Approximating the constellation constrained capacity of the MIMO channel with discrete input

Yankov, Metodi Plamenov; Forchhammer, Søren; Larsen, Knud J.; Christensen, Lars P. B.

Published in:
Proceedings of IEEE International Conference on Communicaitons

Link to article, DOI:
10.1109/ICC.2015.7248952

Publication date:
2015

Document Version
Peer reviewed version

Approximating the Constellation Constrained Capacity of the MIMO Channel with Discrete Input

Metodi P. Yankov*, Søren Forchhammer*, Knud J. Larsen* and Lars P. B. Christensen†
* Technical University of Denmark, Department of Photonics Engineering, 2800 Kgs Lyngby, Denmark
† Fingerprint Cards, Lyskær 3CD, 2730 Herlev, Denmark
Email: meya@fotonik.dtu.dk

***Abstract*—In this paper the capacity of a Multiple Input Multiple Output (MIMO) channel is considered, subject to average power constraint, for multi-dimensional discrete input, in the case when no channel state information is available at the transmitter. We prove that when the constellation size grows, the QAM constrained capacity converges to Gaussian capacity, directly extending the AWGN result from [1]. Simulations show that for a given constellation size, a rate close to the Gaussian capacity can be achieved up to a certain SNR point, which can be found efficiently by optimizing the constellation for the equivalent orthogonal channel, obtained by the singular value decomposition. Furthermore, lower bounds on the constrained capacity are derived for the cases of square and tall MIMO matrix, by optimizing the constellation for the equivalent channel, obtained by QR decomposition.**

I. Introduction

In next generation communication systems, high spectral efficiency will be needed in order to satisfy the exponential increase in data rate. The Multiple Input Multiple Output (MIMO) principle with large number of transmit and receive antennas (massive MIMO) will be a key technology to achieving this high spectral efficiency [2]. The ergodic capacity of the MIMO channel was found in [3], and is achieved when the input is a continuous Gaussian with variance given by the power constraint. However, practical transceivers demand signaling with constellations from a finite alphabet, making the analytical calculation of the Constellation Constrained Capacity (CCC) difficult. In [4][5] Blahut and Arimoto derive an iterative algorithm for finding the capacity and the capacity achieving input distribution on an AWGN channel. This algorithm was later modified in [6] to cover MIMO fading channels. However, the complexity of the algorithm grows exponentially both with the constellation size and the number of transmit antennas, making it impractical to calculate the CCC beyond e.g. 2x2 64QAM transmission. In order to cope with this problem, the authors in [6] conjecture, that when the input is taken as a Cartesian product of 1D PAM constellations, the Probability Mass Function (PMF) of the discrete points factorizes into the PMFs of each dimension, thus reducing the complexity of the optimization to 1D. As part of [7] we proved this conjecture to be true. On the other hand, Mutual Information (MI) is usually calculated using Monte-Carlo estimation of the normalized likelihood functions. When the number of receive antennas (the dimensionality of the observations) grows, the complexity still increases dramatically. In [8] the authors derive an analytical approximation of the MI when the input is uniformly distributed. However, it is seen to be inaccurate at low and high SNR. In [9] a better approximation is derived for high SNR via expansion of the MI. Lower and upper bounds are derived in [9] [10] based on the relation between the MI and the Minimum Mean Squared Error. Those bounds are only valid for uniform PMF, and are also quite inaccurate for low-to-moderate SNR. High SNR asymptotic behavior of the MI is studied for arbitrary input distribution in [11], where only AWGN channel is considered.

To our knowledge, the CCC of MIMO channels in the moderate SNR region, which is where practical communication systems tend to operate, is yet to be characterized. Furthermore, in that region the largest shaping gain can be expected for constellations of practical size [12]. The main contributions of this paper are as follows:

- It is proven, that as the constellation size grows, the CCC of MIMO channels approaches the Gaussian capacity, directly extending the AWGN result from [1]. The convergence rate is also the same as in [1]. For a given constellation size, information rates close to Gaussian capacity can be achieved up to a certain SNR point, which can be efficiently found by optimizing the constellation for the equivalent orthogonal channel, obtained by the Singular Value Decomposition (SVD).
- Lower bounds on the CCC of MIMO channels are derived for any SNR, based on the QR decomposition of the channel, using the diagonal elements of the upper-triangular $\mathbf{R}$ matrix. The bounds hold for the cases of square and tall MIMO matrix.
- It is shown empirically that in the low-to-mid SNR region the CCC is the same as the capacity of the equivalent orthogonal channel, obtained by the SVD, whereas in the mid-to-high SNR the above-mentioned lower bound can be used to characterize the MI.

II. Channel model and CCC of orthogonal channels

We consider a standard MIMO channel model:

$$Y = \mathbf{H}X + W, \tag{1}$$

where X is M-dimensional complex random variable vector $X = [X_1, X_2, \dots X_M]^T$, which can be continuous, or discrete taking values from the complex-valued set $\mathcal{X}^M$, obtained as

the Cartesian product of the basic 1D set $\mathcal{X}$. This can be a QAM, APSK, PAM etc. complex-valued set. The matrix $\mathbf{H}$ represents the $[N\mathbf{x}M]$ complex-valued channel, assumed here to have full rank $R = \min(M, N)$. The N dimensional complex AWGN W is assumed here to have unit variance and Y is the N dimensional channel observation. The received SNR with these assumptions is defined as SNR$= P_{av}/M$. We assume the channel realization is known at the receiver, but not at the transmitter. When the input is continuous, the ergodic capacity is achieved by Gaussian input distribution and can be found as [13]:

$$C_G = \mathrm{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathbf{I}^N + \frac{P_{av}}{M}\mathbf{H}\mathbf{H}^H\right)\right], \tag{2}$$

where $\mathrm{E}_{\mathbf{H}}[\cdot]$ denotes the expectation operator w.r.t. $\mathbf{H}$, $\mathbf{I}^N$ is a $[N\mathbf{x}N]$ identity matrix, $(\cdot)^H$ means conjugate transpose and $P_{av} = \sum_{i=1}^{|\mathcal{X}^M|} p(\mathbf{x_i})\alpha|\mathbf{x_i}|^2$ is the average power constraint at the transmitter. Here α is some scaling coefficient, and $\mathbf{x_i}$ is the $i-th$ point in the constellation. In this paper the focus is on finding the capacity and capacity achieving PMF when a fixed input constellation is used and without channel state information at the trasmitter, i.e. uniform power allocation and no pre-coding are employed. The channel capacity when signaling with $\mathcal{X}^M$ and averaging among the possible channel realizations can be expressed as [6]:

$$\begin{aligned} C &= \max_{p(X),\alpha} \mathrm{E}_{\mathbf{H}}\left[\mathcal{I}(X;Y|\mathbf{H})\right] = \\ &= \max_{p(X),\alpha} \sum_{i=1}^{|\mathcal{X}^M|} p(\mathbf{x_i})\left(\log_2 \frac{1}{p(\mathbf{x_i})} + T_i\right), \end{aligned} \tag{3}$$

$$\text{s.t.} \sum_{i=1}^{|\mathcal{X}^M|} p(\mathbf{x_i})\alpha|\mathbf{x_i}|^2 = P_{av} \text{ and } \sum_{i=1}^{|\mathcal{X}^M|} p(\mathbf{x_i}) = 1$$

where $\mathcal{I}(\cdot;\cdot)$ is the MI and:

$$T_i = \mathrm{E}_{\mathbf{H}}\left[\int_{\mathbf{y}} p(\mathbf{y}|\mathbf{x_i},\mathbf{H}) \log_2 p(\mathbf{x_i}|\mathbf{y},\mathbf{H}) d\mathbf{y}\right]. \tag{4}$$

As mentioned before, when the order of modulation and number of dimensions of the signal grow, maximizing (3) is impractical even when performed offline due to the exponential increase in the number of parameters to be optimized, and because it involves numerically calculating the expectation and integration in (4). We found that sufficiently accurate calculation of the MI on e.g. 64QAM 2x2 requires more than 10^5 samples of the observation space, resulting in a likelihood matrix of size [10^5x$2^{2\cdot 6}$]. Assuming larger constellations and larger antenna arrays, e.g. 256QAM and 8x8 set-up, which is already of interest in practical scenarios for single user MIMO in the e.g. IEEE 802.11ac WLAN standard, the number of samples in the output, needed for accurate estimation of the MI grows exponentially, thus making the calculations challenging for a standard computer.

A. Capacity of orthogonal channels

In this section we consider the CCC of orthogonal channels (or set of parallel channels). This means that the channel matrix can be expressed as diagonal, for which $M = N = R$. For each channel realization, the likelihood is Gaussian and factorizes as:

$$p(Y|X,\mathbf{H}) = \prod_{k=1}^{M} p(Y_k|X_k,\mathbf{H_{kk}}), \tag{5}$$

where $\mathbf{H_{kk}}$ is the element on the $k-th$ row and $k-th$ column of $\mathbf{H}$, and X_k is a random variable, representing the $k-th$ dimension of X, taking values from $\mathcal{X}$. As we prove in [7], when the input constellation is constructed as a Cartesian product of M 1D constellations $\mathcal{X}$, the capacity achieving PMF factorizes into its marginal PMFs on each dimension. The conditional distribution $p(X|Y,\mathbf{H})$ then also factorizes as $p(X|Y,\mathbf{H}) = \prod_{k=1:M} p(X_k|Y_k,\mathbf{H_{kk}})$ [7]. The capacity can then be expressed as:

$$\begin{aligned} \hat{C} &= \max_{p(X)} \mathrm{E}_{\mathbf{H}}\left[\mathcal{I}(X;Y|\mathbf{H})\right] = \\ &\max_{p(X)} \mathrm{E}_{\mathbf{H}}\left[\mathcal{H}(X|\mathbf{H}) - \mathcal{H}(X|Y,\mathbf{H})\right] = \\ &\max_{p(X)} \mathrm{E}_{\mathbf{H}}\left[\sum_{k=1}^{M}\left[\mathcal{H}(X_k|\mathbf{H_{kk}}) - \mathcal{H}(X_k|Y_k,\mathbf{H_{kk}})\right]\right] = \\ &\sum_{k=1}^{M} \max_{p(X_k)} \mathrm{E}_{\mathbf{H_{kk}}}\left[\mathcal{I}(X_k;Y_k|\mathbf{H_{kk}})\right], \end{aligned} \tag{6}$$

where $\mathcal{H}$ is the entropy function. When the channel matrix elements are identically distributed, (6) simplifies to:

$$\hat{C} = M \max_{p(X_i)} \mathrm{E}_{\mathbf{H_{ii}}}\left[\mathcal{I}(X_i;Y_i|\mathbf{H_{ii}})\right] \tag{7}$$

for any $i \in [1;M]$, subject to power constraint on the $i-th$ input $P_{av}^i = P_{av}/M$.

III. Capacity of interference channels

In this section the core results of the paper are derived. Let $\mathbf{U}, \mathbf{S}$ and $\mathbf{V}$ be the SVD components of $\mathbf{H} : \mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^H$. We assume that $\mathbf{S}$ is ordered, so that its first R diagonal elements are non-zero. Let us then consider 3 different channel models:

1) $Y = \mathbf{H}X + W$: one realization of the channel from (1)
2) $\hat{Y} = \mathbf{S}\hat{X} + W$: the channel, obtained by the SVD, where $\hat{Y} = \mathbf{U}^H Y$ and $\hat{X} = \mathbf{V}^H X$
3) $\tilde{Y} = \mathbf{S}X + W$: orthogonal channel, where $\mathbf{S}$ is the diagonal channel matrix.

We denote the MI on each channel as a function of the input distribution with $\mathcal{I}_1(\cdot), \mathcal{I}_2(\cdot)$ and $\mathcal{I}_3(\cdot)$, respectively. Let $\delta_{(\cdot)}$ denote the Dirac-delta function. We will also need the following PDFs:

1) $p_1(X) = \sum_{i=1:|\mathcal{X}^M|} w_i \delta_{\mathbf{x_i}}$
2) $q_1(X) = \sum_{i=1:|\mathcal{X}^M|} w_i \delta_{\mathbf{S}\mathbf{x_i}}$
3) $p_2(X) = \sum_{i=1:|\mathcal{X}^M|} w_i \delta_{\mathbf{V}^H\mathbf{x_i}}$
4) $q_2(X) = \sum_{i=1:|\mathcal{X}^M|} w_i \delta_{\mathbf{S}\mathbf{V}^H\mathbf{x_i}}$
5) $p_G(X) = \mathcal{N}(0, diag(P_{av}/M))$
6) $q_G(X) = \mathcal{N}(0, \mathbf{S}\mathbf{S}^H diag(P_{av}/M))$

In the first 4 PDFs, $w_i \geq 0$ for all i and $\sum_{i=1:|\mathcal{X}^M|} w_i = 1$. In the last 2 PDFs, $diag(P_{av}/M)$ is the covariance matrix

of the Gaussian, which is a diagonal matrix with elements P_{av}/M or ${\mathbf{S}_{\mathbf{ii}}}^2 P_{av}/M$, respectively. Let p_1^* denote the optimal PMF (or the PMF with optimal weights w_i) for Channel 3), i.e. $p_1^* = \arg\max \mathcal{I}_3(p_1(X))$. Likewise, p_2^* is the PMF with the same weights on the rotated version of the original QAM constellation. We will need the following auxiliary theorems:

Theorem 1: For any input PDF, $p_1(X)$, the mutual information on the non-orthogonal channel is the same as on the equivalent orthogonal channel with rotated input: $\mathcal{I}_1(p_1(X)) = \mathcal{I}_2(p_2(X))$

Proof: Given in the Appendix. ■

Theorem 2: When $P_{av}^1 = P_{av}^2 = \cdots = P_{av}^M = P_{av}/M$, the MI on all three channels with a continuous Gaussian input is the same: $\mathcal{I}_1(p_G(X)) = \mathcal{I}_2(p_G(X)) = \mathcal{I}_3(p_G(X))$

Proof: Given in the Appendix. ■

In [1] the authors prove that as the size of the constellation grows, Shannon capacity can be approached for AWGN channels. The proof relies on the fact that the MI is continuous in the *quadratic Wasserstein space*. The loss, incurring from the discrete nature of the input is then continuous in the quadratic Wasserstein distance W_2 between the discrete PMF and the continuous Gaussian distribution. The quadratic Wasserstein distance between two probability measures μ and v from the same space is defined as:

$$W_2(\mu, v) = \inf\left\{\left(\mathrm{E}_{X,Y}\left[||X-Y||^2\right]\right)^{1/2}\right\}, \tag{8}$$

where X and Y are governed by laws μ and v respectively, and the minimum is taken over all joint distributions of (X, Y) [1][14]. In Section II-A we showed that the capacity of an orthogonal channel is the sum of the capacities on each parallel channel, and therefore:

$$\lim_{|\mathcal{X}|\to\infty} \mathcal{I}_3(p_1^*(X)) = \mathcal{I}_3(p_G(X)). \tag{9}$$

Using the above results, we state the following theorem:

Theorem 3: Consider the interference channel model 1). As the size of the constellation grows, the CCC is $\lim_{|\mathcal{X}|\to\infty} \mathcal{I}_1(p_1^*(X)) = \mathcal{I}_1(p_G(X))$

Proof: The continuity of the MI in the W_2 space result from [1], together with the strict concavity of the MI in the input distribution [15] mean that when SNR> 0:

$$\lim_{|\mathcal{X}|\to\infty} \mathcal{I}_3(p_1^*(X)) = \mathcal{I}_3(p_G(X)) \Leftrightarrow \lim_{|\mathcal{X}|\to\infty} W_2(q_1(X), q_G(X)) = 0. \tag{10}$$

When $\mathbf{S}$ is full rank, if for some joint PDF $p(X, Y)$ we have that $\mathrm{E}_{X,Y}\left[||X-Y||^2\right] = 0 \Rightarrow \mathrm{E}_{\mathbf{S}X,\mathbf{S}Y}\left[||\mathbf{S}X - \mathbf{S}Y||^2\right] = \mathrm{E}_{\mathbf{S}X,\mathbf{S}Y}\left[||\mathbf{S}(X-Y)||^2\right] = 0$. Then the following is true:

$$\lim_{|\mathcal{X}|\to\infty} W_2(q_1(X), q_G(X)) = 0 \Leftrightarrow \lim_{|\mathcal{X}|\to\infty} W_2(p_1(X), p_G(X)) = 0. \tag{11}$$

The distribution $p_G(X)$ is rotationally invariant, and therefore $W_2(p_1^*(X), p_G(X)) = W_2(p_2^*(X), p_G(X))$. Then applying (10) we have:

$$\lim_{|\mathcal{X}|\to\infty} W_2(q_2(X), q_G(X)) = 0 \Rightarrow \lim_{|\mathcal{X}|\to\infty} \mathcal{I}_2(p_2^*(X)) = \mathcal{I}_3(p_G(X)). \tag{12}$$

By Theorems 1 and 2 we have $\mathcal{I}_1(p_1^*(X)) = \mathcal{I}_2(p_2^*(X))$ and $\mathcal{I}_3(p_G(X)) = \mathcal{I}_1(p_G(X))$ which proves the theorem. ■

Consequences of Theorem 3: The main consequence is that as the constellation size grows, Gaussian capacity can be approached for the interference channel for any SNR> 0. This is a direct extension of the result in [1] for AWGN channel. However, there is also a very practical application of Theorem 3: if for a fixed constellation size the CCC of the orthogonal channel approaches the Gaussian capacity, similar CCC can be expected on the equivalent interference channel *without precoding*. Using similar arguments as in [1], the rate at which the gap to Gaussian capacity vanishes can be estimated as $\mathcal{O}(1/|\mathcal{X}|)$.

Theorems 2 and 3 are proven for one realization of the channel. However, they can be extended to cover the ergodic case:

Theorem 4: $\lim_{|\mathcal{X}|\to\infty} \max_{p_1(X)} \mathrm{E}_{\mathbf{H}}\left[\mathcal{I}_1(p_1(X))\right] = C_G$

Proof: Let us re-define $p_1^*(X|\mathbf{H}_k) = \arg\max \mathcal{I}_1(p_1(X)|H = \mathbf{H}_k)$ as the optimal PMF for the $k-th$ channel realization. By Theorem 3 we have that:

$$\lim_{|\mathcal{X}|\to\infty} W_2(p_1^*(X|\mathbf{H}_k), p_G(X)) = 0 \tag{13}$$

for all k. The Wasserstein distance is a distance measure, and therefore [14]:

$$W_2(p_1^*(X|\mathbf{H}_k), p_1^*(X|\mathbf{H}_j)) \leq W_2(p_1^*(X|\mathbf{H}_k), p_G(X)) + W_2(p_1^*(X|\mathbf{H}_j), p_G(X)). \tag{14}$$

Taking the limit of large constellations, we get:

$$\lim_{|\mathcal{X}|\to\infty} W_2(p_1^*(X|\mathbf{H}_k), p_1^*(X|\mathbf{H}_j)) \leq \lim_{|\mathcal{X}|\to\infty} W_2(p_1^*(X|\mathbf{H}_k), p_G(X)) + \lim_{|\mathcal{X}|\to\infty} W_2(p_1^*(X|\mathbf{H}_j), p_G(X)) = 0. \tag{15}$$

The continuity of the MI means that due to the vanishing Wasserstein distance, in the limit of infinitely large constellations, if the optimal PMF on channel j achieves Gaussian capacity, it must also achieve similar capacity on channel k:

$$\lim_{|\mathcal{X}|\to\infty} \mathcal{I}_1(p_1^*(X|\mathbf{H}_k)|H = \mathbf{H}_k) = \mathcal{I}_1(p_1^*(X|\mathbf{H}_j)|H = \mathbf{H}_k) = \mathcal{I}_1(p_G(X)|H = \mathbf{H}_k) \tag{16}$$

for any j and k. Then for the average MI we have:

$$\lim_{|\mathcal{X}|\to\infty} \mathrm{E}_{\mathbf{H}}\left[\mathcal{I}_1(p_1^*(X, \mathbf{H}_k)) - \mathcal{I}_1(p_G)\right] = 0 \tag{17}$$

for any k, which proves the theorem. ■

A. Lower bounds via QR decomposition

Let $\mathbf{H} = \mathbf{QR}$ be the QR decomposition of $\mathbf{H}$, where $\mathbf{Q}$ is unitary and $\mathbf{R}$ is upper-triangular. A well known method for detection of MIMO signals uses the form of $\mathbf{R}$ to successively detect each layer by removing the interference from the previously detected layers - Successive Interference Cancellation (SIC). In this section we analyze the maximum rate which can be achieved by SIC under uniform power allocation and i.i.d. on the elements of the channel matrix, for the case of $M \leq N$, and therefore $R = M$.

We introduce two more channel models - Channels 4) and 5), with MI $\mathcal{I}_4(\cdot)$ and $\mathcal{I}_5(\cdot)$, respectively:

4) $\dot{Y} = \mathbf{R}X + W$, where $\dot{Y} = \mathbf{Q}^H Y$
5) $\ddot{Y} = diag(\mathbf{R})X + W$,

where $diag(\mathbf{R})$ means the matrix with the diagonal elements of $\mathbf{R}$ on its diagonal, and zeros elsewhere. Rotation does not change the multivariate Gaussian with i.i.d. on each dimension, and the noise distribution is therefore unchanged. Similarly to $p_1^*(X)$, we define the optimal discrete PMF input to Channel 5) as $p_5^*(X) = \arg\max \mathcal{I}_5(p(X))$.

Theorem 5: $\mathcal{I}_1(p_5^*(X)) \geq \mathcal{I}_5(p_5^*(X))$

Proof: We express the MI on Channel 4) with input $p_5^*(X)$ as:

$$\begin{aligned}\mathcal{I}_4(p_5^*(X)) &= \mathcal{H}(X) - \mathcal{H}(X|\dot{Y}) = \\ &= \mathcal{H}(X) - \mathcal{H}(X_M|\dot{Y}) - \sum_{i\in\{1:M-1\}} \mathcal{H}(X_i|\dot{Y}, X_{\{i+1:M\}}) \geq \\ &\geq \mathcal{H}(X) - \mathcal{H}(X_M|\dot{Y}) - \sum_{i\in\{1:M-1\}} \mathcal{H}(X_i|\dot{Y}), \end{aligned} \tag{18}$$

where the last inequality follows from the fact, that conditioning does not increase the entropy. Using this argument again, we can write:

$$\mathcal{H}(X_M|\dot{Y}) \leq \mathcal{H}(X_M|\dot{Y}_M) = \mathcal{H}(X_M|\ddot{Y}_M), \tag{19}$$

where we have also used the fact, that on the $M-th$ layer of Channel 4) there is no interference, and the conditional distributions $p(\dot{Y}_M|X_M)$ and $p(\ddot{Y}_M|X_M)$ for Channels 4) and 5) are the same. Consequently, for the same input $p_5^*(X)$, $p(X_M|\dot{Y}_M)$ and $p(X_M|\ddot{Y}_M)$ are also the same.

Due to the i.i.d. of the channel elements, and applying the chain rule for entropy multiple times, for any i we have:

$$\begin{aligned}\mathcal{H}(Y|X_M, H) &= \mathcal{H}(Y|X_i, H) = \\ \mathcal{H}(\mathbf{Q}^H Y|X_M, H) &= \mathcal{H}(\mathbf{Q}^H Y|X_i, H) \Rightarrow \\ \mathcal{H}(\dot{Y}|X_M, H) + \mathcal{H}(X_M) &= \mathcal{H}(\dot{Y}|X_i, H) + \mathcal{H}(X_i) \Rightarrow \end{aligned} \tag{20}$$

$$\begin{aligned}\mathcal{H}(\dot{Y}, X_M|H) &= \mathcal{H}(\dot{Y}, X_i|H) \Rightarrow \\ \mathcal{H}(X_M|\dot{Y}, H) + \mathcal{H}(\dot{Y}) &= \mathcal{H}(X_i|\dot{Y}, H) + \mathcal{H}(\dot{Y}) \Rightarrow \\ \mathcal{H}(X_M|\dot{Y}, H) &= \mathcal{H}(X_i|\dot{Y}, H) \leq \\ &\leq \mathcal{H}(X_M|\ddot{Y}, H) = \mathcal{H}(X_i|\ddot{Y}_i, H). \end{aligned} \tag{21}$$

Equation 20 follows from the fact, that due to the i.i.d. of the channel elements the marginal distributions on each dimension of X are identical. In (21) we have used that by definition, Channel 5) is orthogonal. Inserting (21) in (18) we have:

$$\begin{aligned}\mathcal{I}_4(p_5^*(X)) &\geq \mathcal{H}(X) - \sum_{i\in\{1:M\}} \mathcal{H}(X_i|\dot{Y}) \geq \\ &= \mathcal{H}(X) - \sum_{i\in\{1:M\}} \mathcal{H}(X_i|\ddot{Y}_i) = \mathcal{I}_5(p_5^*(X)). \end{aligned} \tag{22}$$

Similarly to Theorem 1, we have that $\mathcal{I}_4(p_5^*(X)) = \mathcal{I}_1(p_5^*(X))$, which proves the theorem. ■

We have arrived at a lower bound for the channel capacity. The MI $\mathcal{I}_5(p(X))$ can be easily optimized and calculated, in a manner, similar to $\mathcal{I}_3(p(X))$, since the channel is orthogonal. We only need the elements on the diagonal of the $\mathbf{R}$ matrix. Even though Theorem 5 was proven for $p_5(X)^*$, it actually follows for any $p(X)$, for which the dimensions of X are i.i.d., e.g. the uniform PMF.

In the case of continuous Gaussian input with uniform power allocation, the proof of Theorem 5 can be simplified. We can notice that the outputs of Channels 4) and 5) in that case are Gaussians, with respective covariance matrices:

$$\dot{\Sigma} = \mathrm{Cov}\left[\dot{Y}\right] = \frac{P_{av}}{M}\mathbf{R}\mathbf{R}^H + \Sigma_W \tag{23}$$

$$\ddot{\Sigma} = \mathrm{Cov}\left[\ddot{Y}\right] = \frac{P_{av}}{M} diag(\mathbf{R}) diag(\mathbf{R})^H + \Sigma_W, \tag{24}$$

where Σ_W is the diagonal covariance matrix of the noise. It is then trivial to show that:

$$\det\dot{\Sigma} \geq \det\ddot{\Sigma} \Rightarrow \mathcal{I}_4(p_G(X)) \geq \mathcal{I}_5(p_G(X)), \tag{25}$$

with equality if SNR $= \infty$.

B. Discussion of the theorems in Section III

The main implication of the theorems in this section is that while the ergodic Gaussian capacity of the orthogonal channels, obtained from the SVD of each channel realization can be approached with a finite size constellation, it can be expected that the ergodic Gaussian capacity of the interference channel is also approached with the same constellation, having the same PMF. As discussed in Section II-A, the CCC and the capacity achieving PMF of the orthogonal channel are easily calculated by the Blahut-Arimoto algorithm, taking $\mathbf{H_{ii}}$ in (7) to have the distribution of the singular values of the MIMO matrix. For large MIMO it is shown in [16] that the singular values distribution of Gaussian distributed channel matrix coefficients follows a quarter-quadratic law, which can be used to generate singular values for the 1D optimization. For small matrices the SVD is simple to calculate, and the distribution can be accurately approximated by Monte Carlo methods. When calculating the QR decomposition based lower bounds, the distribution of the elements on the diagonal of the $\mathbf{R}$ matrix is needed. Even though this distribution is not known, similar approach can be taken - draw matrices $\mathbf{H}$ from their known distribution, perform the QR decomposition on each $\mathbf{H}$, and use the diagonal elements of $\mathbf{R}$ instead of $\mathbf{H_{ii}}$ when maximizing (7).

As we mentioned in the introduction, complex-valued input sets are the focus here. When $\mathcal{X}$ is the popular QAM set, which is a product of two real-valued PAM sets, the reduction in complexity is relevant further down to the PAM set. Theorems 4 and 5 can then be used without loss of generality. Equation (7) becomes $\hat{C} = 2M \max_{p(X_i)} \mathrm{E}_{\hat{\mathbf{H}}_{\mathbf{ii}}} \left[\mathcal{I}(X_i; Y_i|\hat{\mathbf{H}}_{\mathbf{ii}})\right]$, where $\hat{\mathbf{H}}$ is the real-valued equivalent of $\mathbf{H}$, and each dimension of the input is taken from the corresponding PAM set. This is the model we consider in the following sections.

C. Some near-optimal input PMFs

Since we will exclusively use orthogonal channels to approximate the capacity in (3), it is worth examining the implications the distribution of the singular values has on the optimal PMF of X. Figure 1 depicts the optimal 8PAM component PMFs, i.e. $p(X) = \arg\max \mathrm{E}_{\mathbf{S_{11}}} \left[\mathcal{I}(X_1; \hat{Y}_1|\mathbf{S_{11}})\right]$, for transmission of different rank $R = M = N$ at the same average SNR, for which $\mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_3(p_1^*(X))\right] \approx C_G$. It is interesting to see how the shape of the optimal PMF changes when we increase the rank. This can be contributed to the fact, that the distribution of the singular values broadens. The AWGN channel can be considered as MIMO with zero variance of the singular values. The optimal input PMF on the AWGN channel therefore approaches the continuous Gaussian for this SNR. On the scalar fading channel, the singular values are Rayleigh distributed. As the rank of transmission grows, the distribution of singular values approaches the broader quarter-circular law. The optimal PMF in that case must account for broader range of SNR. It is well known, and can also be seen in [6], that uniform PMF approaches capacity at low and high SNR. The optimal PMF is therefore pushed to uniform when the rank of transmission increases. In Fig. 2, the histogram of the elements on the diagonal of the $\mathbf{R}$ matrix is shown. We see that it is narrower than the distribution of the singular values. We will use this fact in the next section to analyze the rates, achieved on the orthogonal Channels 3) and 5).

IV. Numerical calculation of capacity

In this section we provide Monte Carlo based calculation of the capacity for various channels. In Fig. 3, the ergodic capacity for the 2x2 i.i.d. MIMO Rayleigh fading channel, i.e. $\mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_3(p_G(X))\right]$ is shown, together with the 64QAM CCC, i.e. $\max \mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_1(p(X))\right]$, the capacity of the SVD based orthogonal channel, i.e. $\max \mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_3(p(X))\right]$, and the QR decomposition based lower bound - $\max \mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_5(p(X))\right]$. We directly see the region, where the limits in Theorem 4 are approached: up to around SNR$= 10dB$. As the SNR increases, the gap to capacity also increases due to the limited size of the constellation. As shown in [10], when the input to an orthogonal channel is discrete, orthogonal inputs can be suboptimal. In Fig. 3 this effect can be seen, as $\max \mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_1(p_1(X))\right] = \max \mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_2(p_2(X))\right] > \max \mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_3(p_1(X))\right]$, or a rotated version of the original QAM performs better on the orthogonal channel. The QR based lower bound in the low SNR is seen as a worse approximation than the SVD based capacity. In this

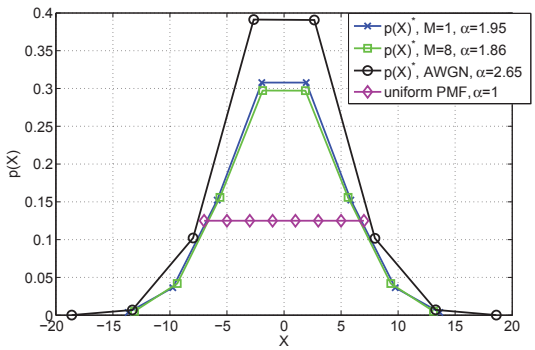

Fig. 1. 8PAM PMFs for different channels at average SNR $= 8dB$. The scaling coeficient α, resulting in the optimal PMF is indicated in the legend. As α increases, more and more weight needs to be put on points with lower energy, in order to keep the average power the same.

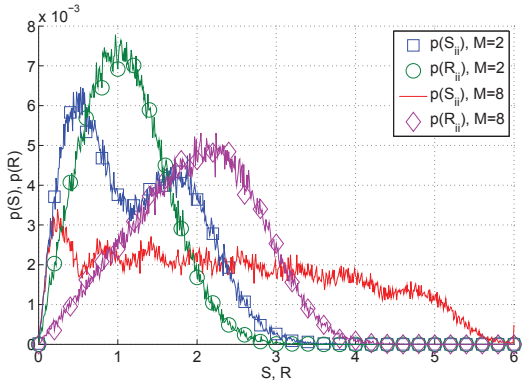

Fig. 2. Histogram of the singular values and the elements on the diagonal of the $\mathbf{R}$ matrix for MIMO of different rank

regime the noise is the limiting factor, and the inequality in (25) becomes more and more strict. In the moderate to high SNR we see that the lower bound becomes tighter, and exceeds the SVD based channel capacity. This is due to the distribution of the diagonal elements of the $\mathbf{R}$ matrix (see Fig. 2). Tighter distribution means that the optimal PMF does not need to account for high and low instantaneous SNR, where uniform PMF is optimal, i.e. the channel is more stable. The SVD based channel on the other hand has optimal PMF, which must be robust to deep fades and vanishing fades. It is therefore pushed to uniform PMF, resulting in lower average MI.

In Fig. 4, the SVD based channel capacity for a 8x8 setup is given, together with the QR decomposition based lower bounds. For SNR $< 10, 16,$ and $24dB$ for 64, 256 and 1024 QAM, respectively, the SVD based capacity is approaching Gaussian capacity, and we can therefore expect that in those SNR regions, $\max \mathrm{E}_{\mathbf{H}} \left[\mathcal{I}_1(p_1(X))\right]$ is also close to the Gaussian capacity. As before, the QR based lower bounds are worse at low SNR, but become better at moderate to high SNR. The envelope of the two curves - the SVD and QR based capacities - can therefore serve as an approximation to the CCC of the MIMO channel for the entire SNR region.

V. Future work

As mentioned before, in this paper lower bounds are derived only for the case of $M \leq N$. When $M > R$, the last layer

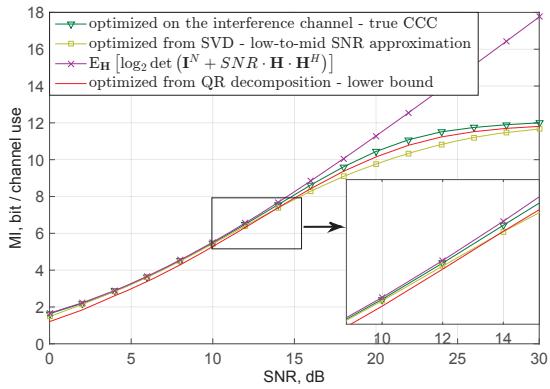

Fig. 3. Capacity curves for i.i.d. Rayleigh fading 2x2 MIMO channel, 64QAM

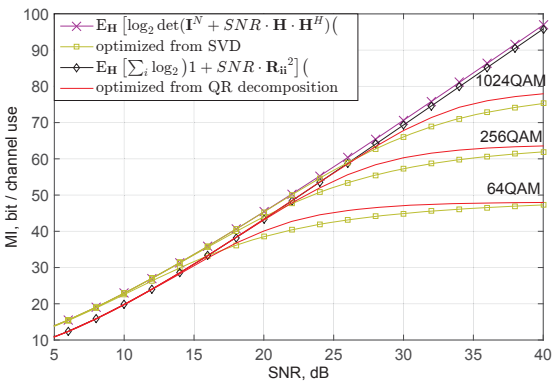

Fig. 4. Capacity curves for i.i.d. Rayleigh fading 8x8 MIMO channel

of Channel 4) will no longer be interference free, which was a necessary condition for stating (19). An interesting area for future research is to provide non-trivial lower bounds for the case of $M > R$.

VI. Conclusion

In this paper the capacity of a MIMO channel with discrete input was considered in the case when no channel knowledge at the transmitter is available. It was proven that as the constellation size grows, the capacity of the interference channel converges to the capacity of the equivalent orthogonal channel, obtained by the SVD, and consequently approaches Gaussian capacity. Simulations showed that values close to that rate can be achieved up to a certain SNR point for QAM constellations of a given size. Using the Blahut-Arimoto algorithm, the capacity of the orthogonal channel can be easily calculated, and the SNR threshold can be obtained, together with the capacity achieving PMFs for each SNR point. These PMFs can then be used on the ergodic interference channel, and similar capacity can be expected up to the SNR threshold. Lower bounds on the constellation constrained capacity were also derived for MIMO channels with square or tall matrix, using the capacity of the QR decomposition based channel. The envelope of the SVD based approximation and the QR based lower bounds can then be used to estimate the constellation constrained capacity for the entire SNR region.

VII. Appendix

A. Proof of Theorem I

Since $\mathbf{U}$ is a rotation matrix, we have:

$$\mathcal{H}(\hat{Y}|\cdot) = \mathcal{H}(\mathbf{U}\hat{Y}|\cdot) - \log_2 |\det \mathbf{U}| = \mathcal{H}(\mathbf{U}\hat{Y}|\cdot) = \mathcal{H}(Y|\cdot).$$

Similarly $\mathcal{H}(X|\cdot) = \mathcal{H}(\hat{X}|\cdot)$. Then it is clear that:

$$\mathcal{I}(X;Y|\mathbf{H}) = \mathcal{I}(\hat{X};Y|\mathbf{H}) = \mathcal{I}(Y;\hat{X}|\mathbf{H}) = \mathcal{I}(\hat{Y};\hat{X}|\mathbf{H}).$$

B. Proof of Theorem II

The distribution $p_G(X)$ is rotationally invariant, i.e. $X \equiv \mathbf{V}^H X \Rightarrow \mathcal{I}_2(p_G(X)) = \mathcal{I}_3(p_G(X))$ and by Theorem 1 $\mathcal{I}_2(p_G(X)) = \mathcal{I}_1(p_G(X))$.

Acknowledgment

The authors would like to thank Prof. Dr. sc. techn. Gerhard Kramer at the Technical University of Munich for the fruitful discussion towards deriving the QR decomposition based lower bounds on the constellation constrained MIMO capacity.

References

[1] Y. Wu and S. Verdú, "Functional Properties of Minimum Mean-Square Error and Mutual Information", *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289-1301, 2012

[2] T. Marzetta, O. Edfors, F. Tufvesson, E. G. Larsson, "Massive MIMO for Next Generation Wireless Systems", *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186-195, 2014

[3] A. Goldsmith and P. Varaiya, "Capacity of Fading Channels With Channel Side Information," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1986-1992, 1997.

[4] R. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460-473, 1972.

[5] S. Arimoto, "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14-20, 1972.

[6] J. Bellorado, S. Ghassemzadeh and A. Kavcic, "Approaching the Capacity of the MIMO Rayleigh Flat-Fading Channel with QAM Constellations, Independent across Antennas and Dimensions", *IEEE Trans. Commun.*, vol. 5, no. 6, pp. 1322-1332, 2006

[7] M. Yankov, S. Forchhammer, K. J. Larsen and L. P. B. Christensen, "Factorization Properties of the Optimal Signaling Distribution of Multidimensional QAM Constellations", *IEEE 6-th International Symposium on Communications, Control and Signal Processing* pp. 384 - 387, 2014

[8] J. Akhtman and L. Hanzo, "Closed-form Approximation of MIMO Capacity," *IET Electronics Letters*, vol. 45, no. 1, pp. 68-69, 2009.

[9] M. Rodrigues, "Multiple-Antenna Fading Channels With Arbitrary Inputs: Characterization and Optimization of the Information Rate", *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 569-585, 2014

[10] F. Prez-Cruz, M. R. D. Rodrigues and S. Verdú, "MIMO Gaussian Channels With Arbitrary Inputs: Optimal Precoding and Power Allocation", *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070-1084, 2010

[11] A. Alvarado, F. Brännström, E. Agrell and T. Koch, "High-SNR Asymptotics of Mutual Information for Discrete Constellations With Applications to BICM", *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1061-1076, 2014

[12] M. Yankov, S. Forchhammer, K. J. Larsen and L. P. B. Christensen, "Rate-adaptive Constellation Shaping for Turbo-coded BICM", *IEEE International Conference on Communications*, pp. 2112 - 2117, 2014

[13] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj and H. V. Poor, *MIMO Wireless Communications*, Cambridge University Press 2007

[14] C. Villani, *Optimal Transport: Old and New*, Berlin, Germany, Springer-Verlag, 2008.

[15] I. C. Abou-Faycal, M. D. Trott and S. Shamai, "The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290-1301, 2001.

[16] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Now Publishers, 2004