Journal of Inequalities in Pure and Applied Mathematics

http://jipam.vu.edu.au/

Volume 5, Issue 2, Article 46, 2004

ON SCHUR-CONVEXITY OF EXPECTATION OF WEIGHTED SUM OF RANDOM VARIABLES WITH APPLICATIONS

HOLGER BOCHE AND EDUARD A. JORSWIECK

FRAUNHOFER INSTITUTE FOR TELECOMMUNICATIONS
HEINRICH-HERTZ-INSTITUT, EINSTEINUFER 37
D-10587 BERLIN, GERMANY.
boche@hhi.de

jorswieck@hhi.de

Received 19 November, 2003; accepted 17 April, 2004
Communicated by P. Bullen

ABSTRACT. We show that the expectation of a class of functions of the sum of weighted identically independent distributed positive random variables is Schur-concave with respect to the weights. Furthermore, we optimise the expectation by choosing extra-weights with a sum constraint. We show that under this optimisation the expectation becomes Schur-convex with respect to the weights. Finally, we explain the connection to the ergodic capacity of some multiple-antenna wireless communication systems with and without adaptive power allocation.

Key words and phrases: Schur-convex function, Optimisation, Sum of weighted random variables.

2000 *Mathematics Subject Classification.* Primary 60E15, 60G50; Secondary 94A05.

1. INTRODUCTION

The Schur-convex function was introduced by I. Schur in 1923 [11] and has many important applications. Information theory [14] is one active research area in which inequalities were extensively used. [2] was the beginning of information theory. One central value of interest is the channel capacity. Recently, communication systems which transmit vectors instead of scalars have gained attention. For the analysis of the capacity of those systems and for analyzing the impact of correlation on the performance we use Majorization theory. The connection to information theory will be further outlined in Section 6.

The distribution of weighted sums of independent random variables was studied in the literature. Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables and let

$$F(c_1, \ldots, c_n; t) = Pr(c_1 X_1 + \cdots + c_n X_n \leq t). \tag{1.1}$$

ISSN (electronic): 1443-5756

164-03

By a result of Proschan [13], if the common density of $X_1, \dots, X_n$ is symmetric about zero and log-concave, then the function F is Schur-concave in $(c_1, \dots, c_n)$. For nonsymmetric densities, analogous results are known only in several particular cases of Gamma distributions [4]. In [12], it was shown for two ($n = 2$) iid standard exponential random variables, that F is Schur-convex on $t \le (c_1 + c_2)$ and Schur-concave on $t \ge \frac{3}{2}(c_1 + c_2)$. Extensions and applications of the results in [12] are given in [9]. For discrete distributions, there are Schur-convexity results for Bernoulli random variables in [8]. Instead of the distribution in (1.1), we study the expectation of the weighted sum of random variables.

We define an arbitrary function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) > 0$ for all $x > 0$. Now, consider the following expectation

$$G(\mu) = G(\mu_1, \dots, \mu_n) = \mathbb{E}\left[f\left(\sum_{k=1}^{n} \mu_k w_k \right) \right] \tag{1.2}$$

with independent identically distributed positive[1] random variables $w_1, \dots, w_n$ according to some probability density function $p(w) : p(x) = 0 \ \ \forall x < 0$ and positive numbers $\mu_1, \dots, \mu_n$ which are in decreasing order, i.e. $\mu_1 \ge \mu_2 \ge \dots \ge \mu_n \ge 0$ with the sum constraint

$$\sum_{k=1}^{n} \mu_k = 1.$$

The function $G(\mu)$ with the parameters $f(x) = \log(1 + \rho x)$ for $\rho > 0$ and with exponentially distributed $w_1, \dots, w_n$ is very important for the analysis of some wireless communication networks. The performance of some wireless systems depends on the parameters $\mu_1, \dots, \mu_n$. Hence, we are interested in the impact of $\mu_1, \dots, \mu_n$ on the function $G(\mu_1, \dots, \mu_n)$. Because of the sum constraint in (1), and in order to compare different parameter sets $\mu^1 = [\mu_1^1, \dots, \mu_n^1]$ and $\mu^2 = [\mu_1^2, \dots, \mu_n^2]$, we use the theory of majorization. Majorization induces a partial order on the vectors μ^1 and μ^2 that have the same l_1 norm.

Our first result is that the function $G(\mu)$ is Schur-concave with respect to the parameter vector $\mu = [\mu_1, \dots, \mu_n]$, i.e. if μ^1 majorizes μ^2 then $G(\mu^1)$ is smaller than or equal to $G(\mu^2)$.

In order to improve the performance of wireless systems, adaptive power control is applied. This leads mathematically to the following objective function

$$H(\mathbf{p}, \mu) = H(p_1, \dots, p_n; \mu_1, \dots, \mu_n) = \mathbb{E}\left[f\left(\sum_{k=1}^{n} p_k \mu_k w_k \right) \right]$$

for fixed parameters $\mu_1, \dots, \mu_n$ and a sum constraint $\sum_{k=1}^{n} p_k = P$. We solve the following optimisation problem

$$I(\mu, P) = I(\mu_1, \dots, \mu_n, P) = \max H(p_1, \dots, p_n; \mu_1, \dots, \mu_n) \tag{1.3}$$
$$\text{s.t.} \sum_{k=1}^{n} p_k = P \quad \text{and} \quad p_k \ge 0 \quad 1 \le k \le n$$

for fixed $\mu_1, \dots, \mu_n$. The optimisation in (1.3) is a convex programming problem which can be completely characterised using the Karush-Kuhn-Tucker (KKT) conditions.

Using the optimality conditions from (1.3), we characterise the impact of the parameters $\mu_1, \dots, \mu_n$ on the function $I(\mu, P)$. Interestingly, the function $I(\mu, P)$ is a Schur-convex function with respect to the parameter vector $\mu = [\mu_1, \dots, \mu_n]$, i.e. if $\mu^{\mathbf{1}}$ majorizes $\mu^{\mathbf{2}}$ then $I(\mu^1, P)$ is larger than $I(\mu^2, P)$ for arbitrary sum constraint P.

[1] A random variable is obviously positive, if $Pr(w_l < 0) = 0$. Those variables are called positive throughout the paper.

The remainder of this paper is organised as follows. In the next, Section 2, we introduce the notation and give definitions and formally state the problems. Next, in Section 3 we prove that $G(\mu)$ is Schur-concave. The optimal solution of a convex programming problem in Section 4 is then used to show that $I(\mu, P)$ is Schur-convex for all $P > 0$. The connection and applications in wireless communications are pointed out in Section 6.

2. BASIC RESULTS, DEFINITIONS AND PROBLEM STATEMENT

First, we give the necessary definitions which will be used throughout the paper.

Definition 2.1. For two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ one says that the vector $\mathbf{x}$ majorizes the vector $\mathbf{y}$ and writes

$$\mathbf{x} \succ \mathbf{y} \ \text{ if } \ \sum_{k=1}^{m} x_k \geq \sum_{k=1}^{m} y_k \ \ , m = 1, \ldots, n-1. \ \text{ and } \ \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k.$$

The next definition describes a function Φ which is applied to the vectors $\mathbf{x}$ and $\mathbf{y}$ with $\mathbf{x} \succ \mathbf{y}$:

Definition 2.2. A real-valued function Φ defined on $\mathcal{A} \subset \mathbb{R}^n$ is said to be *Schur-convex* on $\mathcal{A}$ if

$$\mathbf{x} \succ \mathbf{y} \ \text{ on } \ \mathcal{A} \Rightarrow \Phi(\mathbf{x}) \geq \Phi(\mathbf{y}).$$

Similarly, Φ is said to be *Schur-concave* on $\mathcal{A}$ if

$$\mathbf{x} \succ \mathbf{y} \ \text{ on } \ \mathcal{A} \Rightarrow \Phi(\mathbf{x}) \leq \Phi(\mathbf{y}).$$

Remark 2.1. If the function $\Phi(\mathbf{x})$ on $\mathcal{A}$ is Schur-convex, the function $-\Phi(\mathbf{x})$ is Schur-concave on $\mathcal{A}$.

Example 2.1. Suppose that $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^n$ are positive real numbers and the function Φ is defined as the sum of the quadratic components of the vectors, i.e. $\Phi_2(\mathbf{x}) = \sum_{k=1}^{n} |x_k|^2$. Then, it is easy to show that the function Φ_2 is Schur-concave on $\mathbb{R}_+^n$, i.e. if $\mathbf{x} \succ \mathbf{y} \Rightarrow \Phi_2(\mathbf{x}) \leq \Phi_2(\mathbf{y})$.

The definition of Schur-convexity and Schur-concavity can be extended if another function $\Psi : \mathbb{R} \to \mathbb{R}$ is applied to $\Phi(\mathbf{x})$. Assume that Φ is Schur-concave, if the function Ψ is monotonic increasing then the expression $\Psi(\Phi(\mathbf{x}))$ is Schur-concave, too. If we take for example the function $\Psi(n) = \log(n)$ for $n \in \mathbb{R}_+$ and the function Φ_p from the example above, we can state that the composition of the two functions $\Psi(\Phi_p(\mathbf{x}))$ is Schur-concave on $\mathbb{R}_+^n$. This result can be generalised for all possible compositions of monotonic increasing as well as decreasing functions, and Schur-convex as well as Schur-concave functions. For further reading see [11].

We will need the following lemma (see [11, Theorem 3.A.4]) which is sometimes called Schur's condition. It provides an approach for testing whether some vector valued function is Schur-convex or not.

Lemma 2.2. *Let $\mathcal{I} \subset \mathbb{R}$ be an open interval and let $f : \mathcal{I}^n \to \mathbb{R}$ be continuously differentiable. Necessary and sufficient conditions for f to be Schur-convex on $\mathcal{I}^n$ are*

$$f \quad \textit{is symmetric on} \quad \mathcal{I}^n$$

and

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

Since $f(\mathbf{x})$ is symmetric, Schur's condition can be reduced as [11, p. 57]

$$(x_1 - x_2)\left(\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2}\right) \geq 0. \tag{2.1}$$

From Lemma 2.2, it follows that $f(\mathbf{x})$ is a Schur-concave function on $\mathcal{I}^n$ if $f(\mathbf{x})$ is symmetric and

$$(x_1 - x_2)\left(\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2}\right) \leq 0. \tag{2.2}$$

Finally, we propose the concrete problem statements: At first, we are interested in the impact of the vector μ on the function $G(\mu)$.

This problem is solved in Section 3.

Problem 1. Is the function $G(\mu_1, \ldots, \mu_n)$ in (1.2) a Schur-concave function, i.e. with $\mu^1 = [\mu_1^1, \ldots, \mu_n^1]$ and $\mu^2 = [\mu_1^2, \ldots, \mu_n^2]$ it holds

$$\mu^1 \succeq \mu^2 \Longrightarrow G(\mu^1) \leq G(\mu^2)?$$

Next, we need to solve the following optimisation problem in order to characterise the impact of the vector μ on the function $I(\mu, P)$.

We solve this problem in Section 4.

Problem 2. Solve the following optimisation problem

$$I(\mu_1, \ldots, \mu_n, P) = \max H(p_1, \ldots, p_n; \mu_1, \ldots, \mu_n) \tag{2.3}$$
$$\text{s.t.} \sum_{k=1}^{n} p_k = P \quad \text{and} \quad p_k \geq 0 \quad 1 \leq k \leq n$$

for fixed $\mu_1, \ldots, \mu_n$.

Finally, we are interested in whether the function in (2.3) is Schur-convex or Schur-concave with respect to the parameters $\mu_1, \ldots, \mu_n$. This leads to the last Problem statement 3.

This problem is solved in Section 5.

Problem 3. Is the function $I(\mu, P)$ in (2.3) a Schur-convex function, i.e. for all $P > 0$

$$\mu^1 \succeq \mu^2 \Longrightarrow I(\mu^1, P) \leq I(\mu^2, P)?$$

3. SCHUR-CONCAVITY OF $G(\mu)$

In order to solve Problem 1, we consider first the function $f(x) = \log(1 + x)$. This function naturally arises in the information theoretic analysis of communication systems [14]. That followed, we generalise the statement of the theorem for all concave functions $f(x)$. Therefore, Theorem 3.1 can be seen as a corollary of Theorem 3.3.

Theorem 3.1. *The function*

$$C_1(\mu) = C_1(\mu_1, \ldots, \mu_n) = \mathbb{E}\left[\log\left(1 + \sum_{k=1}^{n} \mu_k w_k\right)\right] \tag{3.1}$$

with iid positive random variables $w_1, \ldots, w_n$ is a Schur-concave function with respect to the parameters $\mu_1, \ldots, \mu_n$.

Proof. We will show that Schur's condition (2.2) is fulfilled by the function $C_1(\mu)$ with $\mu = [\mu_1, \ldots, \mu_n]$. The first derivative of $C_1(\mu)$ with respect to μ_1 and μ_2 is given by

$$\alpha_1 = \frac{\partial C_1}{\partial \mu_1} = \mathbb{E}\left[\frac{w_1}{1 + \sum_{k=1}^{n} \mu_k w_k}\right] \tag{3.2}$$

$$\alpha_2 = \frac{\partial C_1}{\partial \mu_2} = \mathbb{E}\left[\frac{w_2}{1 + \sum_{k=1}^{n} \mu_k w_k}\right]. \tag{3.3}$$

Since $\mu_1 \geq \mu_2$ by definition, we have to show that

$$\mathbb{E}\left[\frac{w_1 - w_2}{z + \mu_1 w_1 + \mu_2 w_2}\right] \leq 0 \tag{3.4}$$

with $z = 1 + \sum_{k=3}^{n} \mu_k w_k$. The expectation operator in (3.4) can be written as a n-fold integral over the probability density functions $p(w_1), \ldots, p(w_n)$. In the following, we show that for all $z \geq 0$

$$\int_0^\infty \int_0^\infty g(w_1, w_2, z) p(w_1) p(w_2) dw_1 dw_2 \leq 0 \tag{3.5}$$

with $g(w_1, w_2, z) = \frac{w_1 - w_2}{z + \mu_1 w_1 + \mu_2 w_2}$. Rewrite the double integral in (3.5) as

$$\begin{aligned}\int_0^\infty \int_0^\infty & g(w_1, w_2, z) p(w_1) p(w_2) dw_1 dw_2 \\ &= \int_{w_1=0}^\infty \int_{w_2=0}^{w_1} \left(g(w_1, w_2, z) + g(w_2, w_1, z)\right) p(w_1) p(w_2) dw_1 dw_2\end{aligned} \tag{3.6}$$

because the random variables w_1 and w_2 are independent identically distributed. In (3.6), we split the area of integration into the area in which $w_1 > w_2$ and $w_2 \geq w_1$ and used the fact, that $g(w_1, w_2, z)$ for $w_1 > w_2$ is the same as $g(w_2, w_1, z)$ for $w_2 \geq w_1$. Now, the expression $g(w_1, w_2, z) + g(w_2, w_1, z)$ can be written for all $z \geq 0$ as

$$\begin{aligned} g(w_1, w_2, z) + g(w_2, w_1, z) &= \frac{(w_1 - w_2)(\mu_1 w_2 + \mu_2 w_1 - \mu_1 w_1 - \mu_2 w_2)}{(z + \mu_1 w_1 + \mu_2 w_2)(z + \mu_1 w_2 + \mu_2 w_1)} \\ &= \frac{(w_1 - w_2)^2 (\mu_2 - \mu_1)}{(z + \mu_1 w_1 + \mu_2 w_2)(z + \mu_1 w_2 + \mu_2 w_1)}. \end{aligned} \tag{3.7}$$

From assumption $\mu_2 \leq \mu_1$ and (3.7) follows (3.5) and (3.4). □

Remark 3.2. Interestingly, Theorem 3.1 holds for all probability density functions which fulfill $p(x) = 0$ for almost every $x < 0$. The main precondition is that the random variables w_1 and w_2 are independent and identically distributed. This allows the representation in (3.6).

Theorem 3.1 answers Problem 1 only for a specific choice of function $f(x)$. We can generalise the statement of Theorem 3.1 in the following way. However, the most important, in practice is the case in which $f(x) = \log(1 + x)$.

Theorem 3.3. *The function $G(\mu)$ as defined in (1.2) is Schur-concave with respect to μ if the random variables $w_1, \ldots, w_n$ are positive identically independent distributed and if the inner function $f(x)$ is monotonic increasing and concave.*

Proof. Let us define the difference of the first derivatives of $f(\sum_{k=1}^{n} \mu_k w_k)$ with respect to μ_1 and μ_2 as

$$\Delta(w_1, w_2) = \left(\frac{\partial f(\sum_{k=1}^{n} \mu_k w_k)}{\partial \mu_1} - \frac{\partial f(\sum_{k=1}^{n} \mu_k w_k)}{\partial \mu_2}\right).$$

Since the function f is monotonic increasing and concave, $f''(x) \leq 0$ and $f'(x)$ is monotonic decreasing, i.e.

$$f'(x_1) \leq f'(x_2) \qquad \text{for all } \; x_1 \geq x_2$$

Note, that $w_1 \geq w_2$ and $\mu_1 \geq \mu_2$ and $\mu_1 w_2 + \mu_2 w_2 \geq \mu_1 w_2 + \mu_2 w_1$. Therefore, it holds

$$(w_1 - w_2)\left(f'(\mu_1 w_1 + \mu_2 w_2 + \sum_{k=3}^{n} \mu_k w_k) - f'(\mu_1 w_2 + \mu_2 w_1 + \sum_{k=3}^{n} \mu_k w_k)\right) \leq 0$$

Using equation (3.6), it follows

$$\int_{w_1=0}^{\infty}\int_{w_2=0}^{w_1}\left(\Delta(w_1,w_2)-\Delta(w_2,w_1)\right)p(w_1)p(w_2)dw_1dw_2 \le 0 \tag{3.8}$$

because the densities are positive. This verifies Schur's condition for (1.2). □

The condition in Theorem 3.3 can be easily checked. Consider for example the function

$$k(x) = \frac{x}{1+x}. \tag{3.9}$$

It is easily verified that the condition in Theorem 3.3 is fulfilled by (3.9). By application of Theorem 3.3 it has been shown that the function $K(\mu)$ defined as

$$K(\mu) = \mathbb{E}\left[\frac{\sum_{k=1}^{n}\mu_k w_k}{1+\sum_{k=1}^{n}\mu_k w_k}\right]$$

is Schur-concave with respect to $\mu_1,\dots,\mu_n$.

4. OPTIMALITY CONDITIONS FOR CONVEX PROGRAMMING PROBLEM $\max H(\mu,\mathbf{p})$

Next, we consider the optimisation problem in (2.3) from Problem 2. Here, we restrict our attention to the case $f(x)=\log(1+x)$. The motivation for this section is to find a characterisation of the optimal $\mathbf{p}$ which can be used to characterise the impact of μ under the optimum strategy $\mathbf{p}$ on $H(\mu,\mathbf{p})$. The results of this section, mainly the KKT optimality conditions are used in the next section to show that $H(\mu,\mathbf{p})$ with the optimal $\mathbf{p}^*(\mu)$ is Schur-convex.

The objective function is given by

$$C_2(\mathbf{p},\mu) = \mathbb{E}\left[\log\left(1+\sum_{k=1}^{n} p_k\mu_k w_k\right)\right] \tag{4.1}$$

and the optimisation problem reads

$$\mathbf{p}^* = \arg\max C_2(\mathbf{p},\boldsymbol{\mu}) \ \text{ s.t. } \sum_{k=1}^{n} p_k = 1 \ \text{ and } \ p_k \ge 0 \ \ 1\le k\le n. \tag{4.2}$$

The optimisation problem in (4.2) is a convex optimisation problem. Therefore, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for the optimality of some $\mathbf{p}^*$ [5]. The Lagrangian for the optimisation problem in (4.2) is given by

$$\mathcal{L}(\mathbf{p},\lambda_1,\dots,\lambda_n,\nu) = C_2(\mathbf{p},\boldsymbol{\mu}) + \sum_{k=1}^{n}\lambda_k p_k + \nu\left(P-\sum_{k=1}^{n}p_k\right) \tag{4.3}$$

with the Lagrangian multiplier ν for the sum constraint and the Lagrangian multipliers $\lambda_1,\dots,\lambda_n$ for the positiveness of $p_1,\dots,p_n$. The first derivative of (4.3) with respect to p_l is given by

$$\frac{d\mathcal{L}}{dp_l} = \mathbb{E}\left[\frac{\mu_l w_l}{1+\sum_{k=1}^{n}\mu_k p_k w_k}\right] + \lambda_l - \nu. \tag{4.4}$$

The KKT conditions are given by

$$\begin{aligned} \mathbb{E}\left[\frac{\mu_l w_l}{1+\sum_{k=1}^n \mu_k p_k w_k}\right] &= \nu - \lambda_l \qquad 1 \le l \le n, \\ \nu &\ge 0, \\ \lambda_k &\ge 0 \qquad\quad 1 \le l \le n, \\ p_k &\ge 0 \qquad\quad 1 \le l \le n, \\ P - \sum_{k=1}^n p_k &= 1. \end{aligned} \tag{4.5}$$

We define the following coefficients

$$\alpha_k(\mathbf{p}) = \int_0^\infty e^{-t} \prod_{l=1, l\neq k}^{n_T} \frac{1}{1+tp_l\mu_l} \cdot \frac{\mu_k}{(1+t\mu_k p_k)^2} dt. \tag{4.6}$$

These coefficients in (4.6) naturally arise in the first derivative of the Lagrangian of (4.2) and directly correspond to the first KKT condition in (4.5) where we have used the fact that

$$\mathbb{E}\left[\frac{w_l}{1+\sum_{k=1}^n p_k\mu_k w_k}\right] = \mathbb{E}\left[w_l \int_0^\infty e^{-t(1+\sum_{k=1}^n p_k\mu_k w_k)} dt\right].$$

Furthermore, we define the set of indices for which $p_i > 0$, i.e.

$$\mathcal{I}(\mathbf{p}) = \{k \in [1, \dots, n_T] : p_k > 0\}. \tag{4.7}$$

We have the following characterisation of the optimum point $\hat{\mathbf{p}}$.

Theorem 4.1. *A necessary and sufficient condition for the optimality of $\hat{\mathbf{p}}$ is*

$$\begin{aligned} \{k_1, k_2 \in \mathcal{I}(\hat{\mathbf{p}}) &\Longrightarrow \alpha_{k_1} = \alpha_{k_2} \ \textit{ and} \\ k \notin \mathcal{I}(\hat{\mathbf{p}}) &\Longrightarrow \alpha_k \le \max_{l \in \mathcal{I}(\hat{\mathbf{p}})} \alpha_l\}. \end{aligned} \tag{4.8}$$

This means that all indices l which obtain p_l greater than zero have the same $\alpha_l = \max_{l\in[1,\dots,n_T]}$. Furthermore, all other α_i are less than or equal to α_l.

Proof. We name the optimal point $\hat{\mathbf{p}}$, i.e. from (4.2)

$$\hat{\mathbf{p}} = \arg \max_{||\mathbf{p}|| \le P, p_i \ge 0} C(\mathbf{p}, \rho, \mu).$$

Let the $\mu_1, \dots, \mu_{n_T}$ be fixed. We define the parametrised point

$$\mathbf{p}(\tau) = (1-\tau)\hat{\mathbf{p}} + \tau\mathbf{p}$$

with arbitrary $\mathbf{p} : ||\mathbf{p}|| \le P,\ p_i \ge 0$. The objective function is given by

$$C(\tau) = \mathbb{E} \log \left(1 + \rho \sum_{l=1}^{n_T} \hat{p}_k \mu_k w_k + \rho\tau \sum_{l=1}^{n_T} (p_k - \hat{p}_k)\mu_k w_k\right). \tag{4.9}$$

The first derivative of (4.9) at the point $\tau = 0$ is given by

$$\left.\frac{dC(\tau)}{d\tau}\right|_{\tau=0} = \sum_{k=1}^{n_T} (p_k - \hat{p}_k)\alpha_k(\hat{\mathbf{p}})$$

with $\alpha_k(\hat{\mathbf{p}})$ defined in (4.6). It is easily shown that the second derivative of $C(\tau)$ is always smaller than zero for all $0 \le \tau \le 1$. Hence, it suffices to show that the first derivative of $C(\tau)$ at the point $\tau = 0$ is less than or equal to zero, i.e.

$$\sum_{k=1}^{n_T} (p_k - \hat{p}_k)\alpha_k(\hat{p}) \le 0. \tag{4.10}$$

We split the proof into two parts. In the first part, we will show that the condition in (4.8) is sufficient. We assume that (4.8) is fulfilled. We can rewrite the first derivative of $C(\tau)$ at the point $\tau = 0$ as

$$\begin{aligned} Q &= \sum_{k=1}^{n_T} (\hat{p}_k - p_k)\alpha_k(\hat{p}_k) \\ &= \sum_{k=1}^{n_T} \hat{p}_k \alpha_k(\hat{p}) - \sum_{k=1}^{n} p_k \alpha_k(\hat{p}) \\ &= \max_{k \in [1,\dots,n_T]} \alpha_k(\hat{p}) \sum_{l \in \mathcal{I}(\hat{p})} \hat{p}_l - \sum_{l=1}^{n_T} p_l \alpha_l(\hat{p}). \end{aligned} \tag{4.11}$$

But we have that

$$\sum_{l=1}^{n_T} p_l \alpha_l(\hat{p}) \le \sum_{l=1}^{n_T} p_l \max_{k \in [1,\dots,n_T]} \alpha_l(\hat{p}).$$

Therefore, it follows for Q in (4.11)

$$Q \ge \max_{k \in [1,\dots,n]} \alpha_k(\hat{p}) \left(\sum_{l \in \mathcal{I}(\hat{p})} \hat{p}_l - \sum_{l=1}^{n} p_l \right) = 0,$$

i.e. (4.10) is satisfied.

In order to show that condition (4.8) is a necessary condition for the optimality of power allocation $\hat{\mathbf{p}}$, we study two cases and prove them by contradiction.

(1) Assume (4.8) is not true. Then we have a $k \in \mathcal{I}(\hat{p})$ and $k_0 \in \mathcal{I}(\hat{p})$ with the following properties:

$$\max_{1 \le k \le n_T} \alpha_k(\hat{p}) = \alpha_{k_0(\hat{p})}$$

and $\alpha_k(\hat{p}) < \alpha_{k_0}(\hat{p})$. We set $\tilde{p}_{k_0} = 1$ and $\tilde{p}_{i \in [1,\dots,n_T] k_0} = 0$. It follows that

$$\sum_{l=1}^{n_T} (\hat{p}_k - \tilde{p}_k)\alpha_k(\hat{p}) < 0$$

which is a contradiction.

(2) Assume there is a $k_0 : \alpha_{k_0} > \alpha_k$ with $k_0 \notin \mathcal{I}(\hat{p})$ and $k \in \mathcal{I}(\hat{p})$, then set $\tilde{p}_{k_0} = 1$ and $\tilde{o}_{l \in [1,\dots,n_T] k_0} = 0$. Then we have the contradiction

$$\sum_{k=1}^{n_T} (\hat{p}_k - \tilde{p}_k)\alpha_k < 0.$$

This completes the proof of Theorem 4.1. □

5. SCHUR-CONVEXITY OF $I(\mu, P)$

We use the results from the previous section to derive the Schur-convexity of the function $I(\mu, P)$ for all $P > 0$. The representation of the $\alpha_k(\mathbf{p})$ in (4.6) is necessary to show that the condition $\frac{p_l}{\mu_l} \geq \frac{p_{l+1}}{\mu_{l+1}}$ is fulfilled for all $1 \leq l \leq n-1$. This condition is stronger than majorization, i.e. it follows that $\mathbf{p} \succeq \mu$ [11, Proposition 5.B.1]. Note that $\sum_{k=1}^n p_k = \sum_{k=1}^n \mu_k = 1$. The result is summarised in the following theorem.

Theorem 5.1. *For all $P > 0$, the function $I(\mu, P)$ is a Schur-convex function with respect to the parameters $\mu_1, \ldots, \mu_n$.*

Proof. The proof is constructed in the following way: At first, we consider two arbitrary parameter vectors μ^1 and μ^2 which satisfy $\mu^1 \succeq \mu^2$. Then we construct all possible linear combinations of μ^1 and μ^2, i.e. $\mu(\theta) = \theta\mu^2 + (1-\theta)\mu^1$. Next, we study the parametrised function $I(\mu(\theta))$ as a function of the linear combination parameter θ. We show that the first derivative of the parametrised capacity with respect to θ is less than or equal to zero for all $0 \leq \theta \leq 1$. This result holds for all μ^1 and μ^2. As a result, we have shown that the function $I(\mu)$ is Schur-convex with respect to μ.

With arbitrary μ^1 and μ^2 which satisfy $\mu^1 \succeq \mu^2$, define the vector

$$\mu(\theta) = \theta\mu^2 + (1-\theta)\mu^1 \tag{5.1}$$

for all $0 \leq \theta \leq 1$. The parameter vector $\mu(\theta)$ in (5.1) has the following properties which will be used throughout the proof.

- The parametrisation in (5.1) is order preserving between the vectors μ^1 and μ^2, i.e.
$$\forall 0 \leq \theta_1 \leq \theta_2 \leq 1 : \mu^2 = \mu(1) \preceq \mu(\theta_2) \preceq \mu(\theta_1) \preceq \mu(0) = \mu^1.$$
This directly follows from the definition of majorization. E.g. the first inequality is obtained by
$$\mu(\theta_2) = \theta_2\mu^2 + (1-\theta_2)\mu^1 \geq \theta_2\mu^2 + (1-\theta_2)\mu^2 = \mu^2.$$
- The parametrisation in (5.1) is order preserving between the elements, i.e. for ordered elements in μ^1 and μ^2, it follows that for the elements in $\mu(\theta)$, for all $0 \leq \theta \leq 1$,
$$\forall 1 \leq l \leq n_T - 1 : \mu_l(\theta) \geq \mu_{l+1}(\theta).$$
This directly follows from the definition in (5.1).

The optimum power allocation is given by $p_1(\theta), \ldots, p_n(\theta)$. The parametrised objective function $H(\mu(\theta), \mathbf{p}(\theta))$ as a function of the parameter θ is then given by

$$\begin{aligned} H(\theta) &= \mathbb{E} \log\left(1 + \rho \sum_{k=1}^n \mu_k(\theta) p_k(\theta) w_k\right) \\ &= \mathbb{E} \log\left(1 + \rho \sum_{k=1}^n (\mu_k^1 + \theta(\mu_k^2 - \mu_k^1)) p_k(\theta) w_k\right). \end{aligned} \tag{5.2}$$

The first derivative of (5.2) with respect to θ is given by

$$\frac{dH(\theta)}{d\theta} = \mathbb{E}\left(\frac{\sum_{k=1}^n (\mu_k^2 - \mu_k^1) p_k(\theta) w_k + \frac{dp_k(\theta)}{d\theta}(\mu_k^2 + \theta(\mu_k^1 - \mu_k^2))}{1 + \sum_{k=1}^n (\mu_k^2 + \theta(\mu_k^1 - \mu_k^2)) p_k(\theta) w_k}\right). \tag{5.3}$$

Let us consider the second term in (5.3) first. Define

$$\phi_k(\theta) = (\mu_k^2 + \theta(\mu_k^1 - \mu_k^2)) \;\; \forall k = 1, \ldots, n.$$

Then we have

$$\sum_{k=1}^{n} \frac{dp_k(\theta)}{d\theta} \mathbb{E}\left(\frac{\phi_k(\theta) w_k}{1+\sum_{k=1}^{n} \phi_k(\theta) p_k(\theta) w_k} \right) = \sum_{k=1}^{n} \frac{dp_k(\theta)}{d\theta} \alpha_k(\theta). \tag{5.4}$$

In order to show that (5.4) is equal to zero, we define the index m for which holds

$$\frac{dp_k(\theta)}{d\theta} \neq 0 \quad \forall 1 \leq k \leq m \quad \text{and} \quad \frac{dp_k(\theta)}{d\theta} = 0 \;\; k \geq m+1. \tag{5.5}$$

We split the sum in (5.4) in two parts, i.e.

$$\sum_{k=1}^{m} \frac{dp_k(\theta)}{d\theta} \alpha_k(\theta) + \sum_{k=m+1}^{n} \frac{dp_k(\theta)}{d\theta} \alpha_k(\theta). \tag{5.6}$$

For all $1 \leq k \leq m$ we have from (5.5) three cases:

- First case: $p_m(\theta) > 0$ and obviously $p_1(\theta) > 0, ..., p_{m-1}(\theta) > 0$. It follows that
$$\alpha_1(\theta) = \alpha_2(\theta) = \cdots = \alpha_m(\theta)$$
- Second case: There exists an $\epsilon_1 > 0$ such that $p_m(\theta) = 0$ and $p_m(\theta + \epsilon) > 0$ for all $0 < \epsilon \leq \epsilon_1$. Therefore, it holds
$$\alpha_1(\theta+\epsilon) = \cdots = \alpha_m(\theta+\epsilon). \tag{5.7}$$
- Third case: There exists an $\epsilon_1 > 0$ such that $p_m(\theta) = 0$ and $p_m(\theta - \epsilon) > 0$ for all $0 < \epsilon \leq \epsilon_1$. Therefore, it holds
$$\alpha_1(\theta-\epsilon) = \cdots = \alpha_m(\theta-\epsilon). \tag{5.8}$$

Next, we use the fact that if f and g are two continuous functions defined on some closed interval $\mathcal{O}$, $f, g : \mathcal{O} \to \mathcal{R}$. Then the set of points $t \in \mathcal{O}$ for which $f(t) = g(t)$ is either empty or closed.

Assume the case in (5.7). The set of points θ for which $\alpha_k(\theta) = \alpha_1(\theta)$ is closed. Hence, it holds

$$\alpha_k(\theta) = \lim_{\epsilon \to 0} \alpha_k(\theta+\epsilon) = \lim_{\epsilon \to 0} \alpha_1(\theta+\epsilon) = \alpha_1(\theta). \tag{5.9}$$

For the case in (5.8), it holds

$$\alpha_k(\theta) = \lim_{\epsilon \to 0} \alpha_k(\theta-\epsilon) = \lim_{\epsilon \to 0} \alpha_1(\theta-\epsilon) = \alpha_1(\theta).$$

The consequence from (5.9) and (5) is that all active k with $p_k > 0$ at point θ and all k which occur or vanish at this point θ fulfill $\alpha_1(\theta) = \alpha_2(\theta) = \cdots = \alpha_m(\theta)$. Therefore, the first addend in (5.6) is

$$\sum_{k=1}^{m} \frac{dp_k(\theta)}{d\theta} = \alpha_1(\theta) \sum_{k=1}^{m} \frac{dp_k(\theta)}{d\theta} = 0.$$

The second addend in (5.6) is obviously equal to zero. We obtain for (5.3)

$$\frac{dH(\theta)}{d\theta} = \mathbb{E}\left(\frac{\sum_{k=1}^{n} (\mu_k^2 - \mu_k^1) p_k(\theta) w_k}{1 + \sum_{k=1}^{n} (\mu_k^2 + \theta(\mu_k^1 - \mu_k^2)) p_k(\theta) w_k} \right).$$

We are going to show that

$$\sum_{k=1}^{n} (\mu_k^2 - \mu_k^1) \mathbb{E}\left(\frac{p_k(\theta) w_k}{1 + \sum_{k=1}^{n} (\mu_k^2 + \theta(\mu_k^1 - \mu_k^2)) p_k(\theta) w_k} \right) \leq 0. \tag{5.10}$$

We define

$$a_k = \mu_k^1 - \mu_k^2$$
$$s_l = \sum_{k=1}^{l} a_k$$
$$s_n = 0$$
$$s_0 = 0.$$

Therefore, it holds that $s_k \geq 0$ for all $1 \leq k \leq n$. We can reformulate (5.10) and obtain

$$\sum_{l=1}^{n-1} s_l(b_l(\theta) - b_{l+1}(\theta)) \geq 0 \tag{5.11}$$

with

$$b_l(\theta) = \mathbb{E}\left(\frac{p_l(\theta)w_l}{1+\sum_{k=1}^n (\mu_k^2 + \theta(\mu_k^1 - \mu_k^2))p_k(\theta)w_k}\right).$$

The inequality in (5.11) is fulfilled if

$$b_l(\theta) \geq b_{l+1}(\theta).$$

The term b_l in (5) is related to α_l from (4.8) by

$$b_l(\theta) = \frac{p_l(\theta)}{\mu_l(\theta)}\alpha_l(\theta).$$

As a result, we obtain the sufficient condition for the monotony of the parametrised function $H(\theta)$

$$\frac{p_l(\theta)}{\mu_l(\theta)} \geq \frac{p_{l+1}(\theta)}{\mu_{l+1}(\theta)}. \tag{5.12}$$

As mentioned above this is a stronger condition than that the vector $\mathbf{p}$ majorizes the vector μ. From (5.12) it follows that $\mu \succeq \mathbf{p}$.

Finally, we show that the condition in (5.12) is always fulfilled by the optimum $\mathbf{p}$. In the following, we omit the index θ. The necessary and sufficient condition for the optimal $\mathbf{p}$ is that for active $p_l > 0$ and $p_{l+1} > 0$ it holds

$$\alpha_l - \alpha_{l+1} = 0,$$

i.e.

$$\int_0^\infty e^{-t} f(t)\frac{\mu_l}{1+\rho t\mu_l p_l}dt - \int_0^\infty e^{-t} f(t)\frac{\mu_{l+1}}{1+\rho t\mu_{l+1} p_{l+1}}dt = 0 \tag{5.13}$$

with

$$f(t) = \prod_{k=1}^n \frac{1}{1+\rho t\mu_k p_k}$$

and

$$g_l(t) = (1+\rho t\mu_l p_l)^{-1}(1+\rho t\mu_{l+1}p_{l+1})^{-1}.$$

From (5.13) it follows that

$$\int_0^\infty e^{-t} f(t) g_l(t)\left(\mu_l - \mu_{l+1} - (\rho t\mu_{l+1}\mu_l)(p_l - p_{l+1})\right)dt = 0.$$

This gives

$$\int_0^\infty e^{-t} f(t) g_l(t)\left(\frac{\mu_l - \mu_{l+1}}{p_l - p_{l+1}}\frac{1}{\rho\mu_l\mu_{l+1}} - t\right)dt = 0$$

and

$$\frac{\mu_l - \mu_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \mu_l \mu_{l+1}} \int_0^\infty e^{-t} f(t) g_l(t) dt - \int_0^\infty e^{-t} f(t) g_l(t) t dt = 0. \tag{5.14}$$

Note the following facts about the functions $f(t)$ and $g_l(t)$

$$\begin{aligned} g_l(t) \geq 0 \quad \forall\, 0 \leq t \leq \infty \quad & f(t) \geq 0 \quad \forall\, 0 \leq t \leq \infty \\ \frac{dg_l(t)}{dt} \leq 0 \quad \forall\, 0 \leq t \leq \infty \quad & \frac{df(t)}{dt} \leq 0 \quad \forall\, 0 \leq t \leq \infty. \end{aligned} \tag{5.15}$$

By partial integration we obtain the following inequality

$$\int_0^\infty f(t) g_l(t)(1-t) e^{-t} dt = \left(f(t) g_l(t) t e^{-t}\right)_{t=0}^\infty - \int_0^\infty \frac{d(f(t) g_l(t))}{dt} t e^{-t} dt \geq 0. \tag{5.16}$$

From (5.16) and the properties of $f(t)$ and $g_l(t)$ in (5.15) follows that

$$\int_0^\infty e^{-t} f(t) g_l(t) dt \geq \int_0^\infty t e^{-t} f(t) g_l(t) dt.$$

Now we can lower bound the equality in (5.14) by

$$\begin{aligned} 0 &= \frac{\mu_l - \mu_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \mu_l \mu_{l+1}} \int_0^\infty e^{-t} f(t) g_l(t) dt - \int_0^\infty e^{-t} f(t) g_l(t) t dt \\ &\geq \frac{\mu_l - \mu_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \mu_l \mu_{l+1}} - 1. \end{aligned} \tag{5.17}$$

From (5.17) it follows that

$$1 \geq \frac{\mu_l - \mu_{l+1}}{p_l - p_{l+1}} \frac{1}{\rho \mu_l \mu_{l+1}}$$

and further on

$$\mu_l - \mu_{l+1} \leq (p_l - p_{l+1}) \rho \mu_l \mu_{l+1}. \tag{5.18}$$

From (5.18) we have

$$\mu_l (1 - \rho \mu_{l+1} p_l) \leq \mu_{l+1} (1 - \rho \mu_l p_{l+1})$$

and finally

$$\rho \mu_{l+1} p_l \geq \rho \mu_l p_{l+1}. \tag{5.19}$$

From (5.19) follows the inequality in (5.12). This result holds for all μ^1 and μ^2 with $\sum_{k=1}^n \mu_k^1 = \sum_{k=1}^n \mu_k^2 = 1$. As a result, $I(\mu)$ is a Schur-convex function of μ. This completes the proof. □

6. APPLICATION AND CONNECTION TO WIRELESS COMMUNICATION THEORY

As mentioned in the introduction, the three problem statements have an application in the analysis of the maximum amount of information which can be transmitted over a wireless vector channel. Recently, the improvement of the performance and capacity of wireless systems employing multiple transmit and/or receive antennae was pointed out in [15, 6]. Three scenarios are practical relevant: The case when the transmitter has no channel state information (CSI), the case in which the transmitter knows the correlation (covariance feedback), and the case where the transmitter has perfect CSI. These cases lead to three different equations for the average mutual information. Using the results from this paper, we completely characterize the impact of correlation on the performance of multiple antenna systems.

We say, that a channel is more correlated than another channel, if the vector of ordered eigenvalues of the correlation matrix majorizes the other vector of ordered eigenvalues. The average mutual information of a so called wireless multiple-input single-output (MISO) system with n_T transmit antennae and one receive antenna is given by

$$C_{noCSI}(\mu_1, \ldots, \mu_{n_T}, \rho) = \mathbb{E} \log_2 \left(1 + \rho \sum_{k=1}^{n_T} \mu_k w_k \right) \tag{6.1}$$

with signal to noise ratio (SNR) ρ and transmit antenna correlation matrix $\mathbf{R}_T$ which has the eigenvalues $\mu_1, \ldots, \mu_{n_T}$ and iid standard exponential random variables $w_1, \ldots, w_{n_T}$. In this scenario it is assumed that the receiver has perfect channel state information (CSI) while the transmit antenna array has no CSI. The transmission strategy that leads to the mutual information in (6.1) is Gaussian codebook with equal power allocation, i.e. the transmit covariance matrix $\mathbf{S} = \mathbb{E}\mathbf{x}\mathbf{x}^H$, with transmit vectors $\mathbf{x}$ that is complex standard normal distributed with covariance matrix $\mathbf{S}$, is the normalised identity matrix, i.e. $\mathbf{S} = \frac{1}{n_T}\mathbf{I}$.

The ergodic capacity in (6.1) directly corresponds to C_1 in (3.1). Applying Theorem 3.1, the impact of correlation can be completely characterized. The average mutual information is a Schur-concave function, i.e. correlation always decreases the average mutual information. See [2] for an application of the results from Theorem 3.1. If the transmitter has perfect CSI, the ergodic capacity is given by

$$C_{pCSI}(\mu_1, ..., \mu_n, \rho) = \mathbb{E} \log_2 \left(1 + \rho \sum_{k=1}^{n} \mu_k w_k \right).$$

This expression is a scaled version of (6.1). Therefore, the same analysis can be applied.

If the transmit antenna array has partial CSI in terms of long-term statistics of the channel, i.e. the transmit correlation matrix $\mathbf{R}_T$, this can be used to adaptively change the transmission strategy according to $\mu_1, \ldots, \mu_{n_T}$. The transmit array performs adaptive power control $\mathbf{p}(\mu)$ and it can be shown that the ergodic capacity is given by the following optimisation problem

$$C_{cvCSI}(\mu_1, \ldots, \mu_{n_T}, \rho) = \max_{||\mathbf{p}||=1} \mathbb{E} \log_2 \left(1 + \rho \sum_{k=1}^{n_T} p_k \mu_k w_k \right). \tag{6.2}$$

The expression for the ergodic capacity of the MISO system with partial CSI in (6.2) directly corresponds to C_2 in (4.1). Finally, the impact of the transmit correlation on the ergodic capacity in (6.2) leads to Problem 3, i.e. to the result in Theorem 5.1. In [10], Theorem 4.1 and 5.1 have been applied. Interestingly, the behavior of the ergodic capacity in (6.2) is the other way round: it is a Schur-convex function with respect to $\boldsymbol{\mu}$, i.e. correlation increases the ergodic capacity.

7. NOTE ADDED IN PROOF

After submission of this paper, we found that the cumulative distribution function (cdf) of the sum of weighted exponential random variables in (1.1) has not the same clear behavior in terms of Schur-concavity like the function (3.1). In [3], we proved that the cdf $F(x) = Pr[\sum_{k=1}^n \mu_k w_k \leq x]$ is Schur-convex for all $x \leq 1$ and Schur-concave for all $x \geq 2$. Furthermore, the behavior of $F(x)$ between 1 and 2 is completely characterized: For $1 \leq x < 2$, there are at most two global minima which are obtained for $\mu_1 = ... = \mu_k = \frac{1}{k}$ and $\mu_{k+1} = ... = \mu_n = 0$ for a certain k. This result verifies the conjecture by Telatar in [15].

REFERENCES

[1] M. ABRAMOWITZ AND I.A. STEGUN, *Handbook of Mathematical Functions*, Dover Publications, 1970.

[2] H. BOCHE AND E.A. JORSWIECK, On the Schur-concavity of the ergodic and outage capacity with respect to correlation in multi-antenna systems with no CSI at the transmitter,*Proceedings of fortieth annual Allerton conference on communication, control, and computing*, 2002.

[3] H. BOCHE AND E.A. JORSWIECK, Outage probability of multiple antenna systems: Optimal transmission and impact of correlation, *Proc. IEEE International Zürich Seminar*, (2004), 116–119, in press.

[4] M.E. BOCK, P. DIACONIS, F.W. HUFFER AND M.D. PERLMAN, Inequalities for linear combinations of Gamma random variables, *Canad. J. Statistics*, **15** (1987), 387–395.

[5] S.P. BOYD AND L. VANDERBERGHE, Convex optimization, *Course reader for EE364 (Stanford) and EE236 (UCLA), and draft of a book that will be published in*, 2003.

[6] G.J. FOSCHINI AND M.J. GANS, On limits of wireless communications in a fading environment when using multiple antennas, *Wireless Personal Communications,* **6** (1998), 311–335.

[7] I.S. GRADSHTEYN AND I.M. RYZHIK, *Table of Integrals, Series and Products*, Academic Press, London, 5th Edition, 1994.

[8] W. HOEFFDING, On the distribution of the number of successes in independent trials, *Ann. Math. Statist.*, **27** (1956), 713–721.

[9] E.A. JORSWIECK AND H. BOCHE, Behaviour of outage probability in MISO systems with no channel state information at the transmitter, *Proc. of IEEE International Information Theory Workshop, Paris*, 2003.

[10] E.A. JORSWIECK AND H. BOCHE, Optimal power allocation for MISO systems and complete characterization of the impact of correlation on the capacity, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003.

[11] A.W. MARSHALL AND I. OLKIN, *Inequalities: Theory of Majorization and Its Application*, Academinc Press, 1979.

[12] M. MERKLE AND L. PETROVIC, On Schur-convexity of some distribution functions, *Publications de L'institut Mathématique,* **56**(70) (1994), 111–118.

[13] F. PROSCHAN, Peakedness of distribution of convex combinations, *Ann. Math. Statist.,* **36** (1965), 1703–1706.

[14] C.E. SHANNON, A mathematical theory of communication, *Bell System Technical Journal*, **27** (1948), 379–423 and 623–656.

[15] E. TELATAR, Capacity of multi-antenna Gaussian channels, *Bell Labs Journal*, **10**(6) (1999).