Majorization Theory for Optimization in Wireless Communications

Hadi Ghauch

Royal Insititue of Technology

ghauch@kth.se

2012-06-11

Outline

1. History and Motivation
2. Definitions: Schur-convex / Schur-concave functions
3. Matrix inequalities
4. Stochastic Majorization: Basic Results
5. Applications
    - Wireless interference avoidance
    - Capacity of fading MAC channel

Brief history and motivation

History

- Finding inequalities has been a major mathematical research area, beginning with Gauss, Cauchy and others.
- Pure and applied mathematical analysis needs inequalities, e.g. triangle inequalities, integral and differential inequalities, etc...
- The building blocks of the theory are in [3].

Applications

Solve communication and information theoretic problems in wireless communications

- impact of spatial correlation in multiantenna systems
- average capacity of channel
- outage probability of channels

Definition

Definition: Majorization for vectors

For two vectors $\boldsymbol{x}$, $\boldsymbol{y} \in \mathcal{R}^n$ with components in descending order $x_1 \geq x_2 \geq ..., x_n \geq 0$ and $y_1 \geq y_2 \geq ... \geq y_n \geq 0$, one says that $\boldsymbol{x}$ *majorizes* $\boldsymbol{y}$, ($\boldsymbol{x} \succeq \boldsymbol{y}$) if,

$$\sum_{k=1}^{m} x_k \geq \sum_{k=1}^{m} y_k, \forall m = 1, 2, ...n-1 \textit{ and } \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k$$

Examples

- $(5, 4, 2, 0) \succeq (5, 4, 1, 1)$
- $(\frac{1}{n}, ..., \frac{1}{n}) \preceq (\frac{1}{n-1}, ..., \frac{1}{n-1}, 0) \preceq \; ... \; \preceq (\frac{1}{2}, \frac{1}{2}, 0, ..., 0) \preceq (1, 0, ..., 0)$

Majorization is a *partial order* relation, that describes how "less spread-out" are the components of two vectors, e.g. $\boldsymbol{x} = (1, 0, ..., 0) \succeq \boldsymbol{y} = (\frac{1}{n}, ..., \frac{1}{n})$

Schur-convex functions

The next logical step would be to characterize order-preserving functions, with the respect to the majorization order.

Definition: Schur-convex / Schur-concave functions

A real-valued function ϕ defined on a subset $\mathcal{A}$ of $\mathcal{R}^n$ is said to be Schur-convex on $\mathcal{A}$ if $\boldsymbol{x} \succeq \boldsymbol{y} \Rightarrow \phi(\boldsymbol{x}) \geq \phi(\boldsymbol{y})$. The definition of Schur-concave functions follows.

Let $\phi(\boldsymbol{x}) = \sum_k |x_k|^2$ is Schur-convex. *Proof:* consider two vectors on $\mathcal{R}^n$ that satisfy $\boldsymbol{x} \succeq \boldsymbol{y}$. It is easy to verify that $\phi(\boldsymbol{x}) \geq \phi(\boldsymbol{y})$.

Definition: Schur's condition

Let $f : \mathcal{I}^n \rightarrow \mathcal{R}, \mathcal{I} \subset \mathcal{R}$ be continuously differentiable. f is Schur-convex *iff* f is symmetric on $\mathcal{I}^n$ *and* $(x_1 - x_2)(\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2}) \geq 0$.

Matrix Inequalities

Majorization theory can also be used to prove some well-known matrix inequalities.

Schur's inequality

Let $\boldsymbol{Q}$ be an n x n Hermitian matrix. The vector of eigenvalues $\lambda[\boldsymbol{Q}]$ majorizes the vector of diagonal entries, $diag[\boldsymbol{Q}]$. We write $\lambda[\boldsymbol{Q}] \succeq diag[\boldsymbol{Q}]$, with equality when $\boldsymbol{Q}$ is diagonal.

Proof: relies on the following property: if $\boldsymbol{xP} = \boldsymbol{y}$ where $\boldsymbol{P}$ is a doubly stochastic matrix, then $\boldsymbol{x} \succeq \boldsymbol{y}$. By writing the eigendecomposition of $\boldsymbol{Q} = \boldsymbol{U\Lambda U}^H$, the proof follows.

Matrix Inequalities

Hadamard's inequality

Let $\boldsymbol{Q}$ be an n x n positive semi-definite matrix with $\boldsymbol{Q}_{11} \geq ... \geq \boldsymbol{Q}_{nn} \geq 0$ and $\lambda_1[\boldsymbol{Q}] \geq ... \geq \lambda_n[\boldsymbol{Q}] \geq 0$. We write $\prod_{k=l}^{n} \boldsymbol{Q}_{kk} \geq \prod_{k=l}^{n} \lambda_k[\boldsymbol{Q}]$, $\forall l = 1, ..., n$ with equality when $\boldsymbol{Q}$ is diagonal.

Proof: From prev theorem, $\lambda[\boldsymbol{Q}] \succeq diag[\boldsymbol{Q}]$. And since $\phi(\boldsymbol{x}) = \sum_i log(x_i)$ $= log(\prod_i x_i)$ is Schur-concave, then $\phi(\lambda[\boldsymbol{Q}]) \leq \phi(diag[\boldsymbol{Q})$. The property follows.

Sum of k-smallest eigenvalues

For a given positive semi-definite matrix $\boldsymbol{Q}$ with ordered eigenvalues, the sum of the largest / smallest k-eigenvalues is given by,

Partial trace

$$\max_{\boldsymbol{U}\boldsymbol{U}^H = I_k} \; tr(\boldsymbol{U}\boldsymbol{Q}\boldsymbol{U}^H) = \sum_{l=1}^{k} \lambda_l[\boldsymbol{Q}]$$

$$\min_{\boldsymbol{U}\boldsymbol{U}^H = I_k} \; tr(\boldsymbol{U}\boldsymbol{Q}\boldsymbol{U}^H) = \sum_{l=n-l+1}^{k} \lambda_l[\boldsymbol{Q}]$$

where $k = 1, ..., n$, and the max / min is taken over all k x n unitary matrices satisfying $\boldsymbol{U}\boldsymbol{U}^H = I_k$.

Proof: provided in [2], and based on the following: if $\boldsymbol{x}\boldsymbol{P} = \boldsymbol{y}$ where $\boldsymbol{P}$ is a doubly-stochastic matrix, then $\boldsymbol{x} \succeq \boldsymbol{y}$

Weyl's Inequalities

In the context of majorization theory, Weyl's inequalities can be stated in the following form

Weyl's inequality

Let $\boldsymbol{Q}$ and $\boldsymbol{P}$ be n x n positive semi-definite matrices with ordered eigenvalues in descending order. Then,

$$\begin{pmatrix} \lambda_1[\boldsymbol{P}] + \lambda_n[\boldsymbol{Q}] \\ \vdots \\ \lambda_n[\boldsymbol{P}] + \lambda_1[\boldsymbol{Q}] \end{pmatrix} \preceq \begin{pmatrix} \lambda_1[\boldsymbol{P} + \boldsymbol{Q}] \\ \vdots \\ \lambda_n[\boldsymbol{P} + \boldsymbol{Q}] \end{pmatrix} \preceq \begin{pmatrix} \lambda_1[\boldsymbol{P}] + \lambda_1[\boldsymbol{Q}] \\ \vdots \\ \lambda_n[\boldsymbol{P}] + \lambda_n[\boldsymbol{Q}] \end{pmatrix}$$

Corollary (Weyl's first inequality): $\lambda_1[\boldsymbol{P} + \boldsymbol{Q}] \leq \lambda_1[\boldsymbol{P}] + \lambda_1[\boldsymbol{Q}]$.

Proof: uses the trace property form prev slide (refer to [2]).

Application: Interference avoidance

$\boldsymbol{Q}$ is the covariance matrix of the *interference signal*, seen by some multiantenna receiver. We need to find the "optimal" linear filter $\boldsymbol{U}$ - a d x n unitary matrix that is applied to the received signal, and that minimizes the resulting interference.

After applying the linear filter U, the resulting interference is given by $tr(\boldsymbol{UQU}^H)$. Thus the we need to solve the following optimization problem: $\underset{\boldsymbol{UU}^H = I_d}{min} \quad tr(\boldsymbol{UQU}^H)$. Using the result of the prev slide, the latter quantity is the sum of the d-smallest eigenvalues of Q, i.e. $\sum_{l=n-l+1}^{d} \lambda_l[\boldsymbol{Q}]$. Moreover, this is achieve by picking the rows of $\boldsymbol{U}$ as the corresponding eigenvectors.

Basic Results - [1/2]

Let $w_1, ..., w_n$ be i.i.d random variables with a given pdf, and $\boldsymbol{\mu}$ have non-negative non-decreasing entries, $\mu_1 \geq \mu_2 \geq ... \geq \mu_n \geq 0$.

Definition: Average of a function of weighted sum

Suppose the function $f : \mathcal{R}_+ \rightarrow \mathcal{R}_+$ is concave. Then $G(\boldsymbol{\mu}) = E_{w_1,...,w_n}[f(\sum_{k=1}^n \mu_k w_k)]$ is Schur-concave in $\boldsymbol{\mu}$. If f is convex, G is Schur-convex.

Proof: It is clear that $f(\sum_{k=1}^n \mu_k w_k)$ is concave because f is assumed to be concave. Moreover, it is obvious that f is symmetric since the elements of $\boldsymbol{\mu}$ can be permuted without changing its value. The above observations ensure that Schur's condition is satisfied. It follows that G is Schur-concave.

Let $w_1, \ldots, w_n$ be i.i.d random variables with a given pdf, and $\boldsymbol{\mu}$ have non-negative non-decreasing entries, $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n \geq 0$.

Definition: Average of a function of weighted sum

Suppose the function $f : \mathcal{R}_+ \to \mathcal{R}_+$ is concave. Then $G(\boldsymbol{\mu}) = E_{w_1,\ldots,w_n}[f(\sum_{k=1}^{n} \mu_k w_k)]$ is Schur-concave in $\boldsymbol{\mu}$. If f is convex, G is Schur-convex.

Proof: It is clear that $f(\sum_{k=1}^{n} \mu_k w_k)$ is concave because f is assumed to be concave. Moreover, it is obvious that f is symmetric since the elements of $\boldsymbol{\mu}$ can be permuted without changing its value. The above observations ensure that Schur's condition is satisfied. It follows that G is Schur-concave.

Application - [2/2]

$G(\boldsymbol{\mu}) = E_{w_1,\ldots,w_n}[f(\sum_{k=1}^{n} \mu_k w_k)]$ is Schur-concave in $\boldsymbol{\mu}$, for f concave.

Application: Capacity of Fast Fading Uplink with no CSIT

n-users transmit data to the base station (each with max power P, for a total of nP). The random channel gain for link k is $h_k \sim \mathcal{CN}(0,1)$, and is unknown to transmitter k. What is the average communication rate (capacity), for this channel? The averaging is done over the joint distributions of $|h_1|^2, \ldots, |h_n|^2$.

We can think of μ_k as the fraction of channel resources allocated to user k, i.e. power, time, bandwidth. Furthermore, replacing f by $log(1+x)$ - a concave function, the average channel capacity is given by $C(\boldsymbol{\mu}) = E_{h_1,\ldots,h_n}[log(1 + \sum_{k=1}^{n} \mu_k |h_k|^2)]$.

Application - [2/2]

Since the transmitter has no knowledge of the channel, the total channel resources, nP, have to be divided equally among the users. All possibles values for $\boldsymbol{\mu}$ satisfy:

$$\boldsymbol{\mu}^{[1]} = (P, ..., P) \preceq (\frac{nP}{n-1}, \frac{nP}{n-1}, ..., 0) \preceq ... \preceq \boldsymbol{\mu} \preceq$$
$$... \preceq (\frac{nP}{2}, \frac{nP}{2}, 0, ..., 0) \preceq (nP, 0, ..., 0) = \boldsymbol{\mu}^{[2]}$$

where $\boldsymbol{\mu}^{[1]}$ and $\boldsymbol{\mu}^{[2]}$ are the extreme ends: resources are divided equally among all users, or are all allocated to one user. Obviously, one of the latter is optimal.

$C(\boldsymbol{\mu})$ is Schur-concave in $\boldsymbol{\mu}$ (previous slide), and since $\boldsymbol{\mu}^{[2]} \succeq \boldsymbol{\mu}^{[1]}$ it follows that $C(\boldsymbol{\mu}^{[2]}) \leq C(\boldsymbol{\mu}^{[1]})$. Thus, equal power allocation is optimal. Then, the uplink channel capacity is given by $C = E[log(1 + \sum_{k=1}^{n} P|h_k|^2)]$

References

E. Jorswrieck, H. Boche (2007)
Majorization and matrix monotone functions, in wireless communications
Foundations and trends in Communication and Information Technology 3(6), 555 – 584.

A. W. Marshall, I. Olkin (1979)
Inequalities: The theory of majorization and its applications
Mathematics in Science and Engineering, Academic press, London

G. Hardy, J.E. Littlewood, G. Polya (1952)
Inequalities
Cambridge Mathematical Libreary

The End