Water-filling is the Limiting Case of a General Capacity Maximization Principle

Anke Feiten
Institute of Theoretical Information Technology
RWTH Aachen University
Sommerfeldstr. 24, 52074 Aachen, Germany
Email: feiten@ti.rwth-aachen.de

Rudolf Mathar
Institute of Theoretical Information Technology
RWTH Aachen University
Sommerfeldstr. 24, 52074 Aachen, Germany
Email: mathar@ti.rwth-aachen.de

***Abstract*— [1] The optimal power allocation for Gaussian vector channels subject to sum power constraints is achieved by the well known water-filling principle. In this correspondence, we show that the discontinuous water filling solution is obtained as the limiting case of p-norm bounds on the power covariance matrix as p tends to one. Directional derivatives are the main vehicle leading to this result. An easy graphical representation of the solution is derived by the level crossing points of simple power functions, which in the limit $p = 1$ gives a nice dual view of the classical representation.**

I. Introduction

Capacity maximization of a vector channel by assigning power levels from a constrained set to subchannels is an important challenge when transmitting and receiving over multiple antennas. The purpose of this correspondence is to explicitly determine the optimum solution for a general family of constraints, namely bounding the ℓ_p-norm of the power covariance matrix. As the central result, the well known water-filling principle for sum power constraints turns out as the discontinuous limiting case of this family as $p \to 1$. Vice versa, the solution for maximum power constraints is also obtained as the limit $p \to \infty$. Cases in-between are covered as well, and may serve as an approximation for combined maximum and sum power constraints.

Directional derivatives are employed to determine the optimum solution. This method turns out to be very powerful allowing for characterizing the optimum point by a system of linear equations.

The general model we adopt is a linear vector channel with Gaussian noise and input distribution, i.e.,

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{n}. \tag{1}$$

$\boldsymbol{H}$ denotes the matrix of channel gains, $\boldsymbol{X}$ the t-dimensional Gaussian input vector, $\boldsymbol{n}$ the Gaussian noise vector, and $\boldsymbol{Y}$ the received vector at r receive antennas. In the following we assume complete channel state information in that $\boldsymbol{H}$ is known at the transmitter and the receiver.

Point-to-point multiple input multiple output (MIMO) antenna systems are covered by this model, see [1] and [2]. Furthermore, transmit beamforming, code division multiple access (CDMA) systems, and broadcast and general multiple-access channels are described by the above model.

The material in this correspondence is organized as follows. First, the general concept of directional derivatives is introduced in Section II. This section, furthermore, contains the precise system model and the directional derivative of mutual information. The corresponding capacity subject to p-norm constraints of the power covariance matrix is determined in Section III. Our results are briefly summarized in Section IV.

[1]Parts of the material in this correspondence are selected from a paper submission to IEEE Transactions on Information Theory in January 2006.

II. Directional Derivatives and System Model

We start with a short overview of the concept of directional derivatives and its application to the optimization of concave functions f with convex domain $\mathcal{C}$. Let $\hat{x}, x \in \mathcal{C}$. The directional derivative of f at $\hat{x}$ in the direction of x is defined as

$$\begin{aligned} Df(\hat{x}, x) &= \lim_{\alpha \to 0+} \frac{1}{\alpha} \Big[f\big((1-\alpha)\hat{x} + \alpha x\big) - f(\hat{x}) \Big] \\ &= \frac{d}{d\alpha} f\big((1-\alpha)\hat{x} + \alpha x\big)\Big|_{\alpha=0+}, \end{aligned} \tag{2}$$

see, e.g., [3]. Since f is concave, $\big(f((1-\alpha)\hat{x}+\alpha x) - f(\hat{x})\big)/\alpha$ is monotone increasing with decreasing $1 \geq \alpha \geq 0$, and the directional derivative always exists.

If $\mathcal{C}$ is a subset of a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, it is well known that

$$Df(\hat{x}, x) = \langle \nabla f(\hat{x}), x - \hat{x} \rangle, \tag{3}$$

whenever ∇f, the derivative of f in the strong sense, exists.

Optimum points are characterized by directional derivatives as follows, for a proof see [3].

Proposition 1: Let $\mathcal{C}$ be a convex set and $f : \mathcal{C} \to \mathbb{R}$ a concave function. Then the maximum of f is attained at $\hat{x}$ if and only if $Df(\hat{x}, x) \leq 0$ for all $x \in \mathcal{C}$.

Transmission over a channel with t transmit and r receive antennas is modeled by (1). The complex $r \times t$ matrix $\boldsymbol{H}$ describes the linear transformation undergone by the signal. The random noise vector $\boldsymbol{n} \in \mathbb{C}^r$ is circularly symmetric complex Gaussian distributed (see [2]) with expectation $\boldsymbol{0}$ and covariance matrix $\mathrm{E}(\boldsymbol{n}\boldsymbol{n}^*) = \boldsymbol{I}_r$. The complex zero mean input vector $\boldsymbol{X}$ is subject to power constraints described by

$$\mathrm{E}(\boldsymbol{X}\boldsymbol{X}^*) = \boldsymbol{Q} \in \mathcal{Q}$$

for some set of nonnegative definite complex matrices $\mathcal{Q}$.

bbThe general model (1) applies to many different communication systems. The most prominent ones are MIMO transmission systems with r receive antennas and t transmit antennas (see [4]). But also transmit beamforming, broadcast and multiple access channels, cellular CDMA radio, ad-hoc networks and digital wireline systems fall within the scope of model (1).

Following the arguments in [2] the capacity of the vector channel (1) is derived as the maximum of the mutual information over all admissible input distributions of $\boldsymbol{X}$ as

$$C = \max_{\boldsymbol{Q}\in\mathcal{Q}} I(\boldsymbol{X}, \boldsymbol{Y}) = \max_{\boldsymbol{Q}\in\mathcal{Q}} \log\det(\boldsymbol{I}_r + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*).$$

In the following we characterize the covariance matrix $\hat{\boldsymbol{Q}}$ which achieves capacity by using directional derivatives of the function

$$f : \mathcal{Q} \to \mathbb{R} : \boldsymbol{Q} \mapsto \log\det(\boldsymbol{I}_r + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*).$$

From Ky Fan's inequality it follows immediately that f is concave on the convex domain $\mathcal{Q}$.

Proposition 2: Let $\mathcal{Q}$ be convex and $\hat{\boldsymbol{Q}}, \boldsymbol{Q} \in \mathcal{Q}$. The directional derivative of f at $\hat{\boldsymbol{Q}}$ in the direction of $\boldsymbol{Q}$ is given by

$$Df(\hat{\boldsymbol{Q}}, \boldsymbol{Q}) = \mathrm{tr}\left(\boldsymbol{H}^*(\boldsymbol{I}_r + \boldsymbol{H}\hat{\boldsymbol{Q}}\boldsymbol{H}^*)^{-1}\boldsymbol{H}\,(\boldsymbol{Q} - \hat{\boldsymbol{Q}})\right). \quad (4)$$

The proof is given in [5]. It exploits the chain rule for real valued functions of some matrix $\boldsymbol{A}$ and the fact that $\frac{d}{d\boldsymbol{A}}\det\boldsymbol{A} = (\det\boldsymbol{A})(\boldsymbol{A}^{-1})^*$.

From (3) and Proposition 2 we also conclude that the strong derivative of f at $\hat{\boldsymbol{Q}}$ in the Hilbert space of all complex $t \times t$ matrices endowed with the inner product $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^*)$, see [6], p. 286, amounts to

$$\nabla f(\hat{\boldsymbol{Q}}) = \boldsymbol{H}^*(\boldsymbol{I}_r + \boldsymbol{H}\hat{\boldsymbol{Q}}\boldsymbol{H}^*)^{-1}\boldsymbol{H}. \quad (5)$$

III. Capacity for p-norm constraints

Achieving capacity for an appropriate power distribution means to maximize $f(\boldsymbol{Q})$ over the set of possible power assignments $\mathcal{Q}$. According to Proposition 1 the point $\hat{\boldsymbol{Q}}$ maximizes $f(\boldsymbol{Q})$ over some convex set $\mathcal{Q}$ if and only if $Df(\hat{\boldsymbol{Q}}, \boldsymbol{Q}) \leq 0$ for all $\boldsymbol{Q} \in \mathcal{Q}$. By (4) this leads to

$$\begin{aligned} &\mathrm{tr}\left(\boldsymbol{H}^*(\boldsymbol{I}_r + \boldsymbol{H}\hat{\boldsymbol{Q}}\boldsymbol{H}^*)^{-1}\boldsymbol{H}\,\boldsymbol{Q}\right) \\ &\leq \mathrm{tr}\left(\boldsymbol{H}^*(\boldsymbol{I}_r + \boldsymbol{H}\hat{\boldsymbol{Q}}\boldsymbol{H}^*)^{-1}\boldsymbol{H}\,\hat{\boldsymbol{Q}}\right) \end{aligned} \quad (6)$$

for all $\boldsymbol{Q} \in \mathcal{Q}$. Hence, we obtain the following proposition.

Proposition 3: $\max_{\boldsymbol{Q}\in\mathcal{Q}} f(\boldsymbol{Q})$ is attained at $\hat{\boldsymbol{Q}}$ if and only if $\hat{\boldsymbol{Q}}$ is a solution of

$$\max_{\boldsymbol{Q}\in\mathcal{Q}} \mathrm{tr}\left(\nabla f(\hat{\boldsymbol{Q}})\,\boldsymbol{Q}\right). \quad (7)$$

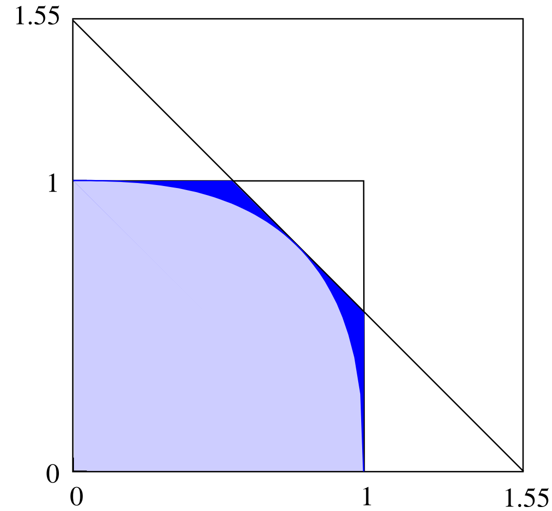

Fig. 1. The intersection of sum power constraints with $L_1 = 1.55$ and max power constraints with $L_2 = 1$ (the union of dark shaded and light shaded areas) approximated by the p-norm constraints with $L = 1$ and $p = 2.71$ (light shaded area).

The central theme in the following are power constraints by matrix p-norms. For a given $1 \leq p \leq \infty$ they are defined on the set of nonnegative Hermitian $t \times t$ matrices as

$$\|\boldsymbol{A}\|_p = \Big(\sum_{i=1}^{t} \lambda_i^p(\boldsymbol{A})\Big)^{1/p}, \quad (8)$$

where $\lambda_i(\boldsymbol{A})$, $i = 1, \ldots, t$, denote the eigenvalues of $\boldsymbol{A}$.

Sum power constraints are contained as the special case $p = 1$. Maximizing capacity here has the well known water-filling principle onto the inverse positive eigenvalues of $\boldsymbol{H}^*\boldsymbol{H}$ as a solution, cf. [2], [7], [8]. The opposite extreme $p = \infty$ corresponds to maximum eigenvalue constraints as $\lim_{p\to\infty} \|\boldsymbol{A}\|_p = \lambda_{\max}(\boldsymbol{A})$, the maximum eigenvalue of $\boldsymbol{A}$. The optimum solution in this case is a multiple of the identity matrix, cp. [5].

The practical motivation for dealing with arbitrary values $1 \leq p \leq \infty$ is to provide a fast numerical method to handle additional constraints, such as peak power constraints. The basic idea is to combine maximum and sum power constraints within a single restriction of the form (8).

Given the dimension t, sum power bound L_1 and normalized max power bound $L_2 = 1$, the optimum choice of p such the approximation error is minimal and neither the max nor the sum power constraints are exceeded is given by the solution of equation $\sum_{i=1}^{t}(L/t)^p = 1$ as

$$p^* = -\frac{\ln t}{\ln(L/t)}.$$

For visualization purposes we confine ourselves to diagonal matrices by considering only eigenvalues $\lambda_1, \ldots, \lambda_t$. Figure 1 shows for dimension $t = 2$ the intersection of sum power constraints with $L_1 = 1.55$ and max power constraints with $L_2 = 1$ (dark shaded jointly with light area) approximated by

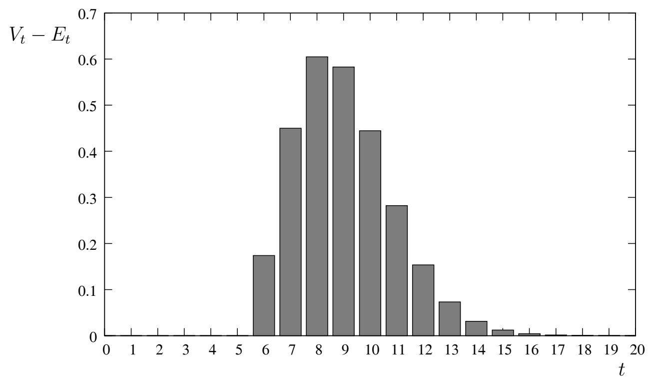

Fig. 2. The absolute error $V_t - E_t$ for $L = 5$.

the optimum p-norm constraints with $L = 1$ and $p^* = 2.71$ (light shaded area).

The volume V_t of the t-dimensional approximating power set $\{(\lambda_1, \dots, \lambda_t) \mid (\sum_{i=1}^t \lambda_i)^{1/t} \le L\}$ is

$$V_t = \frac{\Gamma^t(1 + 1/p^*)}{\Gamma(1 + t/p^*)},$$

where $\Gamma(x)$ denotes the gamma-function. The set combining peak and sum power constraints is given by $\{(\lambda_1, \dots, \lambda_t) \mid 0 \le \lambda_i \le 1, \sum_{i=1}^t \lambda_i \le L\}$. Its volume E_t is

$$E_t = \frac{1}{t!} \sum_{i=0}^{t} (-1)^i \binom{t}{i} (L - i)_+^t,$$

as can be derived from the convolution of uniformly distributed random variables, see [9, Theorem 1, p. 27]. Note that $(x)_+ = \max\{0, x\}$ denotes the positive part of x. The absolute error $V_t - E_t$ is visualized in Figure 2 for $L = 5$ and $t = 0, \dots, 20$. It can be seen that the absolute error is increasing first, and converges to zero with increasing t.

Observe that both E_t and V_t converge to zero with increasing dimension t.

From a mathematical point of view the present approach is conceptually simple, providing an elegant generalization of the concept of water-filling. Classical water-filling is obtained as the discontinuous limiting case as $p \to 1$ by using level crossing points of a class of simple power functions.

For general $1 \le p \le \infty$ and $L > 0$ the constraining set is given by

$$\mathcal{Q}_{p,L} = \{\boldsymbol{Q} \ge \boldsymbol{0} \mid \|\boldsymbol{Q}\|_p \le L\}.$$

The corresponding maximum in (7) can be explicitly determined as follows.

Proposition 4: Let $p, q \ge 1$ be conjugate, i.e., $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\max_{\boldsymbol{Q} \ge \boldsymbol{0},\, \|\boldsymbol{Q}\|_p \le L} \operatorname{tr}\big(\nabla f(\hat{\boldsymbol{Q}})\, \boldsymbol{Q}\big) = L\, \|\nabla f(\hat{\boldsymbol{Q}})\|_q \tag{9}$$

To see this we exploit that $\operatorname{tr}(\boldsymbol{AB}) \le \sum \lambda_{(i)}(\boldsymbol{A}) \lambda_{(i)}(\boldsymbol{B})$ for the ordered eigenvalues of nonnegative definite Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, see [10], H.1.g, p. 248. Together with Hölder's inequality and the fact that $\|\boldsymbol{Q}\| \le L$ over $\mathcal{Q}_{p,L}$ the following chain of inequalities is obtained

$$\begin{aligned}
&\max_{\boldsymbol{Q} \in \mathcal{Q}_{p,L}} \operatorname{tr}\big(\nabla f(\hat{\boldsymbol{Q}})\, \boldsymbol{Q}\big) \\
&\quad \le \max_{\boldsymbol{Q} \in \mathcal{Q}_{p,L}} \sum_{i=1}^{t} \lambda_{(i)}\big(\nabla f(\hat{\boldsymbol{Q}})\big)\, \lambda_{(i)}(\boldsymbol{Q}) \\
&\quad \le \Big(\sum_{i=1}^{t} \lambda_{(i)}^q \big(\nabla f(\hat{\boldsymbol{Q}})\big) \Big)^{1/q} \max_{\boldsymbol{Q} \in \mathcal{Q}_{p,L}} \Big(\sum_{i=1}^{t} \lambda_{(i)}^p (\boldsymbol{Q}) \Big)^{1/p} \\
&\quad \le L\, \|\nabla f(\hat{\boldsymbol{Q}})\|_q
\end{aligned}$$

Equality holds if $\lambda_{(i)}(\boldsymbol{Q}) = \alpha \lambda_{(i)}^{q-1}\big(\nabla f(\hat{\boldsymbol{Q}})\big)$, $\boldsymbol{Q}$ has the same system of unitary eigenvectors, and α is such that $\|\boldsymbol{Q}\|_p = L$. Hence, (9) follows.

Now, in combining Propositions 3 and 4, we get

Theorem 5: Let $p, q \ge 1$ be conjugate. Capacity, i.e., $\max_{\boldsymbol{Q} \in \mathcal{Q}_{p,L}} f(\boldsymbol{Q})$ is attained at power distribution $\hat{\boldsymbol{Q}} \in \mathcal{Q}_{p,L}$ if and only if

$$L\, \|\nabla f(\hat{\boldsymbol{Q}})\|_q = \operatorname{tr}\big(\nabla f(\hat{\boldsymbol{Q}})\, \hat{\boldsymbol{Q}}\big). \tag{10}$$

Once we can solve the above equation for $\hat{\boldsymbol{Q}}$, an optimum power allocation is found. For this purpose let

$$\boldsymbol{H} = \boldsymbol{U} \boldsymbol{\Gamma}^{1/2} \boldsymbol{V}^*$$

denote the singular value decomposition of the channel matrix $\boldsymbol{H}$. Let γ_i denote the identical positive eigenvalues of $\boldsymbol{HH}^*$ and $\boldsymbol{H}^*\boldsymbol{H}$, respectively, augmented by zeros whenever appropriate.

In the following we strive for finding a solution of (10) in the class of power allocations

$$\hat{\boldsymbol{Q}} = \boldsymbol{V} \operatorname{diag}(\hat{q}_1, \dots, \hat{q}_t) \boldsymbol{V}^*,\ \hat{q}_i \ge 0,\ \big(\sum_i \hat{q}_i^p\big)^{1/p} \le L.$$

The first step is to evaluate (10) for $\hat{\boldsymbol{Q}}$ of the above type. In [11] it is shown that the following representations hold.

$$L\, \|\nabla f(\hat{\boldsymbol{Q}})\|_q = L\, \Big(\sum_{i=1}^{t} \big(\frac{\gamma_i}{1 + \gamma_i \hat{q}_i}\big)^q \Big)^{1/q} \tag{11}$$

$$\operatorname{tr}\big(\nabla f(\hat{\boldsymbol{Q}})\, \hat{\boldsymbol{Q}}\big) = \sum_{i=1}^{t} \big(\frac{\gamma_i \hat{q}_i}{1 + \gamma_i \hat{q}_i}\big) \tag{12}$$

We first single out the case $p = \infty$ with $\|\hat{\boldsymbol{Q}}\|_\infty = \max_i \hat{q}_i$. Then, equality of (11) and (12) holds if $\hat{q}_i = L$ for all $i = 1, \dots, t$ with $\gamma_i > 0$, and $\hat{q}_i = 0$ otherwise. Note that for $\gamma_i = 0$ any other $q_i \in [0, L]$ ensures equality and yields an admissible solution as well.

In the case $p = 1$ let $\hat{q}_i = (\nu - 1/\gamma_i)^+$, ν such that $\sum_{i=1}^t \hat{q}_i = L$. Some algebra shows that in this case either (11) and (12) have the same value L/ν and hence are equal.

If in general $1 < p < \infty$ and for some $\nu > 0$ it holds that

$$\frac{\nu \gamma_i}{1 + \gamma_i \hat{q}_i} = \hat{q}_i^{p-1} \tag{13}$$

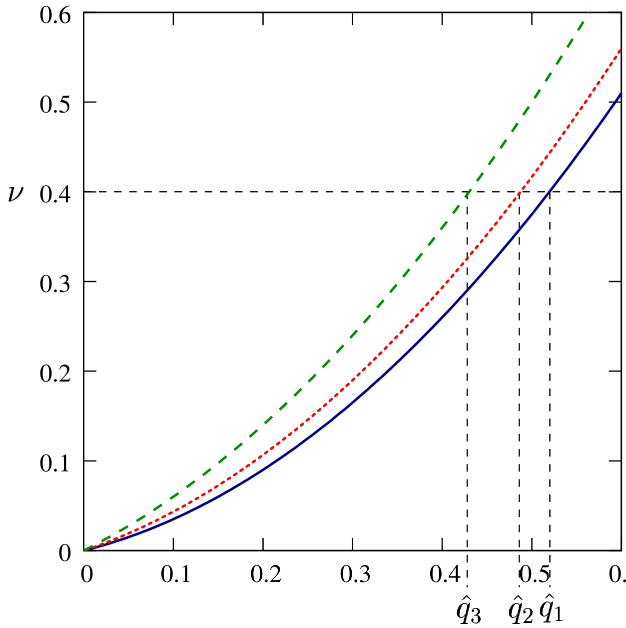

Fig. 3. The curves $\hat{q}_i^p + \hat{q}_i^{p-1}/\gamma_i$ for $p = 2$, $\gamma_1 = 4$ (solid), $\gamma_2 = 3$ (dotted), $\gamma_3 = 2$ (dashed). $\nu = 0.4$ corresponds to the optimum power assignments indicated by $\hat{q}_i$ on the x-axis.

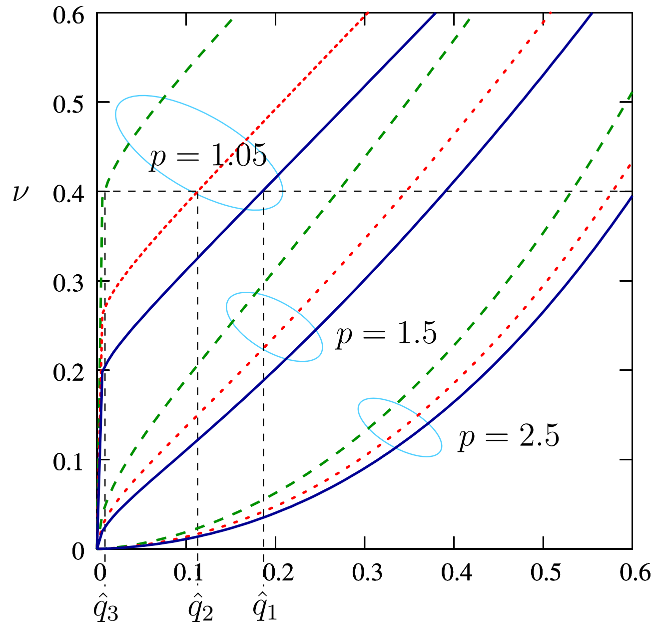

Fig. 4. Visualizing the limiting case $\hat{q}_i^p + \hat{q}_i^{p-1}/\gamma_i$ as $p \to 1$. γ-values are $\gamma_1 = 4$ (solid), $\gamma_2 = 3$ (dotted), $\gamma_3 = 2$ (dashed). $\nu = 0.4$ leads to corresponding $\hat{q}_i$ indicated on the x-axis.

for all $i = 1, \ldots, t$ with $\gamma_i > 0$ and $\hat{q}_i = 0$ otherwise, then (11) equals (12). This can be readily seen from Hölder's inequality, since

$$\sum_{i=1}^{t} \Big(\frac{\gamma_i \hat{q}_i}{1+\gamma_i \hat{q}_i}\Big) \le \Big(\sum_{i=1}^{t} \hat{q}_i^p\Big)^{1/p} \Big(\sum_{i=1}^{t} \big(\frac{\gamma_i}{1+\gamma_i \hat{q}_i}\big)^q\Big)^{1/q}$$

and since $\big(\sum_i \hat{q}_i^p\big)^{1/p} = L$. Equality is attained whenever (13) holds.

For positive γ_i equation (13) can equivalently be written as $\hat{q}_i^p + \frac{1}{\gamma_i}\hat{q}_i^{p-1} = \nu$ such that in summary we have proved the following central result.

Theorem 6: For $1 < p < \infty$ let $\hat{q}_i \ge 0$, $i = 1, \ldots, t$, denote the unique solution of the system of equations

$$\begin{aligned} &\hat{q}_i = 0, \ \text{if } \gamma_i = 0, \\ &\hat{q}_i^p + \frac{1}{\gamma_i}\,\hat{q}_i^{p-1} = \nu, \ \text{if } \gamma_i > 0, \\ &\nu \text{ such that } \Big(\sum_{i=1}^{t} \hat{q}_i^p\Big)^{1/p} = L, \end{aligned} \tag{14}$$

For the limiting case $p = 1$, it holds that

$$\hat{q}_i = \big(\nu - \frac{1}{\gamma_i}\big)^+, \quad \nu \text{ such that } \sum_{i=1}^{t} \hat{q}_i = L, \tag{15}$$

and if $p = \infty$ let $\hat{q}_i = L$ for all $i = 1, \ldots, t$ with $\gamma_i > 0$, and $\hat{q}_i = 0$ otherwise.

Then, for any $1 \le p \le \infty$,

$$\hat{\boldsymbol{Q}} = \boldsymbol{V} \operatorname{diag}(\hat{q}_1, \ldots, \hat{q}_t) \boldsymbol{V}^*$$

is a solution of

$$\max_{\boldsymbol{Q} \ge \boldsymbol{0},\ \|\boldsymbol{Q}\|_p \le L} \log\det(\boldsymbol{I} + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*),$$

and hence represents an optimal power assignment.

The function $q_i^p + \frac{1}{\gamma_i} q_i^{p-1}$ is monotone in q_i for any $p > 1$ such that a solution of (14) always exists for any $L > 0$. Observe that except for the case $p = 1$ all positive eigenvalues γ_i receive a positive amount of power allocated.

A graphical solution of Proposition 6 is represented in Figure 3. The solid, dotted and dashed line correspond to values $p = 2$ and $\gamma_1 = 4$, $\gamma_2 = 3$, $\gamma_3 = 2$. ν is set to 0.4. The optimum arguments can be read from the x-axis as 0.52, 0.48, 0.43, respectively.

The well known water-filling solution (15) corresponding to $p = 1$ is obtained as a special limiting case of (14). Let for positive γ_i

$$g_i(x) = x^p + \frac{1}{\gamma_i}\, x^{p-1}, \quad x \ge 0,$$

denote the functions defined in (14). It holds that

$$\lim_{p \to 1} g_i(x) = \begin{cases} 0, & \text{if } x = 0 \\ x + 1/\gamma_i, & \text{if } x > 0 \end{cases}$$

This fact is readily accessible by the curves plotted in Figure 4 with p-values set to $p = 2.5$, 1.5, 1.05 (from right to left). The discontinuous limit functions are shown in Figure 5. The level crossing points for $\nu = 0.4$ lead to the power levels $\hat{q}_1 = 0.15$, $\hat{q}_2 = 0.067$, and $\hat{q}_3 = 0$. They coincide with the water-filled values to the level ν, as can be seen from the conventional representation on the left.

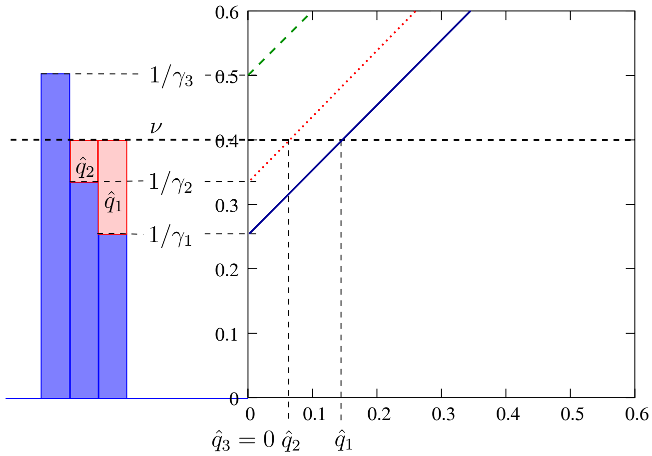

Fig. 5. Visualizing the limit case $p = 1$. γ-values are $\gamma_1 = 4$ (solid), $\gamma_2 = 3$ (dotted), $\gamma_3 = 2$ (dashed). $\nu = 0.4$ leads to the optimum water-filling solution $\hat{q}_1 = 0.15$, $\hat{q}_2 = 0.067$, $\hat{q}_3 = 0$, indicated on the x-axis.

IV. Conclusion

The main purpose of this correspondence is to show that the well known water-filling solution for sum power constraints is derived as a discontinuous limiting case from a more general optimization principle over p-norm constraints. We have employed directional derivatives as a powerful tool for solving the general optimization problem by solving the corresponding linear stationary equations. In summary, we have achieved a general framework for optimal power allocation which extends classical approaches in a natural and unifying way.

Acknowledgment

This work was supported by DFG grant Ma 1184/11-3.

References

[1] G. Foschini and M. Gans, “On limits of wireless communication in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.
[2] I. E. Telatar, “Capacity of multi-antenna gaussian channels,” *European Transactions on Telecommuncations*, vol. 10, no. 6, pp. 585–595, 1999.
[3] P. Whittle, “Some general points in the theory of optimal experimental design,” *J. Roy. Statist. Soc. B*, vol. 35, pp. 123–130, 1973.
[4] E. Biglieri and G. Taricco, *Transmission and Reception with Multiple Antennas: Theoretical Foundations*. Delft: now Publishers, 2004.
[5] A. Feiten, S. Hanly, and R. Mathar, “Derivatives of mutual information in Gaussian vector channels with applications,” Institute of Theoretical Information Technology, RWTH Aachen University, Tech. Rep., 2006.
[6] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia: SIAM, 2000.
[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Whiley, 1991.
[8] R. G. Gallager, *Information Theory and Reliable Communications*. New York: John Wiley & Sons Ltd., 1968.
[9] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons, 1966.
[10] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.
[11] A. Feiten, R. Mathar, and S. Hanly, “Eigenvalue based optimum power allocation for Gaussian vector channels,” Institute of Theoretical Information Technology, RWTH Aachen University, Tech. Rep., 2006.