Majorization and Matrix-Monotone Functions in Wireless Communications

Eduard A. Jorswieck

Signal Processing,
School of Electrical Engineering
Royal Institute of Technology
Osquldas väg 10, 100 44 Stockholm, Sweden

Fall 2007
5th Lecture

Majorization - Applications II
User distributions

- In the signal model, there are $K$ mobile users who are going to receive data from one base station.
- The single-antenna quasi-static block flat-fading channels $h_1, ..., h_K$ between the mobiles and the base are modeled as constant for a block of coherence length $T$ and from block to block as zero-mean independent complex Gaussian distributed (CN$(0, c_k)$).
- The variance is $c_i = \mathbb{E}(h_i^*h_i)$ for $1 \leq i \leq K$.
- In order to guarantee a fair comparison, we constrain the sum variance to be equal to the number of users, i.e. $\sum_{k=1}^{K} c_k = K$.
- **Definition 1. [More spread out user distribution]** A user distribution $c^1$ is called more spread out than a user distribution $c^2$ if the fading variance vector $c^1$ is majorized by $c^2$, i.e. $c^1 \preceq c^2$. 
Symmetric scenario $\mathbf{c}^1 = (1, 1, \ldots, 1)$
One mobile moves to the base and another to the cell edge. The sum of their fading variances stays constant. $c^2 = (1 + \alpha, 1, 1, ..., 1 - \alpha)$
All but one mobile at the cell edge $c^3 = (8, 0, \ldots, 0)$
In the figures, the symmetric scenario is majorized by all other scenarios and scenario three majorizes all other scenarios i.e.

\[ c^1 \preceq c^2 \preceq c^3. \]

Note, that the measure of user distribution and the measure of spatial correlation can be combined for e.g. multiuser MIMO systems. The channel of a user \( k \) can be modelled for all \( 1 \leq k \leq K \) under the Kronecker model assumption as

\[
H_k = c_kR^{1/2}_{R,k}W_kR^{1/2}_{T,k}
\]

with normalized transmit and receive correlation matrix as well as normalized random matrix, i.e. for all \( 1 \leq k \leq K \) \( \text{tr} R_{T,k} = n_T \), \( \text{tr} R_{R,k} = n_R \). Then, the long-term fading is captured by \( c_k \), the spatial correlation by \( R_T \) and \( R_R \) and the rich multi-path environment by \( W \).
Lemma 1. The sum rate with perfect CSI at the base station is achieved by TDMA. The optimal power allocation is to transmit into direction of the best user \(l\) with \(\|h_l\|^2 > \|h_k\|^2\) for all \(1 \leq k \leq K\) and \(l \neq k\). The ergodic sum capacity is then given by

\[
C_{pCSI}(\rho, c) = \mathbb{E} \left( \log \left[ 1 + \rho \max \left( \|h_1\|^2, ..., \|h_K\|^2 \right) \right] \right). \tag{1}
\]

The optimal transmit strategy to achieve the average sum capacity with long-term CSI is TDMA. Only the user with highest channel variance \(c_k\) is allowed to transmit. The achievable average sum capacity is given by

\[
C_{cCSI}(\rho, c) = \mathbb{E} \log(1 + \rho c_1 w_1). \tag{2}
\]

For no CSI at the base, the most robust transmit strategy against worst case user distribution is equal power allocation and the ergodic sum rate\(^1\)

\[^1\text{We do not allow time division within a block.}\]
is given by

\[ C_{\text{noCSI}}(\rho, c) = \mathbb{E} \log \left( 1 + \rho \sum_{k=1}^{K} c_k w_k \right). \]  

\( (3) \)

\textbf{Theorem 1.} Assume perfect CSI at the mobiles. For perfect CSI at the base, the ergodic sum capacity in (1) is a Schur-convex function w.r.t. the fading variance vector \( c \). For a base which knows the fading variances, the ergodic sum capacity in (2) is a Schur-convex function w.r.t. the fading variance vector \( c \). For an uninformed base station, the ergodic sum rate in (3) is a Schur-concave function w.r.t. the fading variance vector \( c \).
Consider TDMA-based opportunistic scheduling first. Only a single user is scheduled.

\[ C_T(\rho) = \mathbb{E} \left[ \log \left( 1 + \rho \max_{1 \leq k \leq K} \max_{1 \leq i \leq n_T} |h_k^H w_i|^2 \right) \right] \]  \hspace{1cm} (4)  

\[ = \int_0^\infty \log(1 + \rho x)p(x)dx \]  \hspace{1cm} (5)  

with pdf \( p(x) \) of the random variable \( \max_{1 \leq k \leq K} \max_{1 \leq i \leq n_T} c_k \lambda_i s_{k,i} \). The CDF is given as

\[ P(x) = \prod_{k=1}^{K} \prod_{n=1}^{n_T} \left( 1 - \exp \left( -\frac{x}{c_k \lambda_n} \right) \right). \]
For convenience, we collect all numbers $\eta_{k+(n-1)K} = c_k \lambda_n$ for $1 \leq k \leq K$ and $1 \leq n \leq n_T$ in one large vector $\eta = [\eta_1, \ldots, \eta_{Kn_T}]$ and parameterize the throughput by $\eta$.

By partial integration the average throughput can be written as

$$C_T(\rho, \eta) = \left| \log(1 + \rho x) [P(x) - 1] \right|_0^\infty - \int_0^\infty \frac{\rho}{1 + \rho x} [P(x) - 1] \, dx$$

$$= \int_0^\infty \frac{\rho}{1 + \rho x} \left[ 1 - \prod_{k=1}^{Kn_T} \left( 1 - \exp \left( -\frac{x}{\eta_k} \right) \right) \right] \, dx. \tag{6}$$

**Theorem 2.** The function $C_T(\rho, \eta)$ is Schur-convex with respect to $\eta$, i.e., $\eta_1 \succeq \eta_2 \Rightarrow C_T(\rho, \eta_1) \geq C_T(\rho, \eta_2)$. 
Average sum rate in bits per channel use (bit/cu) vs. SNR for TDMA-OB: Best-case, worst-case, and intermediate scenario. The high-SNR power difference is 3.2 dB.
In SDMA based opportunistic beamforming: The resulting average throughput can be upper bounded by [Sharif2005]

\[
C(\rho, \mathbf{c}, \mathbf{R}) = n_T \mathbb{E} \log \left( 1 + \max_{1 \leq k \leq K} \text{SINR}_k \right)
\]  

(7)

with \( \text{SINR}_k = \frac{||w_i^H \mathbf{h}_k||^2}{\sigma_n^2 + \sum_{l \neq i} ||w_i^H \mathbf{h}_k||^2} \).

**Theorem 3.** The average throughput of SDMA-OB is upper limited by

\[
B(K, n_T, 1) = \lim_{\rho \to \infty} C(\rho, 1, \mathbf{I}) = \frac{(\Psi(K + 1) + \gamma)n_T}{(n_T - 1) \log(2)}
\]

\[
= \frac{n_T}{(n_T - 1) \log(2)} \sum_{k=1}^{K} \frac{1}{k}
\]

(8)
which is tight for asymptotically high SNR and $\Psi(x)$ is the Psi-function.

**Theorem 4.** The upper bound on the average sum capacity of SDMA-OB $B(K, n_T, c)$ is Schur-concave with respect to $c$, i.e., $c_1 \succeq c_2 \Rightarrow B(K, n_T, c_1) \leq B(K, n_T, c_2)$. The upper bound on the average sum rate of SDMA-OB is largest for the balanced scenario $B(K, n_T, 1)$ and smallest for the completely unequally distributed scenario $B(1, n_T) = B(K, n_T, [K, 0, ..., 0])$. 
Different types of scheduling:

- The first scheduler is called maximum throughput scheduler (MTS) and the achievable average sum rate is given by

\[
R_{sum}^{MT} = \mathbb{E} \left[ \log \left( 1 + \rho \max_{1 \leq k \leq K} ||h_k||^2 \right) \right].
\]  \hspace{1cm} (9)

- The average sum rate of the round-robin scheduler (RRS) is given by

\[
R_{sum}^{RR} = \mathbb{E} \left[ \frac{1}{K} \sum_{k=1}^{K} \log \left( 1 + \rho ||h_k||^2 \right) \right].
\]  \hspace{1cm} (10)

- The proportional fair scheduler (PFS) has achievable sum rate given
by

\[ R_{sum}^{PF} = \mathbb{E} \left[ \log \left( 1 + \rho ||h_{k^*}||^2 \right) \right] \text{ with } \\

k^* = \arg \max_{1 \leq k \leq K} \frac{||h_k||^2}{c_k}. \tag{11} \]

Note that (11) can be rewritten as

\[
R_{sum}^{PF} = \frac{1}{K} \sum_{k=1}^{K} \mathbb{E} \left[ \log \left( 1 + \rho c_k \max_{1 \leq l \leq K} w_l \right) \right] \tag{12}
\]

\[
= \frac{1}{K} \sum_{k=1}^{K} \sum_{l=1}^{K} (-1)^{l-1} \binom{K}{l} \text{Ei} \left( 1, \frac{l}{\rho c_k} \right) e^{\frac{l}{\rho c_k}}
\]

because the scheduling probability of all users is equal to \( \frac{1}{K} \).
The sum rate performance of the opportunistic round robin scheduler (ORS) is given by

$$R_{sum}^{OR} = \frac{1}{K^2} \sum_{n=1}^{K} n \sum_{i=1}^{K} \sum_{j=0}^{n-1} \left( \begin{array}{c} n - 1 \\ j \end{array} \right) (-1)^j \left( \frac{1 + j}{c_i} \right) Ei \left( 1, \frac{1 + j}{c_i} \right).$$  \hspace{1cm} (13)

**Lemma 2.** The average sum rate of the ORS (13) can be written as

$$\int_{0}^{\infty} \left[ 1 - \frac{1}{K^2} \sum_{n=1}^{K} \sum_{i=1}^{K} \left( 1 - e^{-\frac{t}{c_i}} \right)^n \right] \frac{\rho}{1 + \rho t} dt.$$

**Lemma 2.** The average sum rate of the ORS (13) can be written as

$$\int_{0}^{\infty} \left[ 1 - \frac{1}{K^2} \sum_{n=1}^{K} \sum_{i=1}^{K} \left( 1 - e^{-\frac{t}{c_i}} \right)^n \right] \frac{\rho}{1 + \rho t} dt.$$

**Analysis of the sum rate performance**
**Theorem 5.** The average sum rate of the MTS is Schur-convex with respect to the vector of average user powers $c$, i.e.

$$ c \succeq d \implies R_{sum}^{MT}(c) \geq R_{sum}^{MT}(d). $$

**Theorem 6.** The average sum rate of the RRS is Schur-concave with respect to the vector of average user powers $c$, i.e.

$$ c \succeq d \implies R_{sum}^{RR}(c) \leq R_{sum}^{RR}(d). $$

**Theorem 7.** The average sum rate of the PFS is Schur-concave with respect to the vector of average user powers $c$, i.e.

$$ c \succeq d \implies R_{sum}^{PF}(c) \leq R_{sum}^{PF}(d). $$

**Theorem 8.** The average sum rate of the ORS is Schur-concave
with respect to the vector of average user power $c$, i.e.

$$c \succeq d \implies R_{sum}^{OR}(c) \leq R_{sum}^{OR}(d). \quad (18)$$

Fairness analysis: The average worst-case delay $\mathbb{E}[D_{m,K}]$ measures the average number of transmissions that are needed until all $K$ users have been active at least $m$ times. We define $D_1 = \mathbb{E}[D_{1,K}]$.

Average worst-case delay of the deterministic schedulers

$$D^{RRS}_1 = D^{ORS}_1 = K. \quad (19)$$

Average worst-case of PFS

$$D^{PFS}_1 = K \int_0^\infty 1 - (1 - \exp(-x))^K dt \quad (20)$$
Note that (20) can be written as

\[ D_{1}^{PFS} = K (\Psi(K + 1) + \gamma) \]  

(21)

with the \( \Psi \)-function and Euler’s constant \( \gamma \).

\[
\begin{align*}
\text{Average worst-case of MTS} \\
D_{1}^{MTS} &= n \int_{0}^{\infty} \left( 1 - \prod_{k=1}^{K} \left( 1 - \frac{\Gamma(m, d_{k}t)}{\Gamma(m)} \right) \right) dt.
\end{align*}
\]  

(22)

\[
\begin{align*}
\text{Theorem 9.} & \quad \text{The average worst-case delay } \mathbb{E}[D_{1,K}] \text{ of MTS is Schur-convex with respect to } d, \text{ i.e.} \\
d_{1} \succeq d_{2} & \quad \implies D_{1}^{MTS}(d_{1}) \geq D_{1}^{MTS}(d_{2}).
\end{align*}
\]  

(23)
Lemma 3. The mapping from the vector of user distributions to the vector of service probabilities is not order preserving with respect to the partial order Majorization.

Scaling with number of users:

Lemma 4. For symmetrically distributed users $c = 1$, the average sum rate of MTS, PFS, and ORS scale for large $K$ with $\log(K)$, i.e.

$$\lim_{K \to \infty} \frac{R_{\text{sum}}^{MT}(K)}{\log(K)} = \lim_{K \to \infty} \frac{R_{\text{sum}}^{PF}(K)}{\log(K)} = \lim_{K \to \infty} \frac{R_{\text{sum}}^{OR}(K)}{\log(K)} = 1.$$ (24)
Lemma 5. For symmetrically distributed users, the average worst-case delay scales linearly with $K$ for RRS and ORS. For MTS and PFS, it scales as $K \log(K)$, i.e.

$$\lim_{K \to \infty} \frac{D_{1}^{RRS}(K)}{K} = \lim_{K \to \infty} \frac{D_{1}^{ORS}(K)}{K} = 1$$

$$\lim_{K \to \infty} \frac{D_{1}^{MTS}(K)}{K \log(K)} = \lim_{K \to \infty} \frac{D_{1}^{PFS}(K)}{K \log(K)} = 1.$$  \hspace{1cm} (25)

Illustrations:
Observation and main conclusion: Therefore, taking the tradeoff between fairness and average sum rate into account, the PFS and ORS perform reasonable well. PFD is advantageous in symmetric scenarios whereas ORS performs better in asymmetric scenarios.
References I


ii) E. Jorswieck, A. Sezgin, and X. Zhang, ”Framework for analysis of opportunistic schedulers: average sum rate vs. average fairness”, to be submitted to RAWCOM 2008.
Consider the standard frequency selective channel model described in time domain by a channel impulse response \( h_1, \ldots, h_L \) with \( L \) taps probably correlated with channel correlation matrix \( R \), i.e. \( R_{k,l} = \mathbb{E}[h_k h_l^*] \).

Ideal cyclic prefix OFDM is applied with \( N > L \) carriers.

The fading on carrier \( k \) is given by \( H_k = \sum_{l=0}^{L-1} h_l \exp \left( -i 2\pi \frac{k}{N} \right) \).

Therefore, the power of carrier \( k \) is distributed according to an exponential distribution with mean

\[
a_k(R) = \mathbb{E}[|H_k|^2] = 2 \sum_{n=0}^{L-2} \sum_{m=n+1}^{L-1} R_{m,n} \cos \left( 2\pi \frac{k}{N} (m - n) \right) + \sum_{k=1}^{L} R_{k,k}
\]

for all \( 1 \leq k \leq N \).
$N$ parallel flat-fading channels are obtained by ideal CP-OFDM $y_k = H_k x_k + n_k$ for $1 \leq k \leq N$, transmit signals $x_k$ and AWGN $n_k$.

In the following, we focus on the average rate of the system as a function of the power allocation $p = [p_1, \ldots, p_N]$, the inverse noise power $\rho = \frac{1}{\sigma_n^2}$, and the fading variances $a = [a_1, \ldots, a_N]$

$$C(a, \rho) = \mathbb{E}[R(H, \rho, p)] \quad (26)$$

with instantaneous rate which depends on the channel vector $H = [H_1, \ldots, H_N]$ and the power allocation

$$R(H, \rho, p) = \frac{1}{N} \sum_{k=1}^{N} \log \left(1 + \rho H_k p_k \right). \quad (27)$$
We assume that the receiver has perfect CSI and consider three different types of CSI at the transmitter.

At first, assume the transmitter does not have CSI at all. Then the optimal power allocation is equal power allocation and the achievable average rate is given by

\[ C_{no}(a, \rho) = E \left[ \frac{1}{N} \sum_{k=1}^{N} \log \left( 1 + \frac{P}{N} \rho a_k \right) \right]. \]  \hspace{1cm} (28)

If the transmitter has information about \( R \) it can adapt its power allocation to \( a \) and the average rate is given by

\[ C_{part}(a, \rho) = \max_{p_k \geq 0, \sum p_k = P} E \left[ \frac{1}{N} \sum_{k=1}^{N} \log \left( 1 + \rho H_k p_k \right) \right]. \]  \hspace{1cm} (29)
If the transmitter has perfect CSI, spectral power allocation is optimal and the average rate is given by

$$C_p(a, \rho) = \mathbb{E} \left[ \max_{p_k \geq 0, \sum p_k = P} \frac{1}{N} \sum_{k=1}^{N} \log (1 + \rho H_k p_k) \right].$$

(30)

Definition 2. The frequency selective channel described by $\mathbf{R}$ has more correlated taps than the channel described by $\mathbf{S}$ if

$$a(\mathbf{R}) \succeq a(\mathbf{S}),$$

(31)

i.e. $a(\mathbf{R})$ majorizes $a(\mathbf{S})$, i.e.

$$\sum_{k=1}^{n} a_k(\mathbf{R}) \geq \sum_{k=1}^{n} a_k(\mathbf{S}) \text{ for } 1 \leq n \leq N - 1 \text{ and } \sum_{k=1}^{N} a_k(\mathbf{R}) \geq \sum_{k=1}^{N} a_k(\mathbf{S}).$$
Consider for example two taps $L = 2$ with $\mathbf{R}(c) = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$ and $N = 4$ carriers. The average channel power of the four carriers is then given by

$$a(\mathbf{R}(c)) = [2 + 2c, 2, 2, 2 - 2c]. \quad (32)$$

For completely uncorrelated coefficients, $a(\mathbf{R}(0)) = [2, 2, 2, 2]$. For completely correlated coefficients, $a(\mathbf{R}(1)) = [4, 2, 2, 0]$. From (32), it follows that all correlation matrices $\mathbf{R}(c_1), \mathbf{R}(c_2)$ are comparable.

TBD-List:

- Is it possible to compare all channel correlation matrices $\mathbf{R}$?
- Are $C(\mathbf{a}, \rho)$ Schur-convex or Schur-concave? High SNR behavior analysis.
- Maximum Loss/Gain due to spectral correlation?