} \implies \phi(\mathbf{A}) < \phi(\mathbf{B})$$

- If a function is matrix-increasing for all orders $n \geq 1$, it is called matrix-increasing or also matrix-monotone.

Example

- On the set of positive definite matrices, the function $\phi(\mathbf{A}) = \mathbf{A}^{-1}$ is strictly decreasing.

Proof

- Let $\mathbf{Q}_1 > \mathbf{Q}_2 > \mathbf{0}$
- Define $g(t) = (t\mathbf{Q}_1 + (1-t)\mathbf{Q}_2)^{-1} = \mathbf{Q}(t)^{-1}$
- It suffices to prove that $g(t)$ is strictly decreasing in $0 \leq t \leq 1$.

$$\mathbf{Q}(t)\mathbf{Q}(t)^{-1} = \mathbf{I}$$
$$\frac{\partial \mathbf{Q}(t)}{\partial t}\mathbf{Q}(t)^{-1} + \mathbf{Q}(t)\frac{\partial \mathbf{Q}(t)^{-1}}{\partial t} = \mathbf{0}$$
$$\implies \frac{\partial \mathbf{Q}(t)^{-1}}{\partial t} = -\mathbf{Q}(t)^{-1}(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{Q}(t)^{-1} < 0.$$

Matrix Convexity

- Let $\mathcal{A} \subset \mathcal{H}_n$. A function $\phi : \mathcal{A} \to \mathcal{H}_n$ is matrix-convex of order n if

$$\phi(\alpha\mathbf{A} + \bar{\alpha}\mathbf{B}) \leq \alpha\phi(\mathbf{A}) + \bar{\alpha}\phi(\mathbf{B}) \ \forall \ \ \alpha \in [0,1] \text{ and } \mathbf{A}, \mathbf{B} \in \mathcal{A}.$$

- A function is matrix-convex if it is matrix-convex for all orders $n \geq 1$.
- A nonnegative continuous function on $[0, \infty)$ is operator monotone if and only if it is operator concave.
    - However, not every matrix-convex function is necessarily matrix-monotone.
    - $\phi(\mathbf{A}) = \mathbf{A}^2$ is matrix-convex but not matrix-monotone

Proposition:

- Let $\boldsymbol{\phi}$ be a function defined on a convex set $\mathcal{A}$ of $m \times k$ matrices, taking values in $\mathcal{H}_n$ for some n. If $\mathcal{A}$ is open and $\mathbf{g}$ is twice differentiable for all $\mathbf{A}, \mathbf{B} \in \mathcal{A}$ the following are equivalent:
  - ϕ is matrix-convex on $\mathcal{A}$.
  - For all fixed $\mathbf{A}$ and $\mathbf{B}$ in $\mathcal{A}$, the function $\mathbf{g}(\alpha) = \boldsymbol{\phi}(\alpha\mathbf{A} + \bar{\alpha}\mathbf{B})$ is convex in $\alpha \in [0,1]$ in the sense that $\eta\mathbf{g}(\alpha) + \bar{\eta}\mathbf{g}(\beta) - \mathbf{g}(\eta\alpha + \bar{\eta}\beta)$ is positive semidefinite for all $\alpha, \beta, \eta \in [0,1]$.
  - For all fixed $\mathbf{A}, \mathbf{B} \in \mathcal{A}$, $\frac{d^2\mathbf{g}(\alpha)}{d\alpha^2}$ is positive semidefinite for $0 < \alpha < 1$.

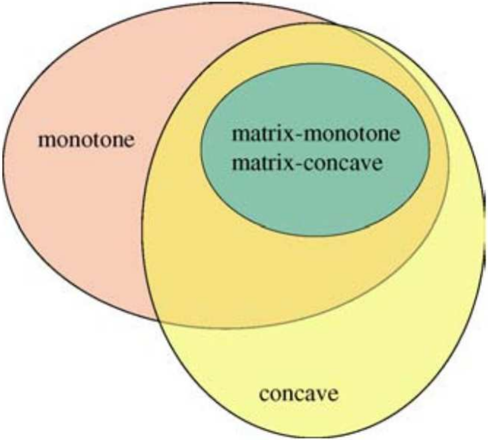
Example:

- $f(\mathbf{X}) = \log(\mathbf{I} + \mathbf{X})$ is matrix-concave

$$\frac{d^2 \log(\mathbf{I} + \alpha\mathbf{X})}{d\alpha^2} = -\mathbf{X}[\mathbf{I} + \alpha\mathbf{X}]^{-2}\mathbf{X} \leq \mathbf{0}$$

Corollary:

- Every matrix-monotone (matrix-convex) function is monotonic (convex) whereas not every monotonic (convex) function is matrix monotone (matrix-convex).



- Matrix-monotone functions are matrix-concave, concave, and monotone.

Frechet Derivative

- Corresponding to the first and second derivatives of scalar functions, there exists a derivative of a matrix valued function ϕ.
- The map ϕ is called (Frechet) differentiable at $\mathbf{A}$ if there exists a linear transformation $D\phi(\mathbf{A})$ on the space of positive semidefinite matrices such that for all $\mathbf{H}$

$$\|\phi(\mathbf{A}+\mathbf{H}) - \phi(\mathbf{A}) - D\phi(\mathbf{A})(\mathbf{H})\| = o(\|\mathbf{H}\|)$$

- A direction $\mathbf{H}$ is needed to define the derivative.

Example

- The first derivative of $\phi(\mathbf{A}) = \mathbf{A}^p$ in direction of $\mathbf{B}$ is given by

$$D\phi(\mathbf{A})(\mathbf{B}) = \sum_{k=1}^{p} \mathbf{A}^{k-1}\mathbf{B}\mathbf{A}^{p-k}$$

Proof:

- Let $\phi(\epsilon) = (\mathbf{A} + \epsilon\mathbf{B})^p$.
- Using product rule,

$$\frac{d\phi(\epsilon)}{d\epsilon} = \left[\frac{d}{d\epsilon}(\mathbf{A} + \epsilon\mathbf{B})\right](\mathbf{A} + \epsilon\mathbf{B})^{p-1} + (\mathbf{A} + \epsilon\mathbf{B})\left[\frac{d}{d\epsilon}(\mathbf{A} + \epsilon\mathbf{B})\right](\mathbf{A} + \epsilon\mathbf{B})^{p-2}$$
$$+ \ldots + (\mathbf{A} + \epsilon\mathbf{B})^{p-1}\left[\frac{d}{d\epsilon}(\mathbf{A} + \epsilon\mathbf{B})\right]$$
$$= \sum_{k=1}^{p} (\mathbf{A} + \epsilon\mathbf{B})^{k-1}\left[\frac{d}{d\epsilon}(\mathbf{A} + \epsilon\mathbf{B})\right](\mathbf{A} + \epsilon\mathbf{B})^{p-k}$$

- Evaluating the differential at $\epsilon = 0$ gives the solution.

Some more examples:

$$\phi(\mathbf{A}) = \mathbf{A}^2 \qquad D\phi(\mathbf{A})(\mathbf{B}) = \mathbf{A}\mathbf{B} + \mathbf{B}\mathbf{A}$$

$$\phi(\mathbf{A}) = \mathbf{A}^{-1} \qquad D\phi(\mathbf{A})(\mathbf{B}) = -\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$$

$$\phi(\mathbf{A}) = \mathbf{A}^H\mathbf{A} : \qquad D\phi(\mathbf{A})(\mathbf{B}) = \mathbf{A}^H\mathbf{B} + \mathbf{B}^H\mathbf{A}$$

Properties

- Linear.
- Composition of two differentiable maps is differentiable.

First Divided Difference:

- Closely related to the Frechet derivative.
- Used to characterize the class of matrix-monotone functions.
- Let $\mathbf{I}$ be an open interval. Let ϕ be a continuously differentiable function on $\mathbf{I}$. Then, we denote by $\phi^{[1]}$ the function on $\mathbf{I} \times \mathbf{I}$ defined as

$$\phi^{[1]}(\lambda_1, \lambda_2) = \frac{\phi(\lambda_1) - \phi(\lambda_2)}{\lambda_1 - \lambda_2}, \qquad \text{if} \quad \lambda_1 \neq \lambda_2$$
$$\phi^{[1]}(\lambda_1, \lambda_1) = \phi'(\lambda_1).$$

- $\phi^{[1]}(\lambda_1, \lambda_2)$ is called the first divided difference of ϕ at (λ_1, λ_2).

Matrix of first divided differences

- The matrix of first divided differences $\phi^{[1]}(\mathbf{A})$ for p.s.d. $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$ is defined as

$$\phi^{[1]}(\mathbf{A}) = \mathbf{U}\phi^{[1]}(\boldsymbol{\Lambda})\mathbf{U}^H.$$

- Applying the diagonal matrix $\boldsymbol{\Lambda}$ with entries $\lambda_1, \ldots, \lambda_n$, the function $\phi^{[1]}$ is defined as an $n \times n$ matrix with

$$\left[\phi^{[1]}(\boldsymbol{\Lambda})\right]_{j,k} = \phi^{[1]}(\lambda_j, \lambda_k).$$

Connection between the matrix of first divided differences and Frechet derivative

- If ϕ is a polynomial function and $\mathbf{A}$ is p.s.d., then

$$D\phi(\mathbf{A})(\mathbf{H}) = \phi^{[1]}(\mathbf{A}) \circ \mathbf{H}.$$

Relation between $\phi^{[1]}$ and derivative ϕ' of the scalar function $\phi(t)$:

- For any matrix-monotone function ϕ and Hermitian matrices $\mathbf{A}$ and $\mathbf{D}$, the following identity holds

$$\mathrm{tr}\left(\phi^{[1]}(\mathbf{A}) \circ \mathbf{D}\right) = \mathrm{tr}\left(\phi'(\mathbf{A}) \cdot \mathbf{D}\right).$$

Proof:

$$\begin{aligned}
\mathrm{tr}\left(\phi^{[1]}(\mathbf{A}) \circ \mathbf{D}\right) &= \frac{d}{d\epsilon}\Big|_{\epsilon=0} \mathrm{tr}\ \phi(\mathbf{A} + \epsilon\mathbf{D}) = \frac{d}{d\epsilon}\Big|_{\epsilon=0} \mathrm{tr}\ \phi(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H + \epsilon\mathbf{D}) \\
&= \mathrm{tr}\left(\mathbf{U}\left[\frac{d}{d\epsilon}\Big|_{\epsilon=0} \phi(\boldsymbol{\Lambda} + \epsilon\mathbf{U}^H\mathbf{D}\mathbf{U})\right]\mathbf{U}^H\right) \\
&= \mathrm{tr}\left(\mathbf{U}D\phi(\boldsymbol{\Lambda})(\mathbf{U}^H\mathbf{D}\mathbf{U})\mathbf{U}^H\right) \\
&= \mathrm{tr}\left(\phi^{[1]} \circ \underbrace{\mathbf{U}^H\mathbf{D}\mathbf{U}}_{\mathbf{Z}}\right) \\
&= \mathrm{tr}\left(\sum_{k=1}^{n}\left[\phi^{[1]}(\boldsymbol{\Lambda})\right]_{k,k}\mathbf{Z}_{k,k}\right) = \mathrm{tr}\left(\sum_{k=1}^{n}\left[\phi'(\boldsymbol{\Lambda})\right]_{k,k}\mathbf{Z}_{k,k}\right) \\
&= \mathrm{tr}(\phi'(\boldsymbol{\Lambda})\mathbf{U}^H\mathbf{D}\mathbf{U}) = \mathrm{tr}(\phi'(\mathbf{A}) \cdot \mathbf{D})
\end{aligned}$$

Representation for Löwner's theory:

- Every matrix-monotone function ϕ can be expressed as

$$\phi(t) = a + bt + \int_0^\infty \frac{st}{s+t} d\mu(s)$$

with a positive measure $\mu \in [0, \infty)$ and real constants $a, b \geq 0$.

- Every matrix-convex function ψ can be represented as

$$\psi(t) = a + bt + ct^2 + \int_0^\infty \frac{st^2}{s+t} d\mu(s)$$

with a positive measure $\mu \in [0, \infty)$ and real constants $a, b, c \geq 0$.

- Every matrix monotone function can be represented by two scalars and a measure as $MM = (a, b, \mu(s))$.
- Matrix valued function can be represented as

$$\phi(\mathbf{A}) = a\mathbf{I} + b\mathbf{A} + \int_0^\infty s\mathbf{A}(s\mathbf{I} + \mathbf{A})^{-1} d\mu(s)$$

Examples

- $MM = (0, 0, \frac{1}{s^2}u(s-1))$ with step function $u(s)$

$$\phi(t) = \int_1^\infty \frac{st}{s+t}\frac{1}{s^2}ds = \log(1+t).$$

- $MM = (-1, 0, \frac{1}{s^3}u(s-1))$ leads to

$$\phi(t) = -1 + \int_1^\infty \frac{st}{s+t}\frac{1}{s^3}ds = -\frac{\log(1+t)}{t}.$$

- $MM = (0, 0, \frac{\sin(r\pi)}{\pi}s^{r-2})$ with $0 < r \leq 1$ yields

$$\phi(t) = \frac{\sin(r\pi)}{\pi}\int_0^\infty \frac{st}{s+t}s^{r-2}ds = t^r.$$

Derivative of matrix-monotone function:

- The first derivative of an arbitrary matrix-monotone function at $\mathbf{A}$ in direction $\mathbf{B}$ is given by

$$D\phi(\mathbf{A})(\mathbf{B}) = b\mathbf{B} + \int_0^\infty s^2[s\mathbf{I}+\mathbf{A}]^{-1}\mathbf{B}[s\mathbf{I}+\mathbf{A}]^{-1}d\mu(s).$$

Proof:

- The directional derivative is defined as $D\phi(\mathbf{A})(\mathbf{B}) = \frac{\partial \phi(\mathbf{A}+\epsilon\mathbf{B})}{\partial \epsilon}\Big|_{\epsilon=0}$

$$\phi(\mathbf{A}+\epsilon\mathbf{B}) = a\mathbf{I} + b(\mathbf{A}+\epsilon\mathbf{B}) + \int_0^\infty s(\mathbf{A}+\epsilon\mathbf{B})(s\mathbf{I}+\mathbf{A}+\epsilon\mathbf{B})^{-1}d\mu(s)$$

$$\frac{\partial \phi(\mathbf{A}+\epsilon\mathbf{B})}{\partial \epsilon}\Big|_{\epsilon=0} = b\mathbf{B} + \int_0^\infty [\mathbf{I} - \mathbf{A}[s\mathbf{I}+\mathbf{A}]^{-1}]s\mathbf{B}[s\mathbf{I}+\mathbf{A}]^{-1}d\mu(s)$$

- By simplifying, we get the result.

- For p.s.d. matrices $\mathbf{A}$ and $\mathbf{B}$ with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_n$, it holds

$$\min_{\pi} \operatorname{tr} \phi(\operatorname{diag}(\alpha_1, \ldots, \alpha_n)\operatorname{diag}(\beta_{\pi_1}, \ldots, \beta_{\pi_n})) \leq \operatorname{tr} \phi(\mathbf{B}^{\frac{1}{2}}\mathbf{A}\mathbf{B}^{\frac{1}{2}})$$
$$\leq \max_{\pi} \operatorname{tr} \phi(\operatorname{diag}(\alpha_1, \ldots, \alpha_n)\operatorname{diag}(\beta_{\pi_1}, \ldots, \beta_{\pi_n}))$$

with permutation π.

[1] Matrix norms covered in Matrix Theory course

Contraction:

- A matrix $\mathbf{C}$ is a contraction if $\mathbf{C}^H\mathbf{C} \leq \mathbf{I}$, or equivalently, $\|\mathbf{C}\|_\infty \leq 1$.
- Let ϕ be a matrix-monotone function on $[0, \infty)$, ψ a matrix convex function on $[0, \infty)$ with $\psi(0) \leq 0$. Then for every contraction $\mathbf{C}$ and every $\mathbf{A} \geq \mathbf{0}$,

$$\phi(\mathbf{C}^H\mathbf{A}\mathbf{C}) \geq \mathbf{C}^H\phi(\mathbf{A})\mathbf{C} \qquad \text{and} \qquad \psi(\mathbf{C}^H\mathbf{A}\mathbf{C}) \leq \mathbf{C}^H\psi(\mathbf{A})\mathbf{C}.$$

Connection:

- A binary operation σ on the class of positive definite matrices $(\mathbf{A}, \mathbf{B}) \rightarrow \mathbf{A}\sigma\mathbf{B}$, is a connection if the following requirements are fulfilled:
  - $\mathbf{A} \leq \mathbf{C}$ and $\mathbf{B} \leq \mathbf{D}$ imply $\mathbf{A}\sigma\mathbf{B} \leq \mathbf{C}\sigma\mathbf{D}$.
  - $\mathbf{C}(\mathbf{A}\sigma\mathbf{B})\mathbf{C} \leq (\mathbf{CAC})\sigma(\mathbf{CBC})$.
  - If a series $\mathbf{A}_n$ converges to $\mathbf{A}$ and a series $\mathbf{B}_n$ converges to $\mathbf{B}$, respectively, then the series $(\mathbf{A}_n\sigma\mathbf{B}_n)$ converges to $\mathbf{A}\sigma\mathbf{B}$.

Main Reference

- Chapter 3 of "Majorization and Matrix-Monotone Functions in Wireless Communications", Eduard Jorswieck, Holger Boche.

Other References

- R. Bhatia, "Matrix Analysis", Springer-Verlag, 1997.
- R. A. Horn and C. R. Johnson, "Matrix Analysis", Cambridge University Press, 1985.
- R. A. Horn and C. R. Johnson, "Topics in Matrix Analysis", Cambridge University Press, 1991.
- A. W. Marshall and I. Olkin, "Inequalities: Theory of Majorization and Its Application" in Mathematics in Science and Engineering, Academic Press, Inc. (London) Ltd., 1979.

THANK YOU!