Optimum Transmit Architecture of a MIMO System under Modulus Channel Knowledge at the Transmitter

Miquel Payaró[1], Xavier Mestre, and Miguel Ángel Lagunas
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
c/ Gran Capità 2 - 4 Rooms 202 - 203
08034 Barcelona, Spain
e-mail: {miquel.payaro, xavier.mestre, m.a.lagunas}@cttc.es

Abstract — **In this paper, we study the ergodic capacity of a multiple input multiple output (MIMO) uncorrelated flat fading channel with perfect channel state information at the receiver and partial channel state information at the transmitter. We focus our attention on the case where the transmitter is informed only with the modulus of the channel matrix coefficients. First, we prove that a simple power allocation strategy among transmitting antennas is the optimal scheme, in the sense that is a capacity achieving architecture. Next, for the particular case where only two antennas are used at each communication end, we derive closed form expressions for the ergodic capacity and the optimal power assigned to each antenna.**

I. Introduction

The ergodic capacity of a multiple input multiple output (MIMO) communications system over a frequency nonselective fading channel with perfect channel state information (CSI) at the receiver has been a topic that has received a lot of attention in the recent literature. In his popular work [1], Telatar stated that, when the transmitter has no CSI, uniform power allocation is the most reasonable approach to achieve capacity. The optimality of the uniform power allocation scheme was later proven in [2] using game-theoretic justifications. For the case where the transmitter has perfect CSI, it is also well known [1],[3] that the optimum transmit strategy consists in splitting the signal among a set of beamformers giving each one a fraction of transmission power according to a "waterfilling" algorithm.

The ergodic capacity of MIMO systems with partial CSI at the transmitter side has also been the focus of investigation of numerous researchers. For example, in [4],[5],[6] a multiple input single output (MISO) architecture is considered, and it is assumed that the transmitter has only access to either the mean value or the covariance of the channel vector. The capacity can be achieved by using a Gaussian codebook with a particular covariance matrix which, in the case of channel covariance feedback, has the same eigenvectors as the true channel covariance matrix. The structure of the corresponding eigenvalues is further investigated in [7]. In [8], the effect of partial side information at the transmitter is analyzed for a generic MIMO case. One of the most important conclusions of their paper is that capacity can be achieved by a structure that first maps source symbols into space-time codewords independently from the CSI, and then weights these codewords as a function of the CSI at the transmitter.

In our paper, a different kind of partial CSI at the transmitter side is considered: the transmitter has perfect knowledge of the modulus of the complex channel coefficients but it does not have access to the channel phases. Many interesting practical situations fit in this model. In time division duplex (TDD) mobile communication systems, for example, if electromagnetic reciprocity between uplink and downlink channels is exploited, then perfect knowledge of the uplink channel yields perfect knowledge of the downlink response. In practice, however, the channel reciprocity is broken because uplink and downlink channels are seen through different radio-frequency (RF) front-ends. To restore reciprocity, calibration architectures that compensate for the amplitude and phase response of each RF front-end must be used. Although calibration of the amplitude response is relatively easy, phase response calibration becomes substantially more complicated (due, for example, to different lengths of the tracks between the oscillator and the different mixers). Hence, in such architectures only perfect knowledge of the channel amplitude can be reasonably assumed.

II. System Model

We consider a narrowband multiplexing system with n_T transmit and n_R receive antennas corrupted with additive white Gaussian noise. Let us define $\mathbf{x} \in \mathbb{C}^{n_T}$ as the transmit signal vector, $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$, as the channel matrix, and $\mathbf{n} \in \mathbb{C}^{n_R}$, as the noise vector, modeled as a circularly symmetric Gaussian distributed random vector with zero mean and covariance matrix $\mathbb{E}\left\{\mathbf{n}\mathbf{n}^H\right\} = \sigma^2 \mathbf{I}$, which is supposed known at the transmitter[2]. The received signal, $\mathbf{y} \in \mathbb{C}^{n_R}$, for this model can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = [\mathbf{h}_1 \mathbf{h}_2 \cdots \mathbf{h}_{n_T}]\, \mathbf{x} + \mathbf{n}. \quad (1)$$

As it was stated in the introduction, we assume that, while $\mathbf{H}$ is perfectly known at the receiver, the transmitter only has access to the modulus of the entries of $\mathbf{H}$. To separate the known from the unknown part of the elements of $\mathbf{H}$, we define

$$\mathbf{H}(\boldsymbol{\Theta}) \equiv \mathbf{M} \odot \mathbf{P}(\boldsymbol{\Theta}),$$

where $[\mathbf{P}(\boldsymbol{\Theta})]_{kl} = e^{\mathrm{j}[\boldsymbol{\Theta}]_{kl}}$ and $\odot$ denotes the Hadamard element-wise matrix product. Following the notation introduced above, the gain between antennas l and k is $h_{kl} = [\mathbf{H}]_{kl} = [\mathbf{M}]_{kl} e^{\mathrm{j}[\boldsymbol{\Theta}]_{kl}} = m_{kl} e^{\mathrm{j}\theta_{kl}}$, where $m_{kl} \in \mathbb{R}^+ \cup \{0\}$, $\forall k, l$. The unknown channel phases, θ_{kl}, are modeled as i.i.d. uniform random variables in $[-\pi, \pi)$ as suggested in, *e.g*, [9].

[1]This work was partially supported by the Spanish Ministry of Science and Technology (MCYT) under projects TIC2002-04594-C02-02 (jointly financed by FEDER) and FIT-070000-2003-257, the European Commission IST-2004-508009, IST-2004-507325, and the Generalitat de Catalunya 2003FI 00195.

[2]In practical implementations σ^2 can be fedback to the transmitter through a low rate channel as σ^2 varies slowly in time.

III. Optimum Transmit Architecture

In [8], the authors showed that the ergodic capacity in the situation described in last section can be achieved if $\mathbf{x}$ is drawn from a circularly symmetric Gaussian distribution with zero mean and covariance $\mathbf{Q}$. The resulting capacity can be obtained by maximizing the ergodic mutual information with respect to $\mathbf{Q}$, *i.e.*

$$C = \sup_{\mathbf{Q}} I(\mathbf{Q}) = \sup_{\mathbf{Q}} \mathbb{E}_{\boldsymbol{\Theta}} \log\det\left(\mathbf{I} + \mathbf{H}(\boldsymbol{\Theta})\mathbf{Q}\mathbf{H}^H(\boldsymbol{\Theta})\right), \quad (2)$$

where the maximization is over the set of positive semidefinite hermitian matrices such that $\operatorname{tr}\mathbf{Q} = P_T\sigma^{-2}$, with P_T being the total transmitted power. The expectation is taken over the distribution of the phases of the entries of $\mathbf{H}$. Next proposition particularizes (2) for a particular case of phase uncertainty at the transmitter.

Proposition 1 *Assume instantaneous perfect channel state information at the receiver and moduli channel state information at the transmitter,* $\mathbf{M}$. *If the channel state, defined as* $\mathbf{H}(\boldsymbol{\Theta}) \equiv \mathbf{M} \odot \mathbf{P}(\boldsymbol{\Theta})$, *is a random matrix whose elements' phases,* $[\boldsymbol{\Theta}]_{kl}$, *are i.i.d. random variables drawn from a uniform distribution in interval* $[-\pi, \pi)$, *then the ergodic capacity of the Gaussian memoryless channel (1), subject to a transmit power constraint* P_T, *is given by:*

$$\begin{aligned} C = \sup_{\boldsymbol{\Lambda}} I(\boldsymbol{\Lambda}) &= \sup_{\boldsymbol{\Lambda}} \mathbb{E}_{\boldsymbol{\Theta}} \log\det\left(\mathbf{I} + \mathbf{H}(\boldsymbol{\Theta})\boldsymbol{\Lambda}\mathbf{H}^H(\boldsymbol{\Theta})\right), \\ & \text{s.t. } \operatorname{tr}\boldsymbol{\Lambda} = P_T\sigma^{-2} \equiv \gamma, \quad (3) \end{aligned}$$

where $\boldsymbol{\Lambda}$ *is a positive definite diagonal matrix.*

Proof. See Appendix A. □

Notice that the diagonal structure of $\boldsymbol{\Lambda}$ implies that the optimal transmission scheme is to perform independent signaling over the transmission antennas (*i.e.* power allocation). In the following sections we will continue our capacity analysis for the particular case of a MIMO channel with $n_R = n_T = 2$.

IV. Ergodic mutual information

As a conclusion from what is proven in the last section, we can state that the capacity of a 2×2 MIMO system with phase uncertainty at the transmitter can be characterized with the following maximization:

$$C = \sup_{\substack{\lambda_1,\lambda_2 \geq 0 \\ \lambda_1+\lambda_2=\gamma}} I(\lambda_1, \lambda_2),$$

where $I(\lambda_1, \lambda_2)$ is the same as defined in (3) with the positive weights λ_1 and λ_2 denoting the diagonal elements of $\boldsymbol{\Lambda}$. Equivalently, $I(\lambda_1, \lambda_2)$ can be expressed as

$$I(\lambda_1, \lambda_2) = \mathbb{E}_{\boldsymbol{\Theta}} \log\det\left(\mathbf{I} + \lambda_1\mathbf{h}_1\mathbf{h}_1^H + \lambda_2\mathbf{h}_2\mathbf{h}_2^H\right), \quad (4)$$

where the dependence of $\mathbf{h}_i$, $i = 1, 2$, on $\boldsymbol{\Theta}$ has not been written explicitly. For the sake of simplicity we define,

$$\begin{aligned} T &= 2\lambda_1\lambda_2 m_{11}m_{12}m_{21}m_{22}, & (5)\\ S &= 1 + \lambda_1\left(m_{11}^2 + m_{21}^2\right) + \lambda_2\left(m_{12}^2 + m_{22}^2\right) + \\ &\quad +\lambda_1\lambda_2\left(m_{11}^2m_{22}^2 + m_{12}^2m_{21}^2\right), & (6)\\ \phi &= \theta_{21} - \theta_{11} + \theta_{12} - \theta_{22}. & (7) \end{aligned}$$

Notice that ϕ is also uniformly distributed in $[-\pi, \pi)$. With the last definitions and expanding the determinant in (4), $I(\lambda_1, \lambda_2)$ can be expressed as

$$I(\lambda_1, \lambda_2) = \log S + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(1 - \frac{T\cos\phi}{S}\right) \mathrm{d}\phi. \quad (8)$$

This last integral can be solved using [10, p.526]

$$\frac{1}{\pi}\int_0^{\pi} \log\left(1 + \beta\cos x\right) \mathrm{d}x = \log\left(\tfrac{1}{2}\left(1 + \sqrt{1-\beta^2}\right)\right).$$

Thus, the mutual information in (8) can be expressed as

$$I(\lambda_1, \lambda_2) = \log\left(\tfrac{1}{2}\left(S + \sqrt{S^2 - T^2}\right)\right). \quad (9)$$

At this point, we can characterize the capacity of the system and the optimum power allocation scheme by carrying out the maximization of (9) with respect to the positive weights λ_1, λ_2 under the constraint $\lambda_1 + \lambda_2 = \gamma$. In the next section we state formally this optimization problem and characterize its optimum solution.

V. Power Allocation

Let us now investigate the design of the power weights corresponding to the two transmit antennas (or, up to a constant, λ_1 and λ_2) in order to achieve capacity. Recalling the main result in the last section, the optimization problem can be reformulated as follows,

$$\begin{aligned} \{\lambda_1, \lambda_2\} &= \arg\sup_{\lambda_1,\lambda_2} I(\lambda_1, \lambda_2) \quad (10)\\ &\quad \text{s.t. } \lambda_1 + \lambda_2 = \gamma,\ \lambda_i \geq 0. \end{aligned}$$

We make the slight abuse of notation and denote the optimization variables and their optimum values with the same symbol. It is easy to see that this optimization problem is convex. In order to investigate the properties of the optimum solution, let us first formulate the Lagrangian of the corresponding minimization problem,

$$\mathcal{L} = -S - \sqrt{S^2 - T^2} - \mu_1\lambda_1 - \mu_2\lambda_2 + \nu\left(\lambda_1 + \lambda_2 - \gamma\right).$$

Imposing the Karush-Kuhn-Tucker (KKT) conditions, we see that, the following are necessary and sufficient conditions for λ_1, λ_2 to be the global maximum of (10)

$$\begin{aligned} \lambda_i &\geq 0, \\ \mu_i &\geq 0, \\ \lambda_i\mu_i &\geq 0, \\ \lambda_1 + \lambda_2 &= \gamma, \\ \frac{\mathrm{d}}{\mathrm{d}\lambda_i}\left(S + \sqrt{S^2 - T^2}\right) + \mu_i &= \nu, \end{aligned}$$

for $i = 1, 2$. Now, observe that μ_1 and μ_2 are just slack variables that can be eliminated, leaving

$$\begin{aligned} &\nu \geq \frac{\mathrm{d}}{\mathrm{d}\lambda_i}\left(S + \sqrt{S^2 - T^2}\right), \\ &\left(\nu - \frac{\mathrm{d}}{\mathrm{d}\lambda_i}\left(S + \sqrt{S^2 - T^2}\right)\right)\lambda_i = 0, \quad (11) \end{aligned}$$

again for $i = 1, 2$. Let us now analyze the different situations regarding power allocation.

A. Situation 1: One transmit antenna is switched off.

Assume that $\lambda_2 = 0$ ($\lambda_1 = \gamma$) so that the second antenna is switched off. Our objective here is to find necessary and sufficient conditions for this situation. Let us first concentrate on the necessary conditions. Obviously, from (11) one must have,

$$\nu = \frac{\mathrm{d}}{\mathrm{d}\lambda_1}\left(S + \sqrt{S^2 - T^2}\right)$$

and consequently

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_1}\left(S + \sqrt{S^2 - T^2}\right) \geq \frac{\mathrm{d}}{\mathrm{d}\lambda_2}\left(S + \sqrt{S^2 - T^2}\right).$$

This is clearly our candidate condition for sufficiency, and can alternatively be written as

$$\gamma \leq \frac{\left(m_{11}^2 + m_{21}^2\right) - \left(m_{12}^2 + m_{22}^2\right)}{m_{11}^2 m_{22}^2 + m_{21}^2 m_{12}^2} \equiv \xi. \tag{12}$$

From this point, we see that the necessary and sufficient conditions for transmitting with the first antenna are (12) together with $\left(m_{11}^2 + m_{21}^2\right) > \left(m_{12}^2 + m_{22}^2\right)$. By the symmetry of the problem, one can also see that, whenever $\left(m_{11}^2 + m_{21}^2\right) < \left(m_{12}^2 + m_{22}^2\right)$ and $\gamma \leq -\xi = |\xi|$, the first antenna will be switched off and the whole power will be allocated to the second one (i.e. $\lambda_1 = 0$, $\lambda_2 = \gamma$).

B. Situation 2: Both antennas are active.

Without loss of generality, we assume here that $\xi > 0$ (the case $\xi < 0$ can readily be studied exploiting the inherent symmetry of the problem, and the case $\xi = 0$ leads to the trivial solution $\lambda_1 = \lambda_2 = \gamma/2$). We define $\lambda_1 = \lambda$ and then $\lambda_2 = \gamma - \lambda$. The last KKT conditions tell us that

$$\nu = \frac{\mathrm{d}}{\mathrm{d}\lambda_1}\left[S + \sqrt{S^2 - T^2}\right] = \frac{\mathrm{d}}{\mathrm{d}\lambda_2}\left[S + \sqrt{S^2 - T^2}\right].$$

This condition can alternatively be written as

$$\frac{S}{T} + \sqrt{\frac{S^2}{T^2} - 1} = \frac{2m_{11}m_{12}m_{21}m_{22}}{m_{11}^2 m_{22}^2 + m_{21}^2 m_{12}^2}\frac{\gamma - 2\lambda}{\xi + \gamma - 2\lambda}. \tag{13}$$

From this equation, the expression of the ergodic mutual information in (9), and the fact that $T > 0$ by its definition in (5), one readily sees that $(\gamma - 2\lambda)$ and $(\xi + \gamma - 2\lambda)$ must always have the same sign (otherwise, the logarithm could not be properly defined). In addition, it can be stated that $\gamma - 2\lambda < 0$, because if the contrary was assumed it would lead to $\sqrt{S^2 - T^2} < 0$. Thus, since $(\gamma - 2\lambda)$ and $(\xi + \gamma - 2\lambda)$ have the same sign, it implies $\lambda > (\gamma + \xi)/2$, *i.e.* the optimum power allocation algorithm gives more power to the best antenna (namely the first one whenever $\xi > 0$).

Note that, up to now, we have only proven that the optimum λ is such that $(\gamma + \xi)/2 < \lambda < \gamma$. Next, we give a more specific characterization of the optimum transmit power allocation. Our objective now is to obtain a closed form expression for the optimum power allocation strategy in the case where the two antennas are active. So, we define

$$\begin{aligned} x &= 2\lambda - \gamma, \qquad (14)\\ \Delta &= \frac{m_{11}^2 m_{22}^2 - m_{21}^2 m_{12}^2}{m_{11}^2 m_{22}^2 + m_{21}^2 m_{12}^2},\\ \Upsilon &= \frac{2 + \gamma\left(m_{11}^2 + m_{21}^2\right) + \gamma\left(m_{12}^2 + m_{22}^2\right)}{\left(m_{11}^2 + m_{21}^2\right) - \left(m_{12}^2 + m_{22}^2\right)}. \end{aligned}$$

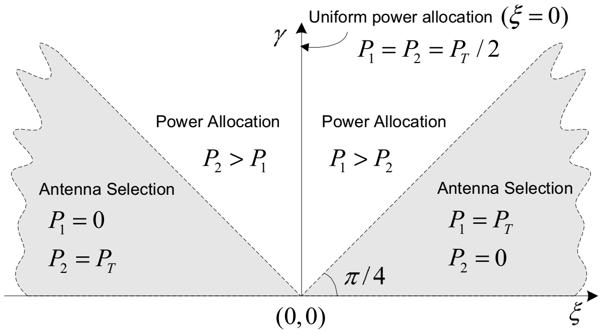

Fig. 1: Graphical scheme of the optimal power allocation as a function of γ and ξ

Using basic algebraic manipulations, one can show that (13) can be reformulated (adding, of course, new solutions) as:

$$\Delta^2 x^4 - 4\xi x^3 + \left[(3\xi - 4\Upsilon)\xi - \Delta^2\gamma^2\right]x^2 + 4\Upsilon\xi^2 x + \gamma^2\xi^2 = 0. \tag{15}$$

The power assignment can be determined finding the root of the fourth order polynomial in (15) in the range $x \in (\xi, \gamma)$. Note that there exist closed form expressions for the roots of a fourth order polynomial, and therefore the optimum power allocation strategy can be derived in exact form (no root-searching procedures are needed). In Appendix C we prove that there is a single root of the polynomial (15) located on $x \in (\xi, \gamma)$. Thereby, using (14), the optimal λ can be found.

C. Summary: Optimum power allocation scheme.

Let us now briefly summarize the optimum power allocation policy in a 2×2 MIMO system under phase uncertainty at the transmitter. We use P_1 and P_2 to denote the power allocated to the first and second antennas respectively (i.e. $P_1 = \sigma^2\lambda$, $P_2 = \sigma^2(\gamma - \lambda)$.

- If $\gamma < |\xi|$, then the best transmit antenna should be selected (namely $P_1 = P_T$, $P_2 = 0$ if $\xi > 0$ and vice versa if $\xi < 0$).
- If $\gamma \geq |\xi|$, and $\xi > 0$, then $P_1 = \sigma^2\lambda = (P_T + \sigma^2 x)/2$, where x is the unique root of the fourth order polynomial (15) in the region $\xi < x < \gamma$. If $\xi < 0$, an equivalent characterization can be given simply by swapping the role of antenna 1 and antenna 2.

In Fig. 1, the optimum power allocation scheme above described is depicted as a function of ξ and γ.

VI. Conclusions

For the case where the transmitter has knowledge only about the modulus of the channel matrix coefficients and the noise parameter σ^2, we have proven that the optimal transmission scheme for a flat fading MIMO channel, whose phases are independent and uniformly distributed in $[-\pi, \pi)$, is to perform independent signaling over the transmission antennas, *i.e.* capacity is always achieved with power allocation. In addition, we have derived closed form expressions for the ergodic capacity and the optimal power assigned to each antenna for the case of $n_R = n_T = 2$. It is also important to state that the optimum power weights depend nontrivially on the modulus of the channel coefficients.

A. Proof of proposition 1

In Section III we stated that the ergodic capacity was given by

$$C = \sup_{\mathbf{Q}} I(\mathbf{Q}) = \sup_{\mathbf{Q}} \mathbb{E}_{\mathbf{\Theta}} \log\det\left(\mathbf{I} + \mathbf{H}(\mathbf{\Theta})\mathbf{Q}\mathbf{H}^H(\mathbf{\Theta})\right).$$

In the following, the explicit dependence of $\mathbf{H}$ on $\mathbf{\Theta}$ will be omitted for the sake of notation. Our objective here is to show that the maximization with respect to $\mathbf{Q}$ can be restricted to the set of diagonal positive semidefinite matrices. We will follow the same procedure as in [4].

Let us first consider the following optimization problem above, restricting the maximization to the set of positive semidefinite diagonal matrices such that $\operatorname{tr}\mathbf{Q} = P_T\sigma^{-2}$. The optimization function is strictly concave in $\mathbf{Q}$, while the constraint set is convex[1] and compact. This implies that there exists a unique $\mathbf{\Lambda}$, diagonal and positive semidefinite, that solves the restricted optimization problem. We now show that this $\mathbf{\Lambda}$ is actually the solution of the first optimization problem. Indeed, a necessary and sufficient condition for the overall optimality of $\mathbf{\Lambda}$ is $dI(\mathbf{\Lambda};\mathbf{Q}-\mathbf{\Lambda}) \le 0$, where $dI(\mathbf{A};\mathbf{B})$ is the Fréchet differential of I in the direction of $\mathbf{B}$ evaluated at $\mathbf{A}$ (see further [11]). This condition is equivalent to:

$$\mathbb{E}_{\mathbf{\Theta}}\operatorname{tr}\left[\mathbf{H}^H\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{H}\left(\mathbf{Q}-\mathbf{\Lambda}\right)\right] \le 0.$$

Let $\mathbf{\Lambda_Q}$ be a diagonal matrix containing the diagonal entries of $\mathbf{Q}$, and $\tilde{\mathbf{Q}} = \mathbf{Q}-\mathbf{\Lambda_Q}$. The above condition can be expressed as

$$\begin{aligned}&\mathbb{E}_{\mathbf{\Theta}}\operatorname{tr}\left[\mathbf{H}^H\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{H}\left(\mathbf{\Lambda_Q}-\mathbf{\Lambda}\right)\right]\\ &+\mathbb{E}_{\mathbf{\Theta}}\operatorname{tr}\left[\mathbf{H}^H\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{H}\tilde{\mathbf{Q}}\right] \le 0. \qquad (16)\end{aligned}$$

Notice that the first term is always negative or zero, thanks to the optimality of $\mathbf{\Lambda}$ ($\operatorname{tr}\mathbf{\Lambda_Q} = \operatorname{tr}\mathbf{Q} = P_T\sigma^{-2}$, so that the constraint is fulfilled). As for the second, it can be written as

$$\begin{aligned}&\mathbb{E}_{\mathbf{\Theta}}\operatorname{tr}\left[\mathbf{H}^H\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{H}\tilde{\mathbf{Q}}\right] =\\ &= \sum_{i=1}^{n_T}\sum_{\substack{k=1\\k\neq i}}^{n_T} q_{ik}\mathbb{E}_{\mathbf{\Theta}}\left[\mathbf{h}_i^H\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{h}_k\right].\end{aligned}$$

From symmetry considerations, notice that it is sufficient to show that, for $1 < i \le n_T$,

$$\mathbb{E}_{\mathbf{\Theta}}\left[\mathbf{h}_i^H\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{h}_1\right] = 0 \qquad (17)$$

in order to prove that the second term in (16) is identically zero. With this, we will have proven that a diagonal covariance matrix $\mathbf{Q}$ is the optimum structure to attain capacity. Equation (17) can be alternatively written as:

$$\frac{1}{J}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\mathbf{h}_i^H\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{h}_1 \mathrm{d}\theta_{11}\ldots\mathrm{d}\theta_{n_Rn_T} = 0,$$

where $J = (2\pi)^{n_Rn_T}$. This is equivalent to

$$\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\mathbf{h}_i^H\mathbf{g}\mathrm{d}\theta_{12}\ldots\mathrm{d}\theta_{n_Rn_T} = 0, \qquad (18)$$

with

$$\mathbf{g} = \int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\left(\mathbf{I}+\mathbf{H\Lambda H}^H\right)^{-1}\mathbf{h}_1\mathrm{d}\theta_{11}\ldots\mathrm{d}\theta_{n_R1}.$$

Defining $\mathbf{f} = (\mathbf{I}+\mathbf{H\Lambda H}^H)^{-1}\mathbf{h}_1$, it can be noticed that each element of vector function $\mathbf{f} = [f_1\ldots f_{n_R}]^T$ can be considered as a function of arguments $(\theta_{11},\ldots,\theta_{n_R1})$ with the remaining $(\theta_{12},\ldots,\theta_{n_Rn_T})$ considered constant parameters, *i.e.* $f_k : \mathbb{R}^{n_R} \to \mathbb{C}$. Each f_k, $1 \le k \le n_R$, is a measurable function and has the following properties:

$$\begin{aligned}&f_k(\theta_{11}+\pi,\ldots,\theta_{n_R1}+\pi) = -f_k(\theta_{11},\ldots,\theta_{n_R1}),\\ &f_k(\theta_{11},\ldots,\theta_{i1}\pm 2\pi,\ldots,\theta_{n_R1}) = f_k(\theta_{11},\ldots,\theta_{i1},\ldots,\theta_{n_R1}),\end{aligned}$$

for all $1 \le i \le n_R$. Next lemma gives a result concerning functions with the above properties.

Lemma 2 *Let $f : \mathbb{R}^n \to \mathbb{C}$ be a measurable function such that $f(\theta_1,\ldots,\theta_n) = -f(\theta_1+\pi,\ldots,\theta_n+\pi)$ and that $f(\theta_1,\ldots,\theta_i\pm 2\pi,\ldots,\theta_n) = f(\theta_1,\ldots,\theta_i,\ldots,\theta_n)$, with $1 \le i \le n$. Then*

$$\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}f(\theta_1,\ldots,\theta_n)\mathrm{d}\theta_1\ldots\mathrm{d}\theta_n = 0.$$

Proof. See appendix B. □

Last lemma implies that $\mathbf{g} = \mathbf{0}$, which is a sufficient condition for (18) to be true. With this, we have proven that a diagonal $\mathbf{Q}$ is enough to attain the MIMO channel capacity.

B. Proof of lemma 2

Although lemma 2 is quite intuitive, its proof is given here for completeness. The proof will be done by induction.

- $n = 1$

Let $f : \mathbb{R} \to \mathbb{C}$ be a measurable function such that $f(\theta) = -f(\theta+\pi)$ and that $f(\theta) = f(\theta \pm 2\pi)$. Then,

$$\int_{-\pi}^{\pi}f(\theta)\mathrm{d}\theta = \int_{-\pi}^{0}f(\theta)\mathrm{d}\theta + \int_{0}^{\pi}f(\theta)\mathrm{d}\theta. \qquad (19)$$

Defining the change $\omega = \theta - \pi$ in the second term of the right hand side of last equation and using the property $f(\theta) = -f(\theta+\pi)$, we can write

$$\int_{-\pi}^{\pi}f(\theta)\mathrm{d}\theta = \int_{-\pi}^{0}f(\theta)\mathrm{d}\theta - \int_{-\pi}^{0}f(\omega)\mathrm{d}\omega = 0. \qquad (20)$$

- n true $\to n+1$ true

Let $f : \mathbb{R}^n \to \mathbb{C}$ be a measurable function such that $f(\theta_1,\ldots,\theta_n) = -f(\theta_1+\pi,\ldots,\theta_n+\pi)$ and that $f(\theta_1,\ldots,\theta_i\pm 2\pi,\ldots,\theta_n) = f(\theta_1,\ldots,\theta_i,\ldots,\theta_n)$, with $1 \le i \le n$, then

$$\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}f(\theta_1,\ldots,\theta_n)\mathrm{d}\theta_1\ldots\mathrm{d}\theta_n = 0. \qquad (21)$$

Let us now consider a measurable function $f : \mathbb{R}^{n+1} \to \mathbb{C}$ such that $f(\theta_1,\ldots,\theta_{n+1}) = -f(\theta_1+\pi,\ldots,\theta_{n+1}+\pi)$ and that $f(\theta_1,\ldots,\theta_i\pm 2\pi,\ldots,\theta_{n+1}) = f(\theta_1,\ldots,\theta_i,\ldots,\theta_{n+1})$, with $1 \le i \le n+1$. Then

$$\begin{aligned}&\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}f(\theta_1,\ldots,\theta_{n+1})\mathrm{d}\theta_1\ldots\mathrm{d}\theta_{n+1} = \qquad (22)\\ &= \int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\varphi(\theta_1,\ldots,\theta_n)\mathrm{d}\theta_1\ldots\mathrm{d}\theta_n, \qquad (23)\end{aligned}$$

[1] It is obviously bounded due to the constraint, and it is closed because of the way it is defined (preimage of a closed set by a continuous application).

where $\varphi(\theta_1,\ldots,\theta_n) = \int_{-\pi}^{\pi} f(\theta_1,\ldots,\theta_{n+1})\mathrm{d}\theta_{n+1}$. Evidently, $\varphi(\theta_1,\ldots,\theta_i \pm 2\pi,\ldots,\theta_n) = \varphi(\theta_1,\ldots,\theta_i,\ldots,\theta_n)$ holds for all i such that $1 \leq i \leq n$, and from the Fubini-Torelli theorem we get that $\varphi(\theta_1,\ldots,\theta_n)$ is a measurable function. Thus, from n case assumption, it is now sufficient to prove that $\varphi(\theta_1,\ldots,\theta_n) = -\varphi(\theta_1+\pi,\ldots,\theta_n+\pi)$. By definition

$$\varphi(\theta_1+\pi,\ldots,\theta_n+\pi) = \tag{24}$$

$$= \int_{-\pi}^{\pi} f(\theta_1+\pi,\ldots,\theta_n+\pi,\theta_{n+1})\mathrm{d}\theta_{n+1}. \tag{25}$$

Considering the change $\omega = \theta_{n+1} - \pi$ last equation reads

$$\varphi(\theta_1+\pi,\ldots,\theta_n+\pi) = \tag{26}$$

$$= \int_{-2\pi}^{-\pi} f(\theta_1+\pi,\ldots,\theta_n+\pi,\omega+\pi)\mathrm{d}\omega + \tag{27}$$

$$+ \int_{-\pi}^{0} f(\theta_1+\pi,\ldots,\theta_n+\pi,\omega+\pi)\mathrm{d}\omega. \tag{28}$$

Using the change $\varpi = \omega + 2\pi$ in the first term of last equation and the property $f(\theta_1,\ldots,\theta_i \pm 2\pi,\ldots,\theta_{n+1}) = f(\theta_1,\ldots,\theta_i,\ldots,\theta_{n+1})$, with $1 \leq i \leq n$, we get

$$\varphi(\theta_1+\pi,\ldots,\theta_n+\pi) = \tag{29}$$

$$\int_{0}^{\pi} f(\theta_1+\pi,\ldots,\theta_n+\pi,\varpi+\pi)\mathrm{d}\varpi + \tag{30}$$

$$+ \int_{-\pi}^{0} f(\theta_1+\pi,\ldots,\theta_n+\pi,\omega+\pi)\mathrm{d}\omega = \tag{31}$$

$$= \int_{-\pi}^{\pi} f(\theta_1+\pi,\ldots,\theta_n+\pi,\omega+\pi)\mathrm{d}\omega. \tag{32}$$

Using that $f(\theta_1,\ldots,\theta_{n+1}) = -f(\theta_1+\pi,\ldots,\theta_{n+1}+\pi)$ we can finally write

$$\varphi(\theta_1+\pi,\ldots,\theta_n+\pi) = -\int_{-\pi}^{\pi} f(\theta_1,\ldots,\theta_n,\omega)\mathrm{d}\omega = \tag{33}$$

$$= -\varphi(\theta_1,\ldots,\theta_n). \tag{34}$$

C. Roots of polynomial (15) in (ξ, γ)

Let us express (15) as

$$(x^2\Delta^2 - \xi^2) = \frac{4x\xi(x-\xi)(x+\Upsilon)}{x^2-\gamma^2}. \tag{35}$$

When $\xi < x < \gamma$ the derivative of the left hand side of last equation is $2x\Delta^2 > 0$, while the derivative of the right hand side can be expressed as

$$4\xi\frac{(x-\xi)(x+\Upsilon)+x(x+\Upsilon)+x(x-\xi)}{x^2-\gamma^2} - $$

$$-\frac{8x^2\xi(x-\xi)(x+\Upsilon)}{(x^2-\gamma^2)^2} < 0.$$

Consequently, it can be seen that for the range of values of x under consideration, the left hand side of (35) is monotonically increasing, while the right hand side is monotonically decreasing. Since there is always a solution within $\xi < x < \gamma$ for which these two side coincide, the solution must be unique.

References

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[2] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. on Information Theory*, vol. 49, pp. 1707–1727, July 2003.

[3] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. on Information Theory*, vol. 45, pp. 2007–2019, Sep. 1999.

[4] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Information Theory*, vol. 47, pp. 2632–2639, Sep. 2001.

[5] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Comm.*, vol. 16, pp. 1423–1436, Oct. 1998.

[6] S. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Intl. Conf. Commun. (ICC)*, vol. 7, pp. 2266–2270, June 2001.

[7] E. Jorswieck and H. Boche, "Optimal transmission with imperfect channel state information at the transmit antenna array," *Wireless Personal Commun.*, vol. 27, pp. 33–56, 2003.

[8] M. Skoglund and G. Jöngren, "On the capacity of a multiple-antenna communication link with channel side information," *IEEE J. Select. Areas Comm.*, vol. 21, pp. 395–405, Apr. 2003.

[9] W. C. Jakes, Jr., *Microwave Mobile Communications*, New York: John Wiley & Sons, Inc., 1974.

[10] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, London: Academic Press, 2000.

[11] D. Luenberger, *Optimization by Vector Space Methods*, New York: John Wiley & Sons, Inc., 1998.