# Interference Cancelation and Information Processing in Communication Systems

## Majorization Theory for Wireless Systems

**Charles Casimiro Cavalcante**
(in cooperation with A. A. P. Guimarães)

Wireless Telecommunications Research Group – GTEL
Federal University of Ceará – UFC
Fortaleza-CE, Brazil

http://www.gtel.ufc.br

charles@gtel.ufc.br

July 2nd, 2009

# About GTEL...

- Wireless Telecommunications Reseacrh Group (GTEL)
- Lab. of research on several aspects of telecommunications
- Part of the Federal University of Ceará (UFC) and the Departament of Teleinformatics Engineering (DETI)
- Several funded projects (national agencies and industry)
- Ericsson Research is the major partner, so far

# Research Lines in GTEL

1. Signal processing for communications
2. Physical layer optimization
3. Resource allocation and management
4. System and link level performance evaluation

# People at GTEL

1. 6 faculty members
2. 2 post-doc researchers
3. 11 PhD students
4. 12 MSc students
5. 15 trainees (undergraduate students)
6. 6 administrative and support people

# Some numbers...

- Since 2000 - more than 20 funded projects
- A private area of around 450m2
- Computational cluster: the second biggest in the universities in Brazil
- Around US$ 6 millions in research funding
- Several partner in academia and industry
- Still open for partnership!

# My main interests

- Signal processing problems: mainly for communications
- Physical and cross-layer optimization
- Blind source separation (mostly in the past)
- Mathematical tools for signal processing
- Interference management

# Motivation and historical aspects

- Hadamard determinant theorem.

# Motivation and historical aspects

- Hadamard determinant theorem.

**Teorema**

If $A = [a_{ij}]_{n\times n}$ is a positive-defined matrix, then:

$$\det A \leq \prod_{i=1}^{n} a_{ii}$$

# Motivation and historical aspects (cont.)

- Schur theorem (1923).

# Motivation and historical aspects (cont.)

- Schur theorem (1923).

**Theorem**

Let $A = [a_{ij}]_{n\times n}$ be an hermitian semidefined positive matrix, whose eigenvalues are $\lambda_1, \ldots, \lambda_n$. If $a_{11} \geq a_{22} \geq \cdots \geq a_{nn}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then

$$\sum_{i=1}^{k} a_{ii} \leq \sum_{i=1}^{k} \lambda_i \qquad 1 \leq k \leq n$$
$$\sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$$

# Motivation and historical aspects (cont.)

- Inequalities theory $\longrightarrow$ Majorization theory.

# Motivation and historical aspects (cont.)

- Inequalities theory $\longrightarrow$ Majorization theory.
- Notation and definition that a vector is majorized by another one was first introduced in the book *Inqualities* de Hardy, Littlewood, and Pólya (1934)

# Motivation and historical aspects (cont.)

- Inequalities theory $\rightarrow$ Majorization theory.
- Notation and definition that a vector is majorized by another one was first introduced in the book *Inqualities* de Hardy, Littlewood, and Pólya (1934)
- Reference: Inequalities: Theory of Majorization and Its Applications (Olkin and Marshall, 1979).

## Notation

Ordering the coordinates of the vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ in the following descendent order

$$x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}$$

we obtain the vector $[\mathbf{x}]$ of $\mathbb{R}^n$ defined by

$$[\mathbf{x}] = (x_{[1]}, x_{[2]}, \cdots, x_{[n]}).$$

## Example

If $\mathbf{x} = (-1, 2, 0, 4)$, then $[\mathbf{x}] = (4, 2, 0, -1)$.

# Majorization of vectors

## definition (Majorization)

Given the vectors $\mathbf{x},\ \mathbf{y} \in \mathbb{R}^n$, we say that $\mathbf{x}$ is majorized by $\mathbf{y}$ or that $\mathbf{y}$ majorizes $\mathbf{x}$ when

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \qquad 1 \leq k \leq n$$

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$$

## Notation

When $\mathbf{x}$ is majorized by $\mathbf{y}$ we write $\mathbf{x} \preceq \mathbf{y}$ or $\mathbf{y} \succeq \mathbf{x}$.

# Majorization of vectors (cont.)

## Definition (Weak majorization)

Given the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we say that $\mathbf{x}$ is weakly majorized by $\mathbf{y}$ or that $\mathbf{y}$ weakly majorizes $\mathbf{x}$ if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \qquad 1 \leq k \leq n$$

## Notation

When $\mathbf{x}$ is weakly majorized by $\mathbf{y}$ we write $\mathbf{x} \preceq_w \mathbf{y}$ or $\mathbf{y} \succeq_w \mathbf{x}$.

# Majorization of vectors (cont.)

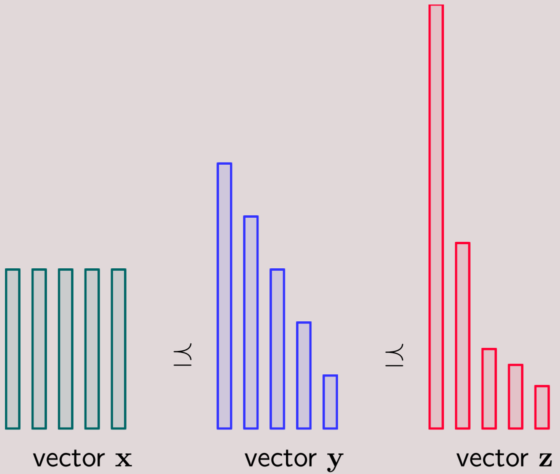
**Example**

Considering the following vectors
$\mathbf{x} = (3;\ 3;\ 3;\ 3;\ 3)$, $\mathbf{y} = (5;\ 4;\ 3;\ 2;\ 1)$ and
$\mathbf{z} = (8;\ 3,5;\ 1,5;\ 1,2;\ 0,8)$, we have

$$\mathbf{x} \preceq \mathbf{y} \preceq \mathbf{z}.$$

# Majorization of vectors (cont.)

## Geometric interpretation



# Majorization of vectors (cont.)

## Other examples

- $\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) \preceq (1, 0, 0, \cdots, 0)$.
- $\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) \preceq (a_1, \ldots, a_n) \preceq (1, 0, 0, \cdots, 0)$, for all vector $\mathbf{a} \in \mathbb{R}^n$ such that $a_i \geq 0$ and $\sum_{i=1}^{n} a_i = 1$.
- $\frac{X}{n}\mathbf{1} \preceq \mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^n$ such that $x_i \geq 0$ and $\sum_{i=1}^{n} x_i = X$.

# Majorization of vectors (cont.)

## Demonstration...

$$
\begin{aligned}
\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) &\preceq \left(\frac{1}{n-1}, \cdots, \frac{1}{n-1}, 0\right) \\
&\preceq \left(\frac{1}{n-2}, \cdots, \frac{1}{n-2}, 0, 0\right) \preceq \ldots \\
&\preceq \left(\frac{1}{2}, \frac{1}{2}, 0, \cdots, 0\right) \\
&\preceq (1, 0, 0, \cdots, 0).
\end{aligned}
$$

# Majorization of vectors (cont.)

## Properties

1. If $\mathbf{x} \preceq \mathbf{y}$, then $k\mathbf{x} \preceq k\mathbf{y}$, for $k \in \mathbb{R}_+$.
2. If $\mathbf{x} \preceq \mathbf{y}$, then $\mathbf{x} \preceq_w \mathbf{y}$.
3. $\mathbf{x} \preceq \mathbf{y}$ if and only if, $\sum_{i=1}^{n} |x_i - t| \leq \sum_{i=1}^{n} |y_i - t|$, for all $t \in \mathbb{R}$.
4. If $\mathbf{x} \preceq \mathbf{y}$, then $\min\{x_i\} \geq \min\{y_i\}$.
5. If $\mathbf{x} \preceq \mathbf{1}$, then $\mathbf{x} = \mathbf{1}$.
6. $\mathbf{x} \preceq \mathbf{y}$ if and only if, $\sum_{i=1}^{n} \varphi(x_i) \leq \sum_{i=1}^{n} \varphi(y_i)$, for all convex function $\varphi : \mathbb{R} \to \mathbb{R}$.

# Definition

## Definition (Doubly stochastic matrix)

A matrix $A = [a_{ij}]_{n \times n}$ is said to be doubly stochastic when it holds the following properties:

- $a_{ij} \geq 0, \ 1 \leq i \leq n, \ \ 1 \leq j \leq n.$
- $\displaystyle\sum_{i=1}^{n} a_{ij} = 1, \quad 1 \leq j \leq n.$
- $\displaystyle\sum_{j=1}^{n} a_{ij} = 1, \quad 1 \leq i \leq n.$

# Main properties

## Properties

Be $A = [a_{ij}]_{n \times n}$ a doubly stochastic matriz, it follows that :

1. The product of two doubly stochastic matrices is also a doubly stochastic matrix.
2. $A\mathbf{1} = \mathbf{1}$.
3. $\lambda = 1$ is an eigenvalue of $A$.
4. If $\lambda$ is an eigenvalue of $A$, then $|\lambda| \leq 1$.
5. $||A|| = 1$.

# Majorization theory × Doubly stochastic matrices

**Theorem**

A matrix $A$ is doubly stochastic if and only if $A\mathbf{x} \preceq \mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^n$.

**Theorem (Hardy, Littlewood, and Pólya, 1929)**

Be $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the following conditions are equivalent:

(a) $\mathbf{x} \preceq \mathbf{y}$.

(b) $\mathbf{x} = A\mathbf{y}$ for some doubly stochastic matrix $A$.

# Application of theorems

## Example

Let $B = [b_{ij}]_{n \times n}$ be a doubly stochastic matrix, $\mathbf{d}(B) = \{b_1, \ldots, b_n\}$ is the main diagonal of $B$ and $\lambda(B) = \{\lambda_1, \ldots, \lambda_n\}$ the eigenvalues of $B$. Then

$$\mathbf{d}(B) \preceq \lambda(B)$$

# Application of theorems (cont.)

## Proof

Since $B$ is hermitian ($B = B^H$), then a unitary matrix $U = [u_{ij}]$ such that

$$B = UD_\lambda U^H$$

where $D_\lambda = \text{diag}\,\{\lambda_1, \ldots, \lambda_n\}$ is the diagonal matrix whose entries are the eigenvalues of $B$. By multiplying we have

$$b_i = \sum_{j=1}^{n} p_{ij}\lambda_j \quad 1 \leq i \leq n$$

where $p_{ij} = |u_{ij}|^2$. Since the matrix $P = [p_{ij}]$ is doubly stochastic and $\mathbf{d}(B) = P \cdot \lambda(B)$, it follows fro the theorem of Hardy, Littlewood, and Polya, that $\mathbf{d}(B) \preceq \lambda(B)$.

# Definition

**Definition (Schur-convex function)**

Let $\phi$ be a real function defined in $A \subset \mathbb{R}^n$. The function $\phi$ is said to be Schur-convex over $A$ if

$$\mathbf{x} \preceq \mathbf{y} \quad \text{in} \quad A, \quad \text{then} \quad \phi(\mathbf{x}) \leq \phi(\mathbf{y}).$$

When $\phi(\mathbf{x}) < \phi(\mathbf{y})$, we say that $\phi$ is strictly Schur-convex.

**Definition (Schur-concave function)**

Let $\phi$ be a real function defined in $A \subset \mathbb{R}^n$. The function $\phi$ is said to be Schur-concave over $A$ if $-\phi$ is Schur-convex.

# Some results

**Theorem**

Be $f : \mathbb{R}^n \to R$ a Schur-convex function which possess only one critical point $\mathbf{a} = a_1, \ldots, a_n$ . If $f(\mathbf{a})$ is a local minimum point then this point is also global minimum.

**Theorem (Schur condition)**

Let $\mathcal{I} \subset \mathbb{R}$ be an open interval and be $f : \mathcal{I}^n \to \mathbb{R}$ a differentiable function. The function $f$ is Schur-convex over $\mathcal{I}^n$ if and only if $f$ is symmetric over $\mathcal{I}^n$ and

$$(x_i - x_j)\left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}\right) \geq 0, \text{for all} 1 \leq i, j \leq n.$$

# Examples

## Examples of Shur-convex functions

- The function $\phi : \mathbb{R}^n \to \mathbb{R}$, defined by $\phi(\mathbf{x}) = |x_i|$.
- The entropy function $H(\mathbf{x}) = -\sum_i x_i \log x_i$, where $\mathbf{x} \in \mathbb{R}^n_+$.
- The variance function $V(\mathbf{x}) = \dfrac{1}{n}\sum_i^n (x_i - \bar{x})^2$, where

  $\bar{x} = \dfrac{1}{n}\sum_i^n x_i$ e $\mathbf{x} \in \mathbb{R}^n$.

# Example

## Observation

A Shcur-convex function is not necessarily convex.

## Example

Let be the interval $\mathcal{I} = (0, 1)$. The function $\phi : \mathcal{I}^2 \to \mathbb{R}$ defined by:

$$\phi(x, y) = \log\left(\frac{1}{x} - 1\right) + \log\left(\frac{1}{y} + 1\right).$$

is Schur-convex but is not convex.

# An optimization problem

Let the Schur-concave function $f : \mathbb{R}_+^n \to \mathbb{R}_+$ and let us consider the following optimization problems:

1. $\max f(\mathbf{x})$, subject to $||\mathbf{x}||_{l_1} = X$.
2. $\min f(\mathbf{x})$, subject to $||\mathbf{x}||_{l_1} = X$.

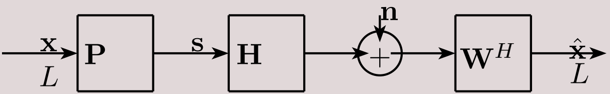
The point $\mathbf{x}_1 = \frac{X}{n}\mathbf{1}$ is solution for the maximization problem while $\mathbf{x}_2 = (X, 0, \cdots, 0)$ is solution for the later one. Indeed, we know that for all $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathbf{x}_1 = \frac{X}{n}\mathbf{1} \preceq \mathbf{x} \preceq (X, 0, \cdots, 0) = \mathbf{x}_2.$$

Since $f$ is Schur-concave it follows that $f(\mathbf{x}_1) \geq f(\mathbf{x}) \geq f(\mathbf{x}_2)$.

# Mean Square Error minimization

## System model




- Data: $\mathbf{x}$ ($L$ symbols)
- Precoder: $\mathbf{P}$
- Transmitted signal: $\mathbf{s} = \mathbf{P}\mathbf{x}$
- Power: $P_T = \mathrm{Tr}(\mathbf{P}\mathbf{P}^H)$
- Linear receiver: $\mathbf{W}$
- Estimated signal: $\hat{\mathbf{x}} = \mathbf{W}^H\mathbf{y}$

# Mean Square Error minimization (cont.)

## Wiener filter

$$\mathbf{W} = (\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H + \mathbf{R}_n)^{-1}\mathbf{H}\mathbf{P}$$

## MSE matrix

$\mathbf{E} = (\mathbf{I} + \mathbf{P}^H\mathbf{R}_H\mathbf{P})^{-1}$ where, $\mathbf{R}_H = \mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}$.

## Nonlinear problem

$$\begin{aligned} \text{minimize} \quad & f(\xi) \\ \text{subject to} \quad & \xi = \mathbf{d}((\mathbf{I} + \mathbf{P}^H\mathbf{R}_H\mathbf{P})^{-1}) \\ & \mathrm{Tr}(\mathbf{P}\mathbf{P}^H) \le P_T \end{aligned}$$

# Mean Square Error minimization (cont.)

## Objective

$$\begin{array}{ll} \text{minimize} & f(\xi) \\ \text{subject to} & (\xi_1, \ldots, \xi_L) \preceq (\alpha_1, \ldots, \alpha_n) \\ & \mathbf{1}^T \mathbf{p} \leq P_T, \quad \mathbf{p} \geq \mathbf{0} \end{array}$$

where

$$\alpha_i = \frac{1}{1 + \sigma_i^2 \lambda_{H,i}} \quad 1 \leq i \leq L.$$

and $\sigma_i^2$ is the allocated power.

# References


- E. Jorswieck and H. Boche
  Majorization and Matrix-Monotone Functions in Wireless Communications
  Foundations and Trends in Communications and Information Theory, vol. 3, no. 6, pp. 553-701, 2006.

- D. P. Palomar and Y. Jiang
  *MIMO transceiver design via majorization theory*
  Foundations and Trends in Communications and Information Theory, vol. 3, no. 4-5, pp. 331-551, 2007.

- X.-M. Zhang,
  *Optimization of Schur-convex functions*,
  Mathematical Inequalities and Applications, vol. 1, no. 3, pp. 319-330, 1998.

# Referências


A. W. Marshall AND I. Olkin,
*Inequalities: Theory of Majorization and Its Applications*,
Academic Press, Inc., 1979.

G. H. Hardy, J.E. Littlewood, and G. Polya,
*Inequalities*,
Cambridge University Press , 1934.