Mutual Information and Minimum Mean-Square Error in Multiuser Gaussian Channels

Samah A. M. Ghanem
Mobile Communications Department, Eurecom,
Sophia Antipolis, France
Email: samah.ghanem@eurecom.fr

1. Introduction

Due to the lack of explicit closed form expressions of the mutual information for binary inputs, which were provided only for the BPSK and QPSK for the single input single output (SISO) case, [1], [2], [3], it is of particular importance to address connections between information theory and estimation theory for the multiuser case.

Connections between information theory and estimation theory dates back to the work of Duncan, in [4] who showed that for the continuous-time additive white Gaussian noise (AWGN) channel, the filtering minimum mean squared error (causal estimation) is twice the input-output mutual information for any underlying signal distribution. Recently, Guo, Shamai, and Verdu have illuminated intimate connections between information theory and estimation theory in a seminal paper, [1]. In particular, Guo et al. have shown that in the classical problem of information transmission through the conventional AWGN channel, the derivative of the mutual information with respect to the SNR is equal to the smoothing minimum mean squared error (noncausal estimation); a relationship that holds for scalar, vector, discrete-time and continuous-time channels regardless of the input statistics. There have been extensions of these results to the case of mismatched input distributions in the scalar Gaussian channel in [5] and [6].

However, the fundamental relation between the derivative of the mutual information and the MMSE, known as I-MMSE identity, and defined for point to point channels with any noise or input distributions in [1] is not anymore suitable for the multiuser case. Therefore, in this paper, we revisit the connections between the mutual information and the MMSE for the multiuser setup. We generalize the I-MMSE relation to the multiuser case. In particular, we prove that the derivative of the mutual information with respect to the signal to noise ratio (SNR) is equal to the minimum mean squared error (MMSE) plus a covariance induced due to the interference, quantified by a term with respect to the cross correlation of the users inputs' estimates, their channels, and their precoding matrices. Further, we capitalize on this unveiled multiuser I-MMSE relation to derive the components of the multiuser mutual information. In particular, we derive the derivative of the conditinal and non-conditional mutual information with respect to the SNR.

Further extensions of this result allows a generalization of the relations of linear vector Gaussian channels in [7] to multiuser channels. In particular, [8], [9] generalize the I-MMSE relation to the per-user gradient of the conditional, non-conditional and joint mutual information with respect to the MMSE, channels and precoders (power allocation) matrices of the user and the interferer.

Such new unveiled relation allows, the derivation of new closed form expressions of the mutual information for single user and multiuser channels driven by BPSK/QPSK inputs, and to provide asymptotic expansions of the mutual information and the MMSE for multiuser setups, [10].

Throughout the paper, the following notation is employed, boldface uppercase letters denote matrices, lowercase letters denote scalars. The superscript, $(.)^{-1}$, $(.)^{T}$, $(.)^{*}$, and $(.)^{\dagger}$ denote the inverse, transpose, conjugate, and conjugate transpose operations. The (∇) denotes the gradient of a scalar function with respect to a variable. The $\mathbb{E}[.]$ denotes the expectation operator. The $||.||$ and $Tr\left\{.\right\}$ denote the Euclidean norm, and the trace of a matrix, respectively.

The rest of the paper is organized as follows; section 2 introduces the system model. Section 3 introduces the new fundamental relation between the multiuser mutual information and the MMSE. Section 4 provides the conditional and non-conditional components of the I-MMSE identity.

2. System Model

Consider the deterministic complex-valued vector channel,

$$\mathbf{y} = \sqrt{snr}\ \mathbf{H_1 P_1 x_1} + \sqrt{snr}\ \mathbf{H_2 P_2 x_2} + \mathbf{n}, \qquad (1)$$

where the $n_r \times 1$ dimensional vector $\mathbf{y}$ and the $n_t \times 1$ dimensional vectors $\mathbf{x_1}$, $\mathbf{x_2}$ represent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vectors from each user input to the multiuser channel[1]. The distributions of both inputs are not fixed, not necessarily Gaussian nor identical. The $n_r \times n_t$ complex-valued matrices $\mathbf{H_1}$, $\mathbf{H_2}$ correspond to the deterministic channel gains for both input channels (known to both encoder and decoder) and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the $n_r \times 1$ dimensional complex Gaussian noise with independent zero-mean unit-variance components. The $n_t \times n_t$ $\mathbf{P_1}$, $\mathbf{P_2}$ are precoding matrices that do not increase the transmitted power.

3. New Fundamental Relation between the Mutual Information and the MMSE

The first contribution is given in the following theorem, which provides a generalization of the I-MMSE identity to the multiuser case.

[1] We consider the two-user case for ease of exploitation. However, the relations apply to the k-user case.

Theorem 1: The relation between the derivative of the mutual information with respect to the SNR and the non-linear MMSE for a multiuser Gaussian channel satisfies:

$$\frac{dI(snr)}{dsnr} = mmse(snr) + \psi(snr) \tag{2}$$

Where,

$$mmse(snr) = Tr\left\{\mathbf{H_1P_1E_1(H_1P_1)}^\dagger\right\} + Tr\left\{\mathbf{H_2P_2E_2(H_2P_2)}^\dagger\right\}, \tag{3}$$

$$\psi(snr) = - Tr\left\{\mathbf{H_1P_1}\mathbb{E}_\mathbf{y}[\mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}]\mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}]^\dagger](\mathbf{H_2P_2})^\dagger\right\} - Tr\left\{\mathbf{H_2P_2}\mathbb{E}_\mathbf{y}[\mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}]\mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}]^\dagger](\mathbf{H_1P_1})^\dagger\right\},$$

Proof: See Appendix A ■

The per-user MMSE is given respectively as follows:

$$\mathbf{E_1} = \mathbb{E}_\mathbf{y}[(\mathbf{x_1} - \widehat{\mathbf{x}}_\mathbf{1})(\mathbf{x_1} - \widehat{\mathbf{x}}_\mathbf{1})^\dagger] \tag{4}$$

$$\mathbf{E_2} = \mathbb{E}_\mathbf{y}[(\mathbf{x_2} - \widehat{\mathbf{x}}_\mathbf{2})(\mathbf{x_2} - \widehat{\mathbf{x}}_\mathbf{2})^\dagger]. \tag{5}$$

The non-linear input estimates of each user input is given respectively as follows:

$$\widehat{\mathbf{x}}_1 = \mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}] = \sum_{\mathbf{x_1,x_2}} \frac{\mathbf{x_1}p_{y|x_1,x_2}(\mathbf{y|x_1,x_2})p_{x_1}(\mathbf{x_1})p_{x_2}(\mathbf{x_2})}{p_y(\mathbf{y})} \tag{6}$$

$$\widehat{\mathbf{x}}_2 = \mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}] = \sum_{\mathbf{x_1,x_2}} \frac{\mathbf{x_2}p_{y|x_1,x_2}(\mathbf{y|x_1,x_2})p_{x_1}(\mathbf{x_1})p_{x_2}(\mathbf{x_2})}{p_y(\mathbf{y})}. \tag{7}$$

The conditional probability distribution of the Gaussian noise is defined as:

$$p_{y|x_1,x_2}(\mathbf{y|x_1,x_2}) = \frac{1}{\pi^{n_r}} e^{-\left\|\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right\|^2} \tag{8}$$

The probability density function for the received vector $\mathbf{y}$ is defined as:

$$p_y(\mathbf{y}) = \sum_{\mathbf{x_1,x_2}} p_{y|x_1,x_2}(\mathbf{y|x_1,x_2})p_{x_1}(\mathbf{x_1})p_{x_2}(\mathbf{x_2}). \tag{9}$$

Henceforth, the system MMSE with respect to the SNR is given by:

$$mmse(snr) = \mathbb{E}_\mathbf{y}\left[\left\|\mathbf{H_1P_1}(\mathbf{x_1} - \mathbb{E}_{\mathbf{x_1|y}}[\mathbf{x_1|y}])\right\|^\mathbf{2}\right] + \mathbb{E}_\mathbf{y}\left[\left\|\mathbf{H_2P_2}(\mathbf{x_2} - \mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2|y}])\right\|^\mathbf{2}\right], \tag{10}$$

$$= Tr\left\{\mathbf{H_1P_1E_1(H_1P_1)}^\dagger\right\} + Tr\left\{\mathbf{H_2P_2E_2(H_2P_2)}^\dagger\right\} \tag{11}$$

Note that the term $mmse(snr)$ is due to the users MMSEs, particularly, $mmse(snr) = mmse_1(snr) + mmse_2(snr)$ and $\psi(snr)$ are covariance terms that appear due to the covariance of the interferers. Those terms are with respect to the channels, precoders, and non-linear estimates of the user inputs.

When the covariance terms vanish to zero, the mutual information with respect to the SNR will be equal to the MMSE with respect to the SNR, this applies to the relation for the single user and point to point communications. Therefore, the result of Theorem 1 is a generalization of such connection between the two canonical operational measures in information theory and estimation theory - the mutual information and the MMSE - and boils down to the result of Guo et. al, [1] under certain conditions which are: (i) when the cross correlation between the inputs estimates equals zero (ii) when interference can be neglected, and (iii) under the single user setup.

Such generalized fundamental relation between the change in the multiuser mutual information and the SNR is of particular relevance. Firstly, such result allows us to understand the behavior of per-user rates with respect to the interference due to the mutual interference and the interference of other users in terms of their power levels and channel strengths. In addition, the result allows us to be able to quantify the losses incurred due to the interference in terms of bits.

Therefore, when the term $\psi(snr)$ equals zero. The derivative of the mutual information with respect to the SNR equals the total $mmse(snr)$:

$$\frac{dI(snr)}{dsnr} = mmse(snr), \tag{12}$$

which matches the result by Guo et. al in [1].

4. The Conditional and Non-Conditional I-MMSE

In this section, we capitalize on the new fundamental relation to extend the derivative with respect to the SNR to the conditional and non-conditional mutual information. To make this more clear, we capitalize on the chain rule of the mutual information which states the following:

$$I(\mathbf{x_1,x_2;y}) = I(\mathbf{x_1;y}) + I(\mathbf{x_2;y|x_1}) \tag{13}$$

Therefore, through this observation we can conclude the following theorem.

Theorem 2: The relation between the derivative of the conditional and the non-conditional mutual information and their corresponding minimum mean squared error satisfies, respectively:

$$\frac{dI(\mathbf{x_2;y|x_1})}{dsnr} = mmse_2(snr) + \psi(snr) \tag{14}$$

$$\frac{dI(\mathbf{x_1;y})}{dsnr} = mmse_1(\gamma snr) \tag{15}$$

Proof: Taking the derivative of both sides of (13), and subtracting the derivative of $I(\mathbf{x_1;y})$ which is equal to $mmse_1(\gamma snr)$, γ is a scaling factor, due to the fact that x_1 is decoded first considering the other users' input x_2 as noise. Therefore, Theorem 2 has been proved. ■

Of particular relevance is the implication of the derived relations on understanding the achievable rates of interference channels. In particular, such relation allows for better understanding of the changes in the rates due the interferer which is either decoded first or considered as noise. Additionally, further details on the generalized relation that expresses the gradient with respect to arbitrary parameters for the joint, conditional, and non-conditional mutual information can be found in [8], [9].

5. Conclusions

We generalize the fundamental relation between the derivative of the mutual information and the MMSE to multiuser setups. We prove that the derivative of the mutual information with respect to the SNR is equal to the MMSE plus a covariance induced due to the interference, quantified by a term with respect to the cross correlation of the multiuser inputs' estimates, their channels, and their precoding matrices. We provide such relations for conditional and non-conditional components of the multiuser mutual information.

Appendix A: Proof of Theorem 1

The conditional probability density for the two-user multiple access Gaussian channel can be written as follows:

$$p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) = \frac{1}{\pi^{n_r}} e^{-\left\|y-\sqrt{snr}\mathbf{H_1P_1x_1}-\sqrt{snr}\mathbf{H_2P_2x_2}\right\|^2} \quad (16)$$

Thus, the corresponding mutual information is:

$$I(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = \mathbb{E}\left[log\left(\frac{p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{p_y(\mathbf{y})}\right)\right] \quad (17)$$

$$I(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = -n_r log(\pi e) - \mathbb{E}\left[log\left(p_y(\mathbf{y})\right)\right] \quad (18)$$

$$I(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = -n_r log(\pi e) - \int p_y(\mathbf{y}) log\left(p_y(\mathbf{y})\right) d\mathbf{y} \quad (19)$$

Then, the derivative of the mutual information with respect to the SNR is as follows:

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = -\frac{\partial}{\partial snr}\int p_y(\mathbf{y}) log\left(p_y(\mathbf{y})\right) d\mathbf{y} \quad (20)$$

$$= -\int \left(p_y(\mathbf{y})\frac{1}{p_y(\mathbf{y})} + log\left(p_y(\mathbf{y})\right)\right)\frac{\partial p_y(\mathbf{y})}{\partial snr} d\mathbf{y} \quad (21)$$

$$= -\int \left(1 + log\left(p_y(\mathbf{y})\right)\right)\frac{\partial p_y(\mathbf{y})}{\partial snr} d\mathbf{y} \quad (22)$$

Where the probability density function of the received vector $\mathbf{y}$ is given by:

$$p_y(\mathbf{y}) = \sum_{\mathbf{x_1},\mathbf{x_2}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) p_{x_1,x_2}(\mathbf{x_1},\mathbf{x_2}) \quad (23)$$

$$= \mathbb{E}_{x_1,x_2}\left[p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] \quad (24)$$

The derivative of the conditional output with respect to the SNR can be written as:

$$\frac{\partial p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\partial snr} = - p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\times \frac{\partial}{\partial snr}\left(\mathbf{y} - \sqrt{snr}\mathbf{H_1P_1x_1} - \sqrt{snr}\mathbf{H_2P_2x_2}\right)^\dagger \times \left(\mathbf{y} - \sqrt{snr}\mathbf{H_1P_1x_1} - \sqrt{snr}\mathbf{H_2P_2x_2}\right) \quad (25)$$

$$= -\frac{1}{\sqrt{snr}}\left((\mathbf{H_1P_1x_1})^\dagger + (\mathbf{H_2P_2x_2})^\dagger\right)\times \left(\mathbf{y} - \sqrt{snr}\mathbf{H_1P_1x_1} - \sqrt{snr}\mathbf{H_2P_2x_2}\right)\times p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \quad (26)$$

$$= -\frac{1}{\sqrt{snr}}\left((\mathbf{H_1P_1x_1})^\dagger + (\mathbf{H_2P_2x_2})^\dagger\right)\nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \quad (27)$$

Therefore, we have:

$$\mathbb{E}_{x_1,x_2}\left[\nabla_{snr} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = \mathbb{E}_{x_1,x_2}[-\frac{1}{\sqrt{snr}}\left((\mathbf{H_1P_1x_1})^\dagger - (\mathbf{H_2P_2x_2})^\dagger\right)\times \nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})] \quad (28)$$

Substitute (28) into (22), we get:

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = \frac{1}{\sqrt{snr}}\int \left(1 + log\left(p_y(\mathbf{y})\right)\right)\times \mathbb{E}_{x_1,x_2}[\left((\mathbf{H_1P_1x_1})^\dagger + (\mathbf{H_2P_2x_2})^\dagger\right)\times \nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})]d\mathbf{y} \quad (29)$$

$$= \frac{1}{\sqrt{snr}}\mathbb{E}_{x_1,x_2}[(\int \left(1 + log\left(p_y(\mathbf{y})\right)\right)\times \left((\mathbf{H_1P_1x_1})^\dagger + (\mathbf{H_2P_2x_2})^\dagger\right)\times \nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})d\mathbf{y})] \quad (30)$$

Using integration by parts applied to the real and imaginary parts of $\mathbf{y}$ we have:

$$\int \left(1 + log\left(p_y(\mathbf{y})\right)\right)\frac{\partial p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\partial t}d\mathbf{t} = \int \left(1 + log\left(p_y(\mathbf{y})\right)\right) p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\frac{1}{p_y(\mathbf{y})}\frac{\partial p_y(\mathbf{y})}{\partial t} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})d\mathbf{t} \quad (31)$$

The first term in (31) goes to zero as $\|\mathbf{y}\| \to \infty$. Therefore,

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = \frac{1}{\sqrt{snr}}\mathbb{E}_{x_1,x_2}[-\int \left(((\mathbf{H_1P_1x_1})^\dagger + (\mathbf{H_2P_2x_2})^\dagger\right)\times \frac{p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{p_y(\mathbf{y})}\times \nabla_{\mathbf{y}} p_y(\mathbf{y})d\mathbf{y})] \quad (32)$$

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = -\frac{1}{\sqrt{snr}}\int \nabla_{\mathbf{y}} p_y(\mathbf{y})\times \mathbb{E}_{x_1,x_2}[((\mathbf{H_1P_1x_1})^\dagger + (\mathbf{H_2P_2x_2})^\dagger)\times \frac{p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{p_y(\mathbf{y})}]d\mathbf{y} \quad (33)$$

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = -\frac{1}{\sqrt{snr}}\int \nabla_{\mathbf{y}} p_y(\mathbf{y})\times \mathbb{E}_{x_1,x_2}[(\mathbf{H_1P_1})^\dagger \mathbb{E}_{x_1|y}\left[\mathbf{x_1}|\mathbf{y}\right]^\dagger - (\mathbf{H_2P_2})^\dagger \mathbb{E}_{x_2|y}\left[\mathbf{x_2}|\mathbf{y}\right]^\dagger]d\mathbf{y} \quad (34)$$

However,

$$\nabla_{\mathbf{y}} p_y(\mathbf{y}) = \nabla_{\mathbf{y}}\mathbb{E}_{x_1,x_2}\left[p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = \mathbb{E}_{x_1,x_2}\left[\nabla_{\mathbf{y}} p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = -\mathbb{E}_{x_1,x_2}[p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\times \left(\mathbf{y} - \sqrt{snr}\mathbf{H_1P_1x_1} - \sqrt{snr}\mathbf{H_2P_2x_2}\right) = -\mathbb{E}_{x_1,x_2}\left[p_y(\mathbf{y})\left(\mathbf{y} - \sqrt{snr}\mathbf{H_1P_1x_1} - \sqrt{snr}\mathbf{H_2P_2x_2}\right)|\mathbf{y}\right] = -p_y(\mathbf{y})\times (\mathbf{y} - \sqrt{snr}\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}] - \sqrt{snr}\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]) \quad (35)$$

Substitute (35) into (34) we get:

$$\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} = \frac{1}{\sqrt{snr}}\int p_y(\mathbf{y})(\mathbf{y}-\sqrt{snr}\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}] + \sqrt{snr}\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}])\times$$
$$\mathbb{E}_{x_1,x_2}\left((\mathbf{H_1P_1})^\dagger\mathbb{E}_{x_1|y}\left[\mathbf{x_1}|\mathbf{y}\right]^\dagger + (\mathbf{H_2P_2})^\dagger\mathbb{E}_{x_2|y}\left[\mathbf{x_2}|\mathbf{y}\right]^\dagger\right)d\mathbf{y}$$

$$\begin{aligned}\frac{d\mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} &= \frac{1}{\sqrt{\mathbf{snr}}}\mathbb{E}_y[\mathbf{y}\mathbf{x_1^\dagger}](\mathbf{H_1P_1})^\dagger + \frac{1}{\sqrt{snr}}\mathbb{E}_y[\mathbf{y}\mathbf{x_2^\dagger}](\mathbf{H_2P_2})^\dagger \\ &- \mathbb{E}_y[\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \\ &- \mathbb{E}_y[\mathbf{H_1P_1}\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ &- \mathbb{E}_y[\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ &- \mathbb{E}_y[\mathbf{H_2P_2}\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \end{aligned} \quad (36)$$

Therefore,

$$\begin{aligned}\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} &= \mathbf{H_1P_1}\mathbb{E}_{x_1}[\mathbf{x_1x_1^\dagger}](\mathbf{H_1P_1})^\dagger \\ &- \mathbf{H_1P_1}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger(\mathbf{H_1P_1})^\dagger \\ &- \mathbf{H_1P_1}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{\mathbf{x_2|y}}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ &+ \mathbf{H_2P_2}\mathbb{E}_{x_2}[\mathbf{x_2x_2^\dagger}](\mathbf{H_2P_2})^\dagger \\ &- \mathbf{H_2P_2}\mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger(\mathbf{H_2P_2})^\dagger \\ &- \mathbf{H_2P_2}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \end{aligned} \quad (37)$$

According to (4) and (5), (37) simplifies to:

$$\begin{aligned}\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} &= \mathbf{H_1P_1E_1}(\mathbf{H_1P_1})^\dagger + \mathbf{H_2P_2E_2}(\mathbf{H_2P_2})^\dagger \\ &- \mathbf{H_1P_1}\mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]^\dagger](\mathbf{H_2P_2})^\dagger \\ &- \mathbf{H_2P_2}\mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x_2}|\mathbf{y}]\mathbb{E}_{x_1|y}[\mathbf{x_1}|\mathbf{y}]^\dagger](\mathbf{H_1P_1})^\dagger \end{aligned} \quad (38)$$

Therefore, the derivative of the mutual information with respect to the SNR and the per users mmse and input estimates (or covariances) is as follows:

$$\begin{aligned}\frac{dI(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{dsnr} &= mmse_1(snr) + mmse_2(snr) \\ &- Tr\left\{\mathbf{H_1P_1}\mathbb{E}_y[\widehat{\mathbf{x}}_\mathbf{1}\widehat{\mathbf{x}}_\mathbf{2}^\dagger](\mathbf{H_2P_2})^\dagger\right\} \\ &- Tr\left\{\mathbf{H_2P_2}\mathbb{E}_y[\widehat{\mathbf{x}}_\mathbf{2}\widehat{\mathbf{x}}_\mathbf{1}^\dagger](\mathbf{H_1P_1})^\dagger\right\}\end{aligned} \quad (39)$$

Therefore, we can write the derivative of the derivative of the mutual information with respect to the SNR as follows:

$$\frac{dI(snr)}{dsnr} = mmse(snr) + \psi(snr) \quad (40)$$

Therefore, Theorem 1 has been proved as a generalization of the I-MMSE identity to the multiuser case.

References

[1] D. Guo and S. Shamai and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels", *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1282, Apr 2005.

[2] S. A. M. Ghanem, "Mutual Information for Generalized Arbitrary Binary Input Constellations", *MAP-Tele Workshop*, Feb 2010.

[3] S. A. M. Ghanem, "Analysis, Modeling, Design, and Optimization of Future Communications Systems: From Theory to Practice", *Ph.D. Thesis*, Dec 2013.

[4] T. E. Duncan, "On the calculation of mutual information", *SIAM Journal on Applied Mathematics*, vol. 19, pp. 215–220, Jul 1970.

[5] S. Verdu, "Mismatched Estimation and Relative Entropy", *IEEE Transactions on Information Theory*, vol. 56, pp. 3712–3720, Aug 2010.

[6] T. Weissman, "The Relationship Between Causal and Noncausal Mismatched Estimation in Continuous-Time AWGN Channels", *IEEE Transactions on Information Theory*, vol. 56, pp. 4256-4273, Sept 2010.

[7] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels", *IEEE Transactions on Information Theory*, vol. 52, pp. 141–154, Jan 2006.

[8] S. A. M. Ghanem, "MAC Gaussian Channels With Arbitrary Inputs: Optimal Precoding and Power Allocation", *IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2012.

[9] S. A. M. Ghanem, "Multiple Access Gaussian Channels With Arbitrary Inputs: Optimal Precoding and Power Allocation", *http://arxiv.org/abs/1411.0446*, 2014.

[10] S. A. M. Ghanem, "Multiuser I-MMSE", *in preparation*