MAC Gaussian Channels with Arbitrary Inputs: Optimal Precoding and Power Allocation

Samah A. M. Ghanem
Instituto de Telecomunicações
Faculdade de Engenharia, Universidade do Porto,
Porto, Portugal
samah.ghanem@fe.up.pt

Abstract—In this paper, (i) we derive closed-form expressions for the gradient of the mutual information with respect to arbitrary parameters of the Two-user multiple-access channel (MAC), (ii) we investigate the linear precoding and power allocation policies that maximize the mutual information for the two-user MAC Gaussian channels with arbitrary input distributions, capitalizing on the relationship between mutual information and minimum mean-square error (MMSE). (i) enforces the fundamental relations between information theory and estimation theory. And (ii) enforces that the optimal design of linear precoders may satisfy a fixed-point equation as a function of the channel and the input constellation under specific setups. We provide an interpretation for the interference with respect to the channel, power, and input estimates of the main user and the interferer.

Keywords—minimum mean-square error (MMSE); multiple-access–channel (MAC); mutual information; precoding.

I. INTRODUCTION

Linear precoding for mutually interfering channels corrupted with independent Gaussian noise can be maximized under a total power constraint [1]. In the multiple user setting, the mutual information is maximized by imposing a covariance structure on the Gaussian input vector that satisfies two conditions: First, transmitting along the right eigenvectors of the channel. Second, the power allocation follows the waterfilling based on the signal-to-noise (SNR) ratio in each non-interacting subchannel [2], [3]. The MAC channel stands as a special case of interference channels. With arbitrary inputs, the mutual information of the MAC channel is minimized due to the fact that at a certain point the inputs may lie in the null space of the channel matrix, and this causes the mutual information to decay at such point. Therefore, the capacity-achieving strategies can be employed either at the transmitter side and/or the receiver side: by orthogonalizing the inputs using time division multiplexing/frequency division multiplexing (TDM/FDM) techniques, or via using interference mitigation techniques, such as successive interference cancellation, dealing with interference as noise, decoding interference, rate splitting, as well as receive diversity techniques. Precoding is another technique that can be used at the transmitter side to maximize the mutual information by exploiting the transmit diversity. On the one hand, precoding plays the role of power allocation diagonalizing the channel while on the other hand rotating the channel eigen vectors; thus shaping the constellation for better detection. And since practical considerations enforce the use of discrete constellations, such as phase shift keying (PSK) and quadrature amplitude modulation (QAM), which depart significantly from the optimal Gaussian distributions, it is of great importance to revisit the linear precoding problem under the constraint that each input follows a specified discrete constellation. In [3], [4], and [5], the optimal power allocation for parallel non-interacting channels with arbitrary inputs, and the optimal precoding for MIMO channels with different setups were obtained exploiting the relation between the mutual information and the MMSE [4], [6]. The results in [1] demonstrate that capacity-achieving strategies for Gaussian inputs may be suboptimal for discrete inputs. In [3] a general solution for the optimum precoder has been derived for the high and low-snr regimes capitalizing on first order expansions for the MMSE and relating it to mutual information. The mercury/waterfilling solution for power allocation maximizes the mutual information compensating for the non-Gaussianness of the discrete constellations [7]. However, linear precoding techniques introduce rescaling among the channel inputs that achieves cleaner detection of the received signals and thus higher information rates. In this paper, we propose a linear precoder structure that aims to maximize the input–output mutual information for the two-user MAC setting with channel state information known at both receiver and transmitter sides.

The rest of this paper is organized as follows, we derive the relation between the gradient of the mutual information and the MMSE for the MAC channel in Section II-A. We derive the optimal precoding policy in Section II-B. We provide a general form of the optimal precoder, and then we study the low SNR regime for the simplified setups in Sections III and IV, respectively. In section V we derive the optimal power allocation. Finally, we provide a set of numerical results in section VI.

1The following notation is employed, the superscript (.)T, and (.)H denote transpose and conjugate transpose operations. (.)T denotes optimum, Tr(.) denotes the trace of a matrix, E(.) denotes the expectation, (X)x and xj denotes the jth element of the jth column of the matrix X, xi denotes the ith column of X. The Frobenius norm of a matrix is denoted by ||X|| = \sqrt{Tr(X'X)} which reduces to the L2 norm ||x|| in the special case of a vector. And pT denotes the gradient with respect to X.

©IEEE 2012
II. THE TWO-USER MAC CHANNEL

A. Problem Definition

Consider the deterministic complex-valued vector channel:

\[ y = \sqrt{snr} H_1 P_1 x_1 + \sqrt{snr} H_2 P_2 x_2 + n \]  

(1)

Where the \( n \)-dimensional vector \( y \) and the \( m \)-dimensional vectors \( x_1, x_2 \) represent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vectors from each user input to the MAC channel. The distributions of the both inputs are not fixed, not necessarily Gaussian nor identical. The \( n\times m \) complex-valued matrices \( H_1, H_2 \) correspond to the deterministic channel gains for both input channels (known to both encoder and decoder) and \( n \) is the \( m\times m \) -dimensional complex Gaussian noise with independent zero-mean unit-variance components. The optimization of the mutual information is carried out over all preceding matrices \( P_1, P_2 \) that do not increase the transmitted power.

The precoding problem can be cast as a constrained nonlinear optimization problem as follows,

\[
\text{max} \ I(x_1, x_2; y) \quad (2)
\]

Subject to

\[
\begin{align*}
&Tr\{E[P_1 x_1 x_1^\dagger P_1]\} = Tr\{E[P_1]\} \leq Q_1 \quad (3) \\
&Tr\{E[P_2 x_2 x_2^\dagger P_2]\} = Tr\{E[P_2]\} \leq Q_2 \quad (4)
\end{align*}
\]

To solve the optimization problem in (2), we need to know the relation between the mutual information and the non-linear MMSE in the context of MAC channels.

**Theorem 1:** The relation between the gradient of the mutual information and the non-linear MMSE for a two-user MAC channel (1) satisfies,

\[
\begin{align*}
&D_1 I(x_1, x_2; y) = H_1 P_1 E P_1^\dagger - H_2 P_2 E [\hat{x}_1 \hat{x}_1^\dagger P_1] \\
&D_2 I(x_1, x_2; y) = H_2 P_2 E P_2^\dagger - H_1 P_1 E [\hat{x}_2 \hat{x}_2^\dagger P_2] \\
\end{align*}
\]

(5)\hfill(6)

With

\[
\begin{align*}
&E_1 = E(x_1 - \hat{x}_1) (x_1 - \hat{x}_1)^\dagger \\
&E_2 = E(x_2 - \hat{x}_2) (x_2 - \hat{x}_2)^\dagger
\end{align*}
\]

(7)\hfill(8)

Such that,

\[
\begin{align*}
\hat{x}_1 &= E(x_1 | y) = \sum_{x_1, x_2} x_1 p(y|x_1, x_2)p(x_1)p(x_2) \\
\hat{x}_2 &= E(x_2 | y) = \sum_{x_1, x_2} x_2 p(y|x_1, x_2)p(x_1)p(x_2)
\end{align*}
\]

(9)\hfill(10)

And,

\[
\begin{align*}
&\frac{p(y|x_1, x_2)}{p(y)} \leq 1/n e^{-\frac{1}{2}(y - H_1 P_1 x_1 - H_2 P_2 x_2)^2} \\
\end{align*}
\]

(11)

Henceforth, we denote the MMSE by,

\[
\text{MMSE}(\text{snr}) = \text{E}[||H_1 P_1 (x_1 - \hat{x}_1)||^2] + \text{E}[||H_2 P_2 (x_2 - \hat{x}_2)||^2] \\
\text{Tr} [H_1 P_1 E_1 (H_1 P_1^\dagger)] + \text{Tr} [H_2 P_2 E_2 (H_2 P_2^\dagger)]
\]

(12)

**Proof:** See Appendix A

Note that we can derive this relation with respect to any arbitrary parameter following similar steps of the proof. Note also that the derived relations in (5) and (6) reduces to the relation between the gradient of the mutual information and the non-linear MMSE derived for the linear vector Gaussian channels [4] if the multiplication of the estimates of the inputs is zero, in other words, if and only if the users inputs were orthogonalized. However, the derived relation between the gradient of the mutual information and the MMSE of the two-user MAC setting in Theorem 1 will lead to a new formulation of \( E(x|y) \), that is discussed in the following Theorem.

**Theorem 2:** The estimates of the inputs \( x_1, x_2 \) of the two-user MAC channel given the output \( y \) can be expressed as,

\[
H_1 P_1 E(x_1|y) + H_2 P_2 E(x_2|y) = y + \frac{p(y|\hat{y})}{p(y)}
\]

(15)

With \( p(y) \) being the probability density function for the received output vector \( y \).

**Proof:** See Appendix B

Equation (15) can be solved numerically to obtain the estimates of the inputs.

**Lemma 1:** If the channel is a linear vector Gaussian channel, Theorem 2 tends to the following formulation,

\[
E(x|y) = \left[ y + \frac{p(y|\hat{y})}{p(y)} \right] (HP)^{-1}
\]

(16)

The non-linear estimates given in (15) give a statistical intuition to the problem, however, in practical setups, such estimates can be found via the linear MMSE, solving the following optimization problem,

\[
\text{mi n} \quad \text{MMSE}(\text{snr})
\]

(17)

**Theorem 3:** The estimates of the inputs \( x_1, x_2 \) of the two-user MAC channel given the output \( y \) can be expressed as,

\[
\begin{align*}
\hat{x}_1 &= P_1^* H_1^\dagger (1 + P_1^* H_1^\dagger H_1 P_1 + P_2^* H_2^\dagger H_2 P_2)^{-1} y \\
\hat{x}_2 &= P_2^* H_2^\dagger (1 + P_1^* H_1^\dagger H_1 P_1 + P_2^* H_2^\dagger H_2 P_2)^{-1} y
\end{align*}
\]

(18)\hfill(19)

**Proof:** The estimates can be found by deriving the optimal Wiener receive filters solving (17), details omitted due to lack of space.

**B. Optimal Precoders**

Solving the optimization problem in (2) subject to the constraints (3) and (4), we get the following theorem.

**Theorem 4:** Let the distribution of the arbitrary inputs \( x_1, x_2 \) to the channel in (1) be \( p(x_1), p(x_2) \) respectively. The optimum precoding matrix satisfies the following,

\[
\begin{align*}
P_1^* &= v_1^{-1} H_1^\dagger H_1 P_1 E_1 - v_1^{-1} H_1^\dagger H_2 P_2 E_2 \\
P_2^* &= v_2^{-1} H_2^\dagger H_2 P_2 E_2 - v_2^{-1} H_2^\dagger H_2 P_2 E_2
\end{align*}
\]

(20)\hfill(21)

With \( v_1 \) and \( v_2 \) are the Lagrangian multiplier of the optimization problem in (2).

**Proof:** The proof of Theorem 4 relies on the Karush–Kuhn–Tucker (KKT) conditions [8] and the relation between the gradient of the mutual information with respect to the precoding matrices and the MMSE.

There are unique \( P_1^*, P_2^* \) that satisfy the KKT conditions when the problem is strictly concave, corresponding to the global maximum. Using Monte Carlo method, we can compute the MMSE matrices (7) and (8) for discrete constellations. To obtain the solution of (20) and (21), we can use an iterative approach, similar to [1], [3], and [5]. It is very important to notice that the optimal precoders of (20) and (21)
satisfies a fixed point equation under two specific setups: **First**, for the multiple-channels-per user in the two-user-MAC if and only if the second term in the gradient of the mutual information with respect to the precoding matrix is zero, this occurs when both inputs are orthogonal, or in this case, we can think about it as if the second user multiple-inputs have no influence to the first user multiple inputs, therefore, each user precoding is over his mutually interfering channels only. If each user is using OFDM signaling, i.e. transmitting over parallel Gaussian channels, the problem breaks into two separate problems, and the precoding solution for each is a fixed point equation. **Second**, for the SISO-two-user-MAC case with each user is mutually interfering with the other, in this case the relation between the gradient of the mutual information with respect to the precoding matrix can be manipulated as in [4] and thus, the precoder structure is a special case of the one introduced in [3].

### III. Precoder Structure

We will define in the following setup a general form of the optimal precoder’s structure for the two-user MAC channel under the two specific setups where the precoder would satisfy a fixed point equation. The following theorem gives a generalization of the result in (20) and (21) of the precoders, thus, the solution applies to the low-snr and high-snr regime. However, it also reduces to the optimal power allocation solution when the precoding is precluded as it will be shown later. Therefore, the precoder should admit a structure that performs matching of the strongest source modes to the weakest noise modes, and this alignment enforces permutation process to appear in the diagonal power allocation setup.

**Theorem 5**: The non-unique first-order optimal precoders that maximizes the mutual information for a two-user MAC subject to per-user average power constraint can be written as follows,

\[ P_1^* = U_1 D_1 R_1 \]
\[ P_2^* = U_2 D_2 R_2 \]

With \( U_1, U_2 \) are the unitary matrices corresponding to the channel right singular vectors, \( D_1, D_2 \) are diagonal matrices.

**Proof**: This Theorem is a direct consequence to [Theorem1, 5], details omitted due to lack of space.

\( D_1 = \text{diag}\{\sigma_{P_{1,1}}, ..., \sigma_{P_{1,m}}\} \) and \( D_2 = \text{diag}\{\sigma_{P_{2,1}}, ..., \sigma_{P_{2,m}}\} \) are power allocation matrices; i.e., correspond to the mercury/waterfilling optimum power allocation.

And \( R_1, R_2 \) are rotation matrices that insure firstly, allocation of power into the strongest channel singular vectors, and secondly, diagonalizes the MMSE matrix insuring uncorrelated error or in other words independency between inputs. Both rotation matrices induces in its structure the eigen vectors of the MMSE matrix enforcing the setup defined in (20) and (21) for the two special setups discussed previously about each user inputs.

### IV. LOW-SNR REGIME

We now consider the optimal precoding policy for the two-user MAC Gaussian channel with arbitrary input distributions in the regime of low-snr. Consider the case of zero-mean uncorrelated complex inputs, with \( E(x_1 x_1^\dagger) = E(x_2 x_2^\dagger) = 0 \), and \( E(x_1 x_1^\dagger) = E(x_2 x_2^\dagger) = 0 \). We consider the low-snr expansion to the MMSE of equation (12), however, it can be easily deduced that the Taylor expansion of the non-linear MMSE in (12) will lead to the first order Taylor expansion of the linear MMSE for the Gaussian inputs setup, thus the low-snr expansion of the MMSE matrix can be expressed as,
\[
E = I - (H_1 P_1)^\dagger H_1 P_1 \text{snr} - (H_2 P_2)^\dagger H_2 P_2 \text{snr} + o(\text{snr}^2), \tag{26}
\]
with \( E = E_1 + E_2 \). Consequently, \[
\text{MMSE(\text{snr})} = \text{Tr}[(H_1 P_1 E_1 (H_1 P_1)^\dagger) + \text{Tr}[(H_2 P_2 E_2 (H_2 P_2)^\dagger)] \tag{27}
\]
\[
\text{MMSE(\text{snr})} = \text{Tr}[(H_1 P_1 (H_1 P_1)^\dagger + \text{Tr}[(H_2 P_2 (H_2 P_2)^\dagger) \text{snr} - \text{Tr}((H_2 P_2 (H_2 P_2)^\dagger)^2 \text{snr} + o(\text{snr}^2) \tag{28}
\]

Capitilizing on the relationship between mutual information and MMSE [4], [6], the low-snr Taylor expansion of the mutual information is given by,
\[
\mathcal{I}(\text{snr}) = \text{Tr}[(H_1 P_1 (H_1 P_1)^\dagger) \text{snr} + \text{Tr}[(H_2 P_2 (H_2 P_2)^\dagger) \text{snr} - \text{Tr}((H_1 P_1 (H_1 P_1)^\dagger)^2 \frac{\text{snr}^2}{2} - \text{Tr}((H_2 P_2 (H_2 P_2)^\dagger)^2 \frac{\text{snr}^2}{2} + o(\text{snr}^3) \tag{29}
\]

The wideband slope which indicates how fast the capacity is achieved in terms of bandwidth, is inversely proportional to the second order derivative term of the mutual information in the low-snr Taylor expansion (26). Therefore, this term is a key low-power performance measure since the bandwidth required to sustain a given rate with a given low power, i.e. minimal energy per bit, is inversely proportional to this term [9]. According to (29), for first-order optimality, subject to (3) and (4), the form of the optimal precoder follows from the low-snr expansions of the form of optimal precoder for the complex Gaussian inputs settings. To prove this claim, let’s redefine our optimization problem as follows,
\[
\max \text{Tr}[(H_1 P_1 (H_1 P_1)^\dagger) \text{snr} + \text{Tr}[(H_2 P_2 (H_2 P_2)^\dagger) \text{snr} \tag{30}
\]
Subject to \[
\text{Tr}[(E P_1^\dagger P_1)] \leq 1 \tag{31}
\]
\[
\text{Tr}[(E P_2^\dagger P_2)] \leq 1 \tag{32}
\]
Let’s do an eigen value decomposition such that,
\[
H_1^\dagger H_1 = U_1 \Omega_1 U_1^\dagger \tag{33}
\]
\[
H_2^\dagger H_2 = U_2 \Omega_2 U_2^\dagger \tag{34}
\]
And Let,
\[
P_1^* = U_1^\dagger P_1 \tag{35}
\]
\[
P_2^* = U_2^\dagger P_2 \tag{36}
\]
Let \( Z_1 \geq 0 \) and \( Z_1 \geq 0 \) both positive definite matrices, such that,
\[
Z_1 = P_1^\dagger P_1 \tag{37}
\]
\[
Z_2 = P_2^\dagger P_2 \tag{38}
\]
Plug (33) to (38) into (30), (31), and (32), the optimization problem can be re-written as,
\[
\max = \text{Tr}[Z_1 \Omega_1] \text{snr} + \text{Tr}[Z_2 \Omega_2] \text{snr} \tag{39}
\]

©IEEE 2012
Subject to
\[ T_r(Z_i) \leq 1 \]  
(40)
\[ T_r(Z_j) \leq 1 \]  
(41)
Solving the KKT conditions of the Lagrangian of the objective function (39) subject to (40) and (41) leads to,
\[ Z_i = \lambda^{-1}\Omega_i \]  
(42)
\[ Z_j = \lambda^{-1}\Omega_j \]  
(43)
The results in (42), (43) proves that the optimal precoders in the low-snr perform mainly two operations, firstly, it aligns the transmit directions with the eigen vectors of each sub-channel. Secondly, it performs power allocation over the sub-channels.

V. OPTIMUM POWER ALLOCATION

The power-allocation problem can be formulated as a linearly constrained nonlinear optimization,
\[ \max I(x_1, x_2; y) \]  
(44)
Subject to,
\[ (P_i)_{ij} = \sqrt{p_{1j} \cdot (P_2)_{ij} = \sqrt{p_{2j}} } \]  
(45)
\[ \sum_{j=1}^{m} p_{1j} \leq 1, p_{1j} \geq 0 \]  
(46)
\[ \sum_{j=1}^{m} p_{2j} \leq 1, p_{2j} \geq 0 \]  
(47)
Where \( p_{1} = [p_{11}, \ldots, p_{1m}] \), and \( p_{2} = [p_{21}, \ldots, p_{2m}] \).

**Theorem 6:** The optimal power allocation that solves (44) subject to (45)–(47) satisfies,
\[ p_{1j}^* = v_{1j}^{-1}p_{1j} \left| H_{1j}\right| H_{1j} E_{1j} - v_{1j}^{-1}p_{1j} \left| H_{1j}\right| H_{1j} P_{1j} E_{2j} \]  
(48)
\[ p_{2j}^* = v_{2j}^{-1}p_{2j} \left| H_{2j}\right| H_{2j} E_{2j} - v_{2j}^{-1}p_{2j} \left| H_{2j}\right| H_{2j} P_{2j} E_{1j} \]  
(49)

**Proof:** This theorem is a direct consequence of Theorem 4.

For the special case where we have a SISO per user in the two-user MAC, (48)–(49) can be simplified to the following,
\[ p_{1j}^* = v_{1j}^{-1}p_{1j} \left| h_{1j}\right| E_{1j} - v_{1j}^{-1}p_{1j} \left| h_{1j}\right| \operatorname{E}[x_2 \hat{x}_2] \]  
(50)
\[ p_{2j}^* = v_{2j}^{-1}p_{2j} \left| h_{2j}\right| E_{2j} - v_{2j}^{-1}p_{2j} \left| h_{2j}\right| \operatorname{E}[x_1 \hat{x}_1] \]  
(51)
In (50) and (51), due to the fact that the mean-square error is less or equal to 1, and consider the special case where the second covariance term is zero, i.e., no user is interfering with the other, we can express the optimal power allocation in the single user mercury/waterfilling form [7] as follows,
\[ p_{1j}^* = \left( \frac{1}{\text{snr}_{1j} \left| h_{1j}\right|} \right) \left( \frac{v_{1j}}{\left| h_{1j}\right|} \right)^2 \]  
(52)
\[ p_{2j}^* = \left( \frac{1}{\text{snr}_{1j} \left| h_{1j}\right|} \right) \left( \frac{v_{1j}}{\left| h_{1j}\right|} \right)^2 \]  
(53)
There are unique set \( p_{1j}^*, p_{2j}^* \) that satisfy the KKT conditions when the problem is strictly concave, corresponding to the global maximum.

VI. NUMERICAL RESULTS

In this section we will introduce few results of the two-user MAC with arbitrary inputs for the specific and simplified case where the two user’s channels are SISO. We used Monte-Carlo method to generate the results. Let \( h_1 = 1 \), and \( h_2 = 1 \).

Figure 1 illustrates the decay in the mutual information at the 45° when the two BPSK inputs cancel each other as they lie in the null space of the channel; called the Voronoi region. If the channel states or the total powers are different, the decay will be shifted into another line in the coordinate axis. However, we can easily check that if we induce orthogonality, for example if \( h_2 = 1 \) and \( h_2 = 1 \), the mutual information will be the maximum achievable one without this decay region.

In Figure 2 we show the total MMSE for the two-user MAC. It is interesting to notice that the MMSE behavior corresponds to swapping the mutual information curve in Figure 1. This can let us visualize the relation between the gradient of the mutual information and the MMSE [4], [6]. For the case of Gaussian inputs the relation still holds, but the decay encountered in the arbitrary input setup will not exist, this can be easily verified.
And the interference from user 2 to user 1 can be interpreted as,
\[ \mathbf{H}_2 \mathbf{H}_1 \mathbf{P}_1^* \mathbf{x}_1 \mathbf{x}_2^+ \]  
(55)

We can write (55) for the SISO input case as,
\[ \frac{1}{SNR_2} |h_2| |h_1| \text{COV}(SNR_2 |h_2| |h_1| \mathbf{P}_1^*) \]

Figure 3 and Figure 4 illustrate the MMSE per user in the two-user MAC channel. MMSE1 and MMSE2 correspond to the mean square error of the first and second user respectively. It is worth to note that the concatenation of the two MMSEs leads to the total MMSE in Figure 2.

Finally, Figure 5 presents the optimal power allocation for the two-user MAC. When the total power for the first user is larger than that for the second user, the optimal power allocation for the first will be to allocate its total power, as far as the second user doesn’t extremely interfere, however, for the second user, the optimal power allocation will start with an allocation of the total power, then iteratively optimize it by decreasing it. Therefore, precoding is of great importance to such cases where we can align the interference and maximize the information rates.

VII. CONCLUSIONS

In this paper, we studied the two-user MAC with arbitrary inputs. We derived the relation between the gradient of the mutual information and the MMSE for any input setup. We build upon derivations for the optimal precoding and optimal power allocation. We specialize our results to different cases for the input users, where for the simple case when each user transmits over a SISO channel, the relations can be mapped to the one for the linear vector Gaussian channels with mutually interfering inputs [1], [4] and then, the optimal precoder can follow a fixed point equation like in [3]. A more simplification to the setup where both inputs doesn’t interfere with each other leads to the mercury/waterfilling interpretation for the single user case with binary inputs [7]. This casts further insights to have an interpretation of the interference as discussed. As well, it casts further research questions in quantifying the losses incurred in the mutual information due to interference with and without precoding, and in conjunction with error correcting codes.
\[
\frac{\partial I(x_1, x_2; y)}{\partial H_1^+} = \int \left( (1 + \log p(y)) E_{x_1,x_2} \left( \nabla_y p_y(x_1, x_2) (y|x_1, x_2) \right) x_1^t P_1^+ dy \right) \\
= E \left( \left( \int (1 + \log p(y)) p_y(x_1, x_2) (y|x_1, x_2) \right) x_1^t P_1^+ \right)
\]

Using integration by parts applied to the real and imaginary parts of \( y \),
\[
\int (1 + \log p(y)) \frac{\partial p_y(x_1, x_2)(y|x_1, x_2)}{\partial t} dt \\
= \int (1 + \log p(y)) p_y(x_1, x_2) (y|x_1, x_2) \big|_0^\infty - \int_0^\infty \frac{\partial p(y)}{\partial t} p_y(x_1, x_2) (y|x_1, x_2) dt
\]

The first term goes to zero as \( ||y|| \to \infty \), Then,
\[
\frac{\partial I(x_1, x_2; y)}{\partial H_1^+} = E \left( - \int \frac{p_y(x_1, x_2)(y|x_1, x_2)}{p(y)} \nabla_y p(y) dy \right) x_1^t P_1^+ \\
= - \int \nabla_y p(y) E_{x_1,x_2} \left( \frac{p_y(x_1, x_2)(y|x_1, x_2)}{p(y)} \right) x_1^t P_1^+ dy \\
= - \int \nabla_y p(y) E(x_1|y) P_1^+ dy
\]

However,
\[
\nabla_y p(y) = \nabla_y E_{x_1,x_2} \left( \frac{p_y(x_1, x_2)(y|x_1, x_2)}{p(y)} \right) \\
= E_{x_1,x_2} \left( \nabla_y p_{y|x_1,x_2}(y|x_1, x_2) \right) \\
= -E_{x_1,x_2} \left( p_{y|x_1,x_2}(y|x_1, x_2) (y - H_1 P_1 x_1 - H_2 P_2 x_2) \right)
\]

Thus,
\[
\frac{\partial I(x_1, x_2; y)}{\partial H_1^+} = \int p_y(y) \left( y - H_1 P_1 E(x_1|y) - H_2 P_2 E(x_2|y) \right) x_1^t P_1^+ dy \\
= E(y|x_1) P_1^+ - E(H_1 P_1 E(x_1|y) E(x_1|y)^t) P_1^+ \\
- E(H_2 P_2 E(x_2|y) E(x_1|y)^t) P_1^+
\]

Therefore,
\[
\frac{\partial I(x_1, x_2; y)}{\partial H_1^+} = H_1 P_1 E(x_1|x_1^t) P_1^+ \\
- H_2 P_2 E(x_2|y) E(x_1|y)^t) P_1^+
\]

Similarly, we can derive the gradient of the mutual information in terms of the channel matrix that corresponds to the second user in the two-user MAC.