Hessian and Concavity of Mutual Information, Differential Entropy, and Entropy Power in Linear Vector Gaussian Channels

Miquel Payaró, *Member, IEEE*, and Daniel P. Palomar, *Senior Member, IEEE*

***Abstract*—Within the framework of linear vector Gaussian channels with arbitrary signaling, the Jacobian of the minimum mean square error and Fisher information matrices with respect to arbitrary parameters of the system are calculated in this paper. Capitalizing on prior research where the minimum mean square error and Fisher information matrices were linked to information-theoretic quantities through differentiation, the Hessian of the mutual information and the entropy are derived. These expressions are then used to assess the concavity properties of mutual information and entropy under different channel conditions and also to derive a multivariate version of an entropy power inequality due to Costa.**

***Index Terms*—Concavity properties, differential entropy, entropy power, Fisher information matrix, Gaussian noise, Hessian matrices, linear vector Gaussian channels, minimum mean-square error (MMSE), mutual information, nonlinear estimation.**

I. Introduction and Motivation

THE availability of expressions for the Hessian matrix of the mutual information with respect to arbitrary parameters of the system is useful from a theoretical perspective but also from a practical standpoint. In system design, if the mutual information is to be numerically optimized through a gradient algorithm as in [1], the Hessian matrix may be used alongside the gradient in the Newton's method to speed up the convergence of the algorithm. Additionally, from a system analysis perspective, the Hessian matrix can also complement the gradient in studying the sensitivity of the mutual information to variations of the system parameters and, more importantly, in the cases where the mutual information is concave with respect to the system design parameters, it can also be used to guarantee the global optimality of a given design.

In this sense and within the framework of linear vector Gaussian channels with arbitrary signaling, the purpose of this work is twofold. First, we calculate the Hessian matrix of the mutual information, differential entropy and entropy power and, second, we study the concavity properties of these quantities. Both goals are intimately related since concavity can be assessed through the negative semidefiniteness of the Hessian matrix. As intermediate results of our study, we compute the Jacobian of the minimum mean-square error (MMSE) and Fisher information matrices, which are interesting results in their own right and contribute to the exploration of the fundamental links between information theory and estimation theory.

Initial connections between information- and estimation-theoretic quantities for linear channels with additive Gaussian noise date back to the late fifties: in the proof of Shannon's entropy power inequality [2], Stam used the fact that the derivative of the output differential entropy with respect to the added noise power is equal to the Fisher information of the channel output and attributed this identity to De Bruijn. More than a decade later, the links between both worlds strengthened when Duncan [3] and Kadota, Zakai, and Ziv [4] represented mutual information as a function of the error in causal filtering.

Much more recently, in [5], Guo, Shamai, and Verdú fruitfully explored further these connections and, among other results, proved that the derivative of the mutual information and differential entropy with respect to the signal-to-noise ratio (SNR) is equal to half the MMSE regardless of the input statistics. The main result in [5] was generalized to the abstract Wiener space by Zakai in [6] and by Palomar and Verdú in two different directions: in [1] they calculated the partial derivatives of the mutual information with respect to the channel matrix and other arbitrary parameters of the system through the chain rule and, in [7], they represented the derivative of mutual information as a function of the conditional marginal input given the output for general channels (not necessarily additive Gaussian).

In this paper we build upon the setting of [1], where loosely speaking, it was proved that, for the linear vector Gaussian channel $\boldsymbol{Y} = \mathbf{G}\boldsymbol{S} + \mathbf{C}\boldsymbol{N}$, i) the gradients of the differential entropy $h(\boldsymbol{Y})$ and the mutual information $I(\boldsymbol{S};\boldsymbol{Y})$ with respect to functions of the linear transformation undergone by the input, $\mathbf{G}$, are linear functions of the MMSE matrix, $\mathbf{E}_{\boldsymbol{S}}$, and ii) the gradient of the differential entropy $h(\boldsymbol{Y})$ with respect to the linear transformation undergone by the noise, $\mathbf{C}$, are linear functions of the Fisher information matrix, $\mathbf{J}_{\boldsymbol{Y}}$. We show that the previous two key quantities $\mathbf{E}_{\boldsymbol{S}}$ and $\mathbf{J}_{\boldsymbol{Y}}$, which completely characterize the first-order derivatives, are not enough to describe the second-order derivatives. For that purpose, we introduce the more refined conditional MMSE matrix $\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$

Manuscript received June 25, 2008; revised February 25, 2009. Current version published July 15, 2009. This work was supported by a RGC 618008 Research Grant. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT'08), Toronto, Canada, July2008.

M. Payaró was with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. He is now with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona 08860, Spain (e-mail: miquel.payaro@cttc.es).

D. P. Palomar is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: palomar@ust.hk).

Communicated by I. Kontoyiannis, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2009.2023749

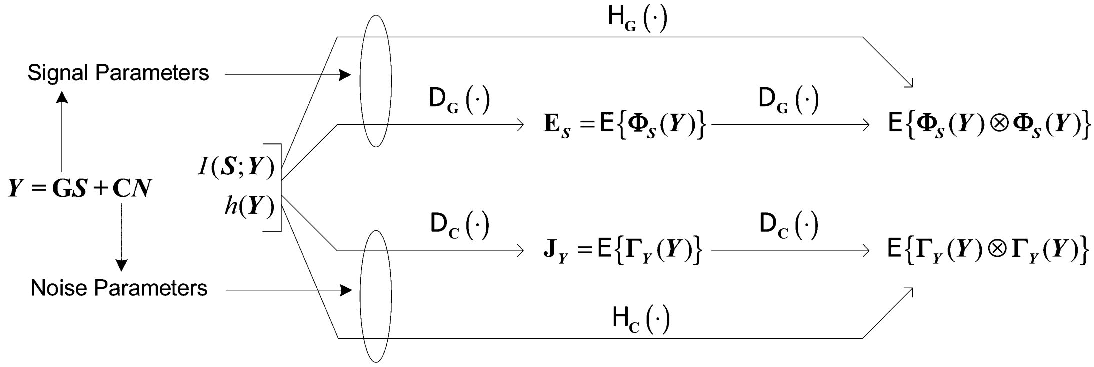

Fig. 1. Simplified representation of the relations between the quantities dealt with in this work. The Jacobian D and Hessian H operators represent first- and second-order differentiation, respectively.

and conditional Fisher information matrix $\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ (when these quantities are averaged with respect to the distribution of the output $\boldsymbol{y}$, we recover $\mathbf{E}_{\boldsymbol{S}} = \mathsf{E}\{\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}$ and $\mathbf{J}_{\boldsymbol{Y}} = \mathsf{E}\{\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}$). In particular, the second-order derivatives depend on $\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$ and $\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ through the terms $\mathsf{E}\{\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y}) \otimes \mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}$ and $\mathsf{E}\{\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}$. See Fig. 1 for a schematic representation of these relations.

Analogous results to some of the expressions presented in this paper particularized to the scalar Gaussian channel were simultaneously derived in [8], [9], where the second and third derivatives of the mutual information with respect to the SNR were calculated.

As an application of the obtained expressions, we show concavity properties of the mutual information and derive a multivariate generalization of the entropy power inequality (EPI) due to Costa [10]. Moreover, our multivariate EPI has already found an application in [11] to derive outer bounds on the capacity region in multiuser channels with feedback.

This paper is organized as follows. In Section II, the model for the linear vector Gaussian channel is given and the quantities dealt with in this work are introduced. The main results of the paper are given in Section III where the expressions for the Jacobian matrix of the MMSE and Fisher information and the Hessian matrix of the mutual information and differential entropy are presented. In Section IV the concavity properties of the mutual information are studied and, finally, in Section V two applications of the results derived in this work are given.

Notation: Straight boldface denote multivariate quantities such as vectors (lowercase) and matrices (uppercase). Uppercase italics denote random variables, and their realizations are represented by lowercase italics. The sets of q-dimensional symmetric, positive semidefinite, and positive definite matrices are denoted by $\mathbb{S}^q$, $\mathbb{S}^q_+$, and $\mathbb{S}^q_{++}$, respectively. The elements of a matrix $\mathbf{A}$ are represented by $\mathbf{A}_{ij}$ or $[\mathbf{A}]_{ij}$ interchangeably, whereas the elements of a vector $\mathbf{a}$ are represented by a_i. The operator $\mathbf{diag}(\mathbf{A})$ represents a column vector with the diagonal entries of matrix $\mathbf{A}$, $\mathbf{Diag}(\mathbf{A})$ and $\mathbf{Diag}(\mathbf{a})$ represent a diagonal matrix whose nonzero elements are given by the diagonal elements of matrix $\mathbf{A}$ and by the elements of vector $\mathbf{a}$, respectively, and $\mathsf{vec}\mathbf{A}$ represents the vector obtained by stacking the columns of $\mathbf{A}$. For symmetric matrices, $\mathsf{vech}\mathbf{A}$ is obtained from $\mathsf{vec}\mathbf{A}$ by eliminating the repeated elements located above the main diagonal of $\mathbf{A}$. The Kronecker matrix product is represented by $\mathbf{A} \otimes \mathbf{B}$ and the Schur (or Hadamard) element-wise matrix product is denoted by $\mathbf{A} \circ \mathbf{B}$. The superscripts $(\cdot)^\mathsf{T}$, $(\cdot)^\dagger$, and $(\cdot)^+$, denote transpose, Hermitian, and Moore–Penrose pseudoinverse operations, respectively. With a slight abuse of notation, we consider that when square root or multiplicative inverse are applied to a vector, they act upon the entries of the vector, we thus have $[\sqrt{\mathbf{a}}]_i = \sqrt{\mathrm{a}_i}$ and $[1/\mathbf{a}]_i = 1/\mathrm{a}_i$.

II. Signal Model

We consider a general discrete-time linear vector Gaussian channel, whose output $\boldsymbol{Y} \in \mathbb{R}^n$ is represented by the model

$$\boldsymbol{Y} = \mathbf{G}\boldsymbol{S} + \boldsymbol{Z} \tag{1}$$

where $\boldsymbol{S} \in \mathbb{R}^m$ is the zero-mean channel input vector with covariance matrix $\mathbf{R}_{\boldsymbol{S}}$, the matrix $\mathbf{G} \in \mathbb{R}^{n\times m}$ specifies the linear transformation undergone by the input vector, and $\boldsymbol{Z} \in \mathbb{R}^n$ represents a zero-mean Gaussian noise with nonsingular covariance matrix $\mathbf{R}_{\boldsymbol{Z}}$.

The channel transition probability density function corresponding to the channel model in (1) is

$$P_{\boldsymbol{Y}\,|\,\boldsymbol{S}}(\boldsymbol{y}\,|\,\boldsymbol{s}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{y}-\mathbf{G}\boldsymbol{s})^\mathsf{T}\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\mathbf{G}\boldsymbol{s})\right)}{\sqrt{(2\pi)^n \det(\mathbf{R}_{\boldsymbol{Z}})}} \tag{2}$$

and the marginal probability density function of the output is given by $P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathsf{E}\{P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\}$, which is an infinitely differentiable continuous function of $\boldsymbol{y}$ regardless of the distribution of the input vector $\boldsymbol{S}$ thanks to the smoothing properties of the added noise [10, Sec. II].

At some points, it may be convenient to write $\boldsymbol{X} = \mathbf{G}\boldsymbol{S}$ and also express the noise vector as $\boldsymbol{Z} = \mathbf{C}\boldsymbol{N}$, where $\mathbf{C} \in \mathbb{R}^{n\times n'}$, such that $n' \geq n$, and where the noise covariance matrix $\mathbf{R}_{\boldsymbol{Z}} = \mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^\mathsf{T}$ has an inverse so that (2) is meaningful.

In the following, we describe the information- and estimation-theoretic quantities whose relations we are interested in.

A. Differential Entropy and Mutual Information

The differential entropy of the continuous random vector $\boldsymbol{Y}$ is defined as $h(\boldsymbol{Y}) = -\mathsf{E}\{\log P_{\boldsymbol{Y}}(\boldsymbol{Y})\}$ [12].[1] For the linear vector

[1]Throughout this paper we work with natural logarithms and thus nats are used as information units.

Gaussian channel in (1), the input–output mutual information is [12]

$$I(\boldsymbol{S};\boldsymbol{Y}) = h(\boldsymbol{Y}) - \frac{1}{2}\mathsf{logdet}(2\pi e\mathbf{R}_{\boldsymbol{Z}}) \quad (3)$$

B. Minimum Mean Square Error (MSE) Matrix

We consider the estimation of the input signal $\boldsymbol{S}$ based on the observation of a realization of the output $\boldsymbol{Y} = \boldsymbol{y}$. The estimator that simultaneously achieves the minimum mean square error (MSE) for all the components of the estimation error vector is given by the conditional mean estimator $\hat{\boldsymbol{S}}(\boldsymbol{y}) = \mathsf{E}\{\boldsymbol{S} \,|\, \boldsymbol{y}\}$ and the corresponding MSE matrix, referred to as the MMSE matrix, is

$$\mathbf{E}_{\boldsymbol{S}} = \mathsf{E}\{(\boldsymbol{S} - \mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{Y}\})(\boldsymbol{S} - \mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{Y}\})^{\mathsf{T}}\}. \quad (4)$$

An alternative and useful expression for the MMSE matrix can be obtained by considering first the MMSE matrix conditioned on a specific realization of the output $\boldsymbol{Y} = \boldsymbol{y}$, which is denoted by $\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$ and defined as

$$\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y}) = \mathsf{E}\{(\boldsymbol{S} - \mathsf{E}\{\boldsymbol{S}|\boldsymbol{y}\})(\boldsymbol{S} - \mathsf{E}\{\boldsymbol{S}|\boldsymbol{y}\})^{\mathsf{T}}\,|\,\boldsymbol{y}\}. \quad (5)$$

Observe from (5) that $\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$ is a positive semidefinite matrix. Finally, the MMSE matrix in (4) can be obtained by taking the expectation of $\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$ with respect to the distribution of the output, $\mathbf{E}_{\boldsymbol{S}} = \mathsf{E}\{\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}$.

C. Fisher Information Matrix

Besides the MMSE matrix, another quantity that is closely related to the differential entropy is the Fisher information matrix. For an arbitrary random vector $\boldsymbol{Y}$, the Fisher information matrix with respect to a translation parameter is [13]

$$\mathbf{J}_{\boldsymbol{Y}} = \mathsf{E}\left\{\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}}\mathsf{log}P_y(\boldsymbol{Y})\mathsf{D}_{\boldsymbol{y}}\mathsf{log}P_y(\boldsymbol{Y})\right\} \quad (6)$$

where D is the Jacobian operator. This operator together with the Hessian operator, H, and other definitions and conventions used for differentiation with respect to multidimensional parameters are described in Appendixes A and B.

The expression of the Fisher information in (6) in terms of the Jacobian of $\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y})$ can be transformed into an expression in terms of its Hessian matrix, thanks to the logarithmic identity

$$\mathsf{H}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\mathsf{H}_{\boldsymbol{y}}P_{\boldsymbol{Y}}(\boldsymbol{y})}{P_{\boldsymbol{Y}}(\boldsymbol{y})} - \mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{D}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y}) \quad (7)$$

together with $\mathsf{E}\{\mathsf{H}_{\boldsymbol{y}}P_{\boldsymbol{Y}}(\boldsymbol{Y})/P_{\boldsymbol{Y}}(\boldsymbol{Y})\} = \int \mathsf{H}_{\boldsymbol{y}}P_{\boldsymbol{Y}}(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \mathbf{0}$, which follows directly from the expression for $\mathsf{H}_{\boldsymbol{y}}P_{\boldsymbol{Y}}(\boldsymbol{y})$ in (75) in Appendix C. The alternative expression for the Fisher information matrix is then

$$\mathbf{J}_{\boldsymbol{Y}} = -\mathsf{E}\{\mathsf{H}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{Y})\}. \quad (8)$$

Similarly to the previous section with the MMSE matrix, it will be useful to define a conditional form of the Fisher information matrix $\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$, in such a way that $\mathbf{J}_{\boldsymbol{Y}} = \mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}$. At this point, it may not be clear which of the two forms (6) or (8) will be more useful for the rest of the paper; we advance that defining $\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ based on (8) will prove more convenient

$$\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y}) = -\mathsf{H}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathbf{R}_{\boldsymbol{Z}}^{-1} - \mathbf{R}_{\boldsymbol{Z}}^{-1}\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})\mathbf{R}_{\boldsymbol{Z}}^{-1} \quad (9)$$

where the second equality is proved in Lemma C.4 in Appendix C and where we have $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y}) = \mathbf{G}\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}^{\mathsf{T}}$.

D. Prior Known Relations

The first known relation between the above described quantities is the De Bruijn identity [2] (see also the alternative derivation in [5]), which couples the Fisher information with the differential entropy according to

$$\frac{\mathrm{d}}{\mathrm{d}t}h(\boldsymbol{X} + \sqrt{t}\boldsymbol{Z}) = \frac{1}{2}\mathsf{Tr}\mathbf{J}_{\boldsymbol{Y}} \quad (10)$$

where in this case $\boldsymbol{Y} = \boldsymbol{X} + \sqrt{t}\boldsymbol{Z}$. A multivariate extension of the De Bruijn identity was found in [1] as

$$\nabla_{\mathbf{C}}h(\boldsymbol{X} + \mathbf{C}\boldsymbol{N}) = \mathbf{J}_{\boldsymbol{Y}}\mathbf{C}\mathbf{R}_{\boldsymbol{N}}. \quad (11)$$

In [5], the more canonical operational measures of mutual information and MMSE were coupled through the identity

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}}I(\boldsymbol{S};\sqrt{\mathsf{snr}}\boldsymbol{S} + \boldsymbol{Z}) = \frac{1}{2}\mathsf{Tr}\mathbf{E}_{\boldsymbol{S}} \quad (12)$$

which was generalized to the multivariate case in [1], yielding

$$\nabla_{\mathbf{G}}I(\boldsymbol{S};\mathbf{G}\boldsymbol{S} + \boldsymbol{Z}) = \mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{G}\mathbf{E}_{\boldsymbol{S}}. \quad (13)$$

From these previous existing results, we realize that the differential entropy function $h(\boldsymbol{X} + \mathbf{C}\boldsymbol{N})$ is related to the Fisher information matrix $\mathbf{J}_{\boldsymbol{Y}}$ through differentiation with respect to the transformation $\mathbf{C}$ undergone by the Gaussian noise $\boldsymbol{N}$ as in (11) and that the mutual information $I(\boldsymbol{S};\mathbf{G}\boldsymbol{S} + \boldsymbol{Z})$ is related to the MMSE matrix $\mathbf{E}_{\boldsymbol{S}}$ through differentiation with respect to the transformation $\mathbf{G}$ undergone by the signal $\boldsymbol{S}$ as in (13) (see also Fig. 1). A comprehensive account of other relations can be found in [5].

Since we are interested in calculating the Hessian matrix of differential entropy and mutual information, in the light of the results in (11) and (13), it is instrumental to first calculate the Jacobian matrix of the MMSE and Fisher information matrices, as considered in the next section.

III. Jacobian and Hessian Results

A. Jacobians of the Fisher Information and MMSE Matrices

As a warm-up, consider first our signal model with Gaussian signaling, $\boldsymbol{Y}_{\mathcal{G}} = \boldsymbol{X}_{\mathcal{G}} + \mathbf{C}\boldsymbol{N}$. In this case, the conditional Fisher information matrix defined in (9) does not depend on the realization of the received vector $\boldsymbol{y}$ and is (e.g., [14, Appendix 3C])

$$\boldsymbol{\Gamma}_{\boldsymbol{Y}_{\mathcal{G}}} = (\mathbf{R}_{\boldsymbol{X}_{\mathcal{G}}} + \mathbf{R}_{\boldsymbol{Z}})^{-1} = \left(\mathbf{R}_{\boldsymbol{X}_{\mathcal{G}}} + \mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}}\right)^{-1}. \quad (14)$$

Consequently, we have that $\mathbf{J}_{\boldsymbol{Y}_{\mathcal{G}}} = \mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}_{\mathcal{G}}}\} = \boldsymbol{\Gamma}_{\boldsymbol{Y}_{\mathcal{G}}}$.

The Jacobian matrix of the Fisher information matrix with respect to the noise transformation $\mathbf{C}$ can be obtained as

$$\begin{aligned}\mathsf{D}_{\mathbf{C}}\mathbf{J}_{\boldsymbol{Y}_{\mathcal{G}}} &= \mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}_{\mathcal{G}}} \cdot \mathsf{D}_{\mathbf{C}}\mathbf{R}_{\boldsymbol{Z}} \quad (15)\\ &= -\mathbf{D}_n^{+}\left(\mathbf{J}_{\boldsymbol{Y}_{\mathcal{G}}} \otimes \mathbf{J}_{\boldsymbol{Y}_{\mathcal{G}}}\right)\mathbf{D}_n \\ &\quad \cdot 2\mathbf{D}_n^{+}\left(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n\right) \quad (16)\\ &= -2\mathbf{D}_n^{+}\left(\mathbf{J}_{\boldsymbol{Y}_{\mathcal{G}}} \otimes \mathbf{J}_{\boldsymbol{Y}_{\mathcal{G}}}\right)\left(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n\right) \quad (17)\\ &= -2\mathbf{D}_n^{+}\mathsf{E}\left\{\mathbf{\Gamma}_{\boldsymbol{Y}_{\mathcal{G}}} \otimes \mathbf{\Gamma}_{\boldsymbol{Y}_{\mathcal{G}}}\right\}\left(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n\right) \quad (18)\end{aligned}$$

where (15) follows from the Jacobian chain rule in Lemma B.5; in (16) we have applied Lemmas B.7.4 and B.7.5 with $\mathbf{D}_n$ being the duplication matrix defined in Appendix A[2]; and finally (17) follows from the fact that $\mathbf{D}_n^{+}(\mathbf{A}\otimes\mathbf{A})\mathbf{D}_n\mathbf{D}_n^{+} = \mathbf{D}_n^{+}(\mathbf{A}\otimes\mathbf{A})$, which can be obtained from (56) and (58) in Appendix A.

In the following theorem, we generalize (18) for the case of arbitrary signaling.

Theorem 1 (Jacobian of the Fisher Information Matrix): Consider the signal model $\boldsymbol{Y} = \boldsymbol{X} + \mathbf{C}\boldsymbol{N}$, where $\mathbf{C}$ is an arbitrary deterministic matrix, the signaling $\boldsymbol{X}$ is arbitrarily distributed, and the noise vector $\boldsymbol{N}$ is Gaussian and independent of the input $\boldsymbol{X}$. Then, the Jacobian of the Fisher information matrix of the n-dimensional output vector $\boldsymbol{Y}$ is

$$\mathsf{D}_{\mathbf{C}}\mathbf{J}_{\boldsymbol{Y}} = -2\mathbf{D}_n^{+}\mathsf{E}\{\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n) \quad (19)$$

where $\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ is defined in (9).

Proof: See Appendix D. □

Remark 1: Due to the fact that, in general, the conditional Fisher information matrix $\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ does depend on the particular value of the observation $\boldsymbol{y}$, it is not possible to express the expectation of the Kronecker product as the Kronecker product of the expectations, as in (17) for the Gaussian signaling case.

Now that Jacobian of the Fisher information matrix has been presented, we proceed with the Jacobian of the MMSE matrix.

Theorem 2 (Jacobian of the MMSE Matrix): Consider the signal model $\boldsymbol{Y} = \mathbf{G}\boldsymbol{S} + \boldsymbol{Z}$, where $\mathbf{G}$ is an arbitrary deterministic matrix, the m-dimensional signaling vector $\boldsymbol{S}$ is arbitrarily distributed, and the noise vector $\boldsymbol{Z}$ is Gaussian and independent of the input $\boldsymbol{S}$. Then, the Jacobian of the MMSE matrix of the input vector $\boldsymbol{S}$ is

$$\mathsf{D}_{\mathbf{G}}\mathbf{E}_{\boldsymbol{S}} = -2\mathbf{D}_m^{+}\mathsf{E}\{\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y}) \otimes \mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}\left(\mathbf{I}_m \otimes \mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right) \quad (20)$$

where $\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$ is defined in (5).

Proof: See Appendix D. □

Remark 2: In light of the two results in Theorems 1 and 2, it is now apparent that $\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ plays an analogous role in the differentiation of the Fisher information matrix as the one played by the conditional MMSE matrix $\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$ when differentiating the MMSE matrix, which justifies the choice made in Section II.C of identifying $\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ with the expression in (8) and not with the expression in (6).

[2]The matrix $\mathbf{D}_n$ appears in (18) and in many successive expressions because we are explicitly taking into account the fact that $\mathbf{J}_{\boldsymbol{Y}}$ is a symmetric matrix.

B. Jacobians With Respect to Arbitrary Parameters

With the basic results for the Jacobian of the MMSE and Fisher information matrices in Theorems 1 and 2, the Jacobians with respect to arbitrary parameters of the system can be found through the chain rule for differentiation. Precisely, we are interested in considering the case where the linear transformation undergone by the signal is decomposed as the product of two linear transformations, $\mathbf{G} = \mathbf{H}\mathbf{P}$, where $\mathbf{H}$ represents the channel, which is externally determined by the propagation environment conditions, and $\mathbf{P}$ represents the linear precoder, which is specified by the system designer.

Theorem 3 (Jacobians With Respect to Arbitrary Parameters): Consider the signal model $\boldsymbol{Y} = \mathbf{H}\mathbf{P}\boldsymbol{S} + \mathbf{C}\boldsymbol{N}$, where $\mathbf{H} \in \mathbb{R}^{n\times p}$, $\mathbf{P} \in \mathbb{R}^{p\times m}$, and $\mathbf{C} \in \mathbb{R}^{n\times n'}$, with $n' \geq n$, are arbitrary deterministic matrices, the signaling $\boldsymbol{S} \in \mathbb{R}^m$ is arbitrarily distributed, the noise $\boldsymbol{N} \in \mathbb{R}^{n'}$ is Gaussian, independent of the input $\boldsymbol{S}$, and has covariance matrix $\mathbf{R}_{\boldsymbol{N}}$, and the total noise, defined as $\boldsymbol{Z} = \mathbf{C}\boldsymbol{N} \in \mathbb{R}^n$, has a positive definite covariance matrix given by $\mathbf{R}_{\boldsymbol{Z}} = \mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}}$. Then, the MMSE and Fisher information matrices satisfy

$$\begin{aligned}\mathsf{D}_{\mathbf{P}}\mathbf{E}_{\boldsymbol{S}} &= -2\mathbf{D}_m^{+}\mathsf{E}\{\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y}) \otimes \mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}\\ &\quad \cdot \left(\mathbf{I}_m \otimes \mathbf{P}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\right) \quad (21)\\ \mathsf{D}_{\mathbf{H}}\mathbf{E}_{\boldsymbol{S}} &= -2\mathbf{D}_m^{+}\mathsf{E}\{\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y}) \otimes \mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}\\ &\quad \cdot \left(\mathbf{P}^{\mathsf{T}} \otimes \mathbf{P}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right) \quad (22)\\ \mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}} &= -\mathbf{D}_n^{+}\mathsf{E}\{\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}\mathbf{D}_n \quad (23)\\ \mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{J}_{\boldsymbol{Y}} &= -\mathbf{D}_n^{+}\mathsf{E}\{\mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \mathbf{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}\\ &\quad \cdot (\mathbf{C}\otimes\mathbf{C})\mathbf{D}_{n'}. \quad (24)\end{aligned}$$

Proof: See Appendix D. □

C. Hessian of Differential Entropy and Mutual Information

Now that we have obtained the Jacobians of the MMSE and Fisher information matrices, we will capitalize on the results in [1] to obtain the Hessians of the mutual information $I(\boldsymbol{S};\boldsymbol{Y})$ and the differential entropy $h(\boldsymbol{Y})$.

Lemma 1 (Entropy Jacobians [1]): Consider the setting of Theorem 3. Then, the differential entropy $h(\boldsymbol{Y})$ satisfies

$$\begin{aligned}\mathsf{D}_{\mathbf{P}}h(\boldsymbol{Y}) &= \mathrm{vec}^{\mathsf{T}}\left(\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{P}\mathbf{E}_{\boldsymbol{S}}\right) \quad (25)\\ \mathsf{D}_{\mathbf{H}}h(\boldsymbol{Y}) &= \mathrm{vec}^{\mathsf{T}}\left(\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{P}\mathbf{E}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}}\right) \quad (26)\\ \mathsf{D}_{\mathbf{C}}h(\boldsymbol{Y}) &= \mathrm{vec}^{\mathsf{T}}\left(\mathbf{J}_{\boldsymbol{Y}}\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\right) \quad (27)\\ \mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}h(\boldsymbol{Y}) &= \frac{1}{2}\mathrm{vec}^{\mathsf{T}}\left(\mathbf{J}_{\boldsymbol{Y}}\right)\mathbf{D}_n \quad (28)\\ \mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}h(\boldsymbol{Y}) &= \frac{1}{2}\mathrm{vec}^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}}\mathbf{J}_{\boldsymbol{Y}}\mathbf{C})\mathbf{D}_{n'}. \quad (29)\end{aligned}$$

Remark 3: Equations (25) and (26) are also valid if the differential entropy $h(\boldsymbol{Y})$ is replaced by the mutual information $I(\boldsymbol{S};\boldsymbol{Y})$. Alternatively, the expressions (27), (28), and (29) do not hold verbatim for the mutual information because, in that case, the differential entropy of the noise vector in (3) does depend on $\mathbf{R}_{\boldsymbol{Z}}$ and, implicitly, on $\mathbf{C}$ and $\mathbf{R}_{\boldsymbol{N}}$. Then, applying basic Jacobian results from [15, Ch. 9], we have

$$\mathsf{D}_{\mathbf{C}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{D}_{\mathbf{C}} h(\boldsymbol{Y}) - \mathsf{vec}^{\mathsf{T}}((\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1}\mathbf{C}\mathbf{R}_{\boldsymbol{N}}) \quad (30)$$

$$\mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}} h(\boldsymbol{Y}) - \frac{1}{2}\mathsf{vec}^{\mathsf{T}}\left(\mathbf{R}_{\boldsymbol{Z}}^{-1}\right)\mathbf{D}_n \quad (31)$$

$$\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}} h(\boldsymbol{Y}) - \frac{1}{2}\mathsf{vec}^{\mathsf{T}}\left(\mathbf{C}^{\mathsf{T}}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1}\mathbf{C}\right)\mathbf{D}_{n'}. \quad (32)$$

With Lemma 1 at hand, and the expressions obtained in the previous section for the Jacobian matrices of the Fisher information and the MMSE matrices, we are ready to calculate the Hessian matrix.

Theorem 4 (Entropy Hessians): Consider the setting of Theorem 3. Then, the differential entropy of the output vector $\boldsymbol{Y}$, $h(\boldsymbol{Y})$, satisfies

$$\mathsf{H}_{\mathbf{P}} h(\boldsymbol{Y}) = \left(\mathbf{E}_{\boldsymbol{S}} \otimes \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\right) - 2\left(\mathbf{I}_m \otimes \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{P}\right)\mathbf{N}_m \cdot \mathsf{E}\{\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y}) \otimes \boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}\left(\mathbf{I}_m \otimes \mathbf{P}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\right) \quad (33)$$

$$\mathsf{H}_{\mathbf{H}} h(\boldsymbol{Y}) = \left(\mathbf{P}\mathbf{E}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}} \otimes \mathbf{R}_{\boldsymbol{Z}}^{-1}\right) - 2\left(\mathbf{P} \otimes \mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{P}\right)\mathbf{N}_m \cdot \mathsf{E}\{\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y}) \otimes \boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}\left(\mathbf{P}^{\mathsf{T}} \otimes \mathbf{P}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right) \quad (34)$$

$$\mathsf{H}_{\mathbf{C}} h(\boldsymbol{Y}) = (\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{J}_{\boldsymbol{Y}}) - 2(\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}} \otimes \mathbf{I}_n)\mathbf{N}_n \cdot \mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n) \quad (35)$$

$$\mathsf{H}_{\mathbf{R}_{\boldsymbol{Z}}} h(\boldsymbol{Y}) = -\frac{1}{2}\mathbf{D}_n^{\mathsf{T}}\mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}\mathbf{D}_n \quad (36)$$

$$\mathsf{H}_{\mathbf{R}_{\boldsymbol{N}}} h(\boldsymbol{Y}) = -\frac{1}{2}\mathbf{D}_{n'}^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}} \otimes \mathbf{C}^{\mathsf{T}}) \cdot \mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}(\mathbf{C} \otimes \mathbf{C})\mathbf{D}_{n'} \quad (37)$$

where $\mathbf{N}_n$ is the symmetrization matrix defined in Appendix A.

Proof: See Appendix D. □

Remark 4: The Hessian results in Theorem 4 are given for the differential entropy. The Hessian matrices for the mutual information can be similarly derived as $\mathsf{H}_{\mathbf{P}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{H}_{\mathbf{P}} h(\boldsymbol{Y})$, $\mathsf{H}_{\mathbf{H}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{H}_{\mathbf{H}} h(\boldsymbol{Y})$, and

$$\mathsf{H}_{\mathbf{C}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{H}_{\mathbf{C}} h(\boldsymbol{Y}) - \mathbf{R}_{\boldsymbol{N}} \otimes (\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1} + 2(\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}} \otimes \mathbf{I}_n)\mathbf{N}_n((\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1} \otimes (\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1})(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n) \quad (38)$$

$$\mathsf{H}_{\mathbf{R}_{\boldsymbol{Z}}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{H}_{\mathbf{R}_{\boldsymbol{Z}}} h(\boldsymbol{Y}) + \frac{1}{2}\mathbf{D}_n^{\mathsf{T}}\left(\mathbf{R}_{\boldsymbol{Z}}^{-1} \otimes \mathbf{R}_{\boldsymbol{Z}}^{-1}\right)\mathbf{D}_n \quad (39)$$

$$\mathsf{H}_{\mathbf{R}_{\boldsymbol{N}}} I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{H}_{\mathbf{R}_{\boldsymbol{N}}} h(\boldsymbol{Y}) + \frac{1}{2}\mathbf{D}_{n'}^{\mathsf{T}}((\mathbf{C}^{\mathsf{T}}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1}\mathbf{C}) \otimes (\mathbf{C}^{\mathsf{T}}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1}\mathbf{C}))\mathbf{D}_{n'}. \quad (40)$$

D. Hessian of Mutual Information With Respect to the Transmitted Signal Covariance

In the previous sections we have purposely avoided calculating the Jacobian and Hessian matrices with respect to covariance matrices of the signal such as the squared precoder $\mathbf{Q}_{\mathbf{P}} = \mathbf{P}\mathbf{P}^{\mathsf{T}}$, the transmitted signal covariance $\mathbf{Q} = \mathbf{P}\mathbf{R}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}}$, or the input signal covariance $\mathbf{R}_{\boldsymbol{S}}$.

The reason is that, in general, the mutual information, the differential entropy, and the MMSE are not functions of $\mathbf{Q}_{\mathbf{P}}$, $\mathbf{Q}$, or $\mathbf{R}_{\boldsymbol{S}}$ alone. It can be seen, for example, by noting that, given $\mathbf{Q}_{\mathbf{P}}$, the corresponding precoder matrix $\mathbf{P}$ is specified up to an arbitrary orthonormal transformation, as both $\mathbf{P}$ and $\mathbf{P}\mathbf{V}$, with $\mathbf{V}$ being orthonormal, yield the same squared precoder $\mathbf{Q}_{\mathbf{P}}$. Now, it is easy to see that the two precoders $\mathbf{P}$ and $\mathbf{P}\mathbf{V}$ need not yield the same mutual information, and, thus, the mutual information is not well defined as a function of $\mathbf{Q}_{\mathbf{P}}$ alone because it cannot be uniquely determined from $\mathbf{Q}_{\mathbf{P}}$. The same reasoning applies to the differential entropy and the MMSE matrix.

There are, however, some particular cases where the quantities of mutual information and differential entropy are indeed functions of $\mathbf{Q}_{\mathbf{P}}$, $\mathbf{Q}$, or $\mathbf{R}_{\boldsymbol{S}}$. We have, for example, the particular case where the signaling is Gaussian, $\boldsymbol{S} = \boldsymbol{S}_{\mathcal{G}}$. In this case, the mutual information is given by

$$I(\boldsymbol{S}_{\mathcal{G}};\boldsymbol{Y}_{\mathcal{G}}) = \frac{1}{2}\mathsf{logdet}\left(\mathbf{I}_n + \mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{P}\mathbf{R}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\right) \quad (41)$$

which is, of course, a function of the transmitted signal covariance $\mathbf{Q} = \mathbf{P}\mathbf{R}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}}$, a function of the input signal covariance $\mathbf{R}_{\boldsymbol{S}}$, and also a function of the squared precoder $\mathbf{Q}_{\mathbf{P}} = \mathbf{P}\mathbf{P}^{\mathsf{T}}$ when $\mathbf{R}_{\boldsymbol{S}} = \mathbf{I}_m$. Upon direct double differentiation with respect to, e.g., $\mathbf{Q}$ we obtain [15, Ch. 9 and 10]

$$\mathsf{H}_{\mathbf{Q}} I(\boldsymbol{S}_{\mathcal{G}};\boldsymbol{Y}_{\mathcal{G}}) = \frac{1}{2}\mathbf{D}_p^{\mathsf{T}}\left(\left(\left(\mathbf{I}_p + \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{Q}\right)^{-1}\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\right) \otimes \left(\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\left(\mathbf{I}_p + \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{Q}\right)^{-1}\right)\right)\mathbf{D}_p. \quad (42)$$

Another particular case where the mutual information is a function of the transmit covariance matrices is in the low-SNR regime [16]. Assuming that $\mathbf{R}_{\boldsymbol{Z}} = N_0\mathbf{I}$, Prelov and Verdú showed that [16, Theorem 3]

$$I(\boldsymbol{S};\boldsymbol{Y}) = \frac{1}{2N_0}\mathsf{Tr}\left(\mathbf{H}\mathbf{P}\mathbf{R}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\right) - \frac{1}{4N_0^2}\mathsf{Tr}\left(\left(\mathbf{H}\mathbf{P}\mathbf{R}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}}\mathbf{H}^{\mathsf{T}}\right)^2\right) + o\left(N_0^{-2}\right) \quad (43)$$

where the dependence of the mutual information with respect to $\mathbf{Q}$ is explicitly shown. The Hessian of the mutual information, for this case becomes [15, Ch. 9 and 10]

$$\mathsf{H}_{\mathbf{Q}} I(\boldsymbol{S};\boldsymbol{Y}) = -\frac{1}{2N_0^2}\mathbf{D}_p^{\mathsf{T}}\left(\mathbf{H}^{\mathsf{T}}\mathbf{H} \otimes \mathbf{H}^{\mathsf{T}}\mathbf{H}\right)\mathbf{D}_p + o\left(N_0^{-2}\right). \quad (44)$$

Even though we have shown two particular cases where the mutual information is a function of the transmitted signal covariance matrix $\mathbf{Q}$, it is important to highlight that care must be taken when calculating the Jacobian matrix of the MMSE and the Hessian matrix of the mutual information or the differential entropy as, in general, these quantities are *not* functions of $\mathbf{Q}_{\mathbf{P}}$, $\mathbf{Q}$, nor $\mathbf{R}_{\boldsymbol{S}}$. In this sense, the results in [1, Theorem 2, eq. (23), (24), (25); Cor. 2, eq. (49); Theorem 4, eq. (56)] only make sense when the mutual information is well defined as a function of the signal covariance matrix (such as when the signaling is Gaussian or the SNR is low).

IV. Mutual Information Concavity Results

As we have mentioned in the introduction, studying the concavity of the mutual information with respect to design parameters of the system is important from both analysis and design perspectives.

The first candidate as a system parameter of interest that naturally arises is the precoder matrix $\mathbf{P}$ in the signal model $\boldsymbol{Y} = \mathbf{H}\mathbf{P}\boldsymbol{S} + \boldsymbol{Z}$. However, one realizes from the expression $\mathsf{H}_{\mathbf{P}} I(\boldsymbol{S};\boldsymbol{Y})$ in Remark 4 of Theorem 4, that for a sufficiently small $\mathbf{P}$ the Hessian is approximately $\mathsf{H}_{\mathbf{P}} I(\boldsymbol{S};\boldsymbol{Y}) \approx \mathbf{E}_{\boldsymbol{S}} \otimes \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}$, which, from Lemma H.3 is positive definite and, consequently, the mutual information is not concave in $\mathbf{P}$ (actually, it is convex). Numerical computations show that the nonconcavity of the mutual information with respect to $\mathbf{P}$ also holds for nonsmall $\mathbf{P}$.

The next candidate is the transmitted signal covariance matrix $\mathbf{Q}$, which, at first sight, is better suited than the precoder $\mathbf{P}$ as it is well known that, for the Gaussian signaling case, the mutual information as in (41) is a concave function of the transmitted signal covariance $\mathbf{Q}$. Similarly, in the low SNR regime we have that, from (44), the mutual information is also a concave function with respect to $\mathbf{Q}$.

Since in this work we are interested in the properties of the mutual information for all the SNR range and for arbitrary signaling, we wish to study if the above results can be generalized. Unfortunately, as discussed in the previous section, the first difference of the general case with respect to the particular cases of Gaussian signaling and low SNR is that the mutual information is not well defined as a function of the transmitted signal covariance $\mathbf{Q}$ only.

Having discarded the concavity of the mutual information with respect to $\mathbf{P}$ and $\mathbf{Q}$, in the following subsections we study the concavity of the mutual information with respect to other parameters of the system.

For the sake of notation we define the channel covariance matrix as $\mathbf{R}_{\mathbf{H}} = \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}$, which will be used in the remainder of the paper.

A. Concavity in the SNR

The concavity of the mutual information with respect to the SNR for arbitrary input distributions can be derived as a corollary from Costa's results in [10], where he proved the concavity of the entropy power of a random variable consisting of the sum of a signal and Gaussian noise with respect to the power of the signal. As a direct consequence, the concavity of the entropy power implies the concavity of the mutual information in the signal power, or, equivalently, in the SNR.

In this section, we give an explicit expression of the Hessian of the mutual information with respect to the SNR, which was previously unavailable for vector Gaussian channels.

Corollary 1 (Mutual Information Hessian With Respect to the SNR): Consider the model $\boldsymbol{Y} = \sqrt{\mathsf{snr}}\mathbf{H}\boldsymbol{S}+\boldsymbol{Z}$, with $\mathsf{snr} > 0$ and where all the terms are defined as in Theorem 3. It then follows that the mutual information is a concave function with respect to snr, $\mathsf{H}_{\mathsf{snr}} I(\boldsymbol{S};\boldsymbol{Y}) \leq 0$.

Moreover, we have that

$$\mathsf{H}_{\mathsf{snr}} I(\boldsymbol{S};\boldsymbol{Y}) = -\frac{1}{2}\mathsf{Tr}\mathsf{E}\{(\mathbf{R}_{\mathbf{H}}\mathbf{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y}))^2\}. \tag{45}$$

Proof: See Appendix G. □

Remark 5: Observe that (45) agrees with [9, Prop. 5] for scalar Gaussian channels.

In the following section, we extend the concavity result in Corollary 1 to more general quantities than the scalar SNR.

B. Concavity in the Squared Singular Values of the Precoder When the Precoder Diagonalizes the Channel

Consider the eigendecomposition of the $p \times p$ channel covariance matrix $\mathbf{R}_{\mathbf{H}} = \mathbf{U}_{\mathbf{H}}\mathbf{Diag}(\boldsymbol{\sigma})\mathbf{U}_{\mathbf{H}}^{\mathsf{T}}$, where $\mathbf{U}_{\mathbf{H}} \in \mathbb{R}^{p\times p}$ is an orthonormal matrix and the vector $\boldsymbol{\sigma} \in \mathbb{R}^{p}$ contains nonnegative entries in no particular order. Note that, in the case where $\mathsf{rank}(\mathbf{R}_{\mathbf{H}}) = p' < p$, then $p - \mathsf{rank}(\mathbf{R}_{\mathbf{H}})$ elements of vector $\boldsymbol{\sigma}$ are zero.

Let us now consider the singular value decomposition (SVD) of the $p \times m$ precoder matrix $\mathbf{P} = \mathbf{U}_{\mathbf{P}}\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}$. Defining $m' = \min\{p, m\}$, we have that the vector $\boldsymbol{\lambda}$ is m'-dimensional, and the matrices $\mathbf{U}_{\mathbf{P}} \in \mathbb{R}^{p\times m'}$ and $\mathbf{V}_{\mathbf{P}} \in \mathbb{R}^{m\times m'}$ contain orthonormal columns such that $\mathbf{U}_{\mathbf{P}}^{\mathsf{T}}\mathbf{U}_{\mathbf{P}} = \mathbf{I}_{m'}$ and $\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\mathbf{V}_{\mathbf{P}} = \mathbf{I}_{m'}$, respectively.

In the following theorem we characterize the concavity properties of the mutual information with respect to the entries of the squared singular values vector $\boldsymbol{\lambda}$ for the particular case where the left singular vectors of the precoder coincide with the first m' eigenvectors of the channel covariance matrix.

Theorem 5 (Mutual Information Hessian With Respect to the Squared Singular Values of the Precoder): Consider $\boldsymbol{Y} = \mathbf{H}\mathbf{P}\boldsymbol{S} + \boldsymbol{Z}$, where all the terms are defined as in Theorem 3, for the particular case where the first m' eigenvectors of the channel covariance matrix $\mathbf{R}_{\mathbf{H}}$ and the left singular vectors of the precoder $\mathbf{P} \in \mathbb{R}^{p\times m}$ coincide. It then follows that the mutual information is a concave function of the squared singular values of the precoder $\boldsymbol{\lambda}$, $\mathsf{H}_{\boldsymbol{\lambda}} I(\boldsymbol{S};\boldsymbol{Y}) \leq \mathbf{0}$.

Moreover, the Hessian of the mutual information with respect to $\boldsymbol{\lambda}$ is

$$\mathsf{H}_{\boldsymbol{\lambda}} I(\boldsymbol{S};\boldsymbol{Y}) = -\frac{1}{2}\mathbf{Diag}(\tilde{\boldsymbol{\sigma}})\mathsf{E}\left\{\mathbf{\Phi}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{Y}) \circ \mathbf{\Phi}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{Y})\right\}\mathbf{Diag}(\tilde{\boldsymbol{\sigma}}) \tag{46}$$

where we have defined $\tilde{\boldsymbol{\sigma}} = (\sigma_1\sigma_2\ldots\sigma_{m'})^{\mathsf{T}}$.

Proof: See Appendix G. □

Remark 6: Observe from the expression for the Hessian in (46) that for the case where the channel covariance matrix $\mathbf{R}_{\mathbf{H}}$ is rank deficient, $\mathsf{rank}(\mathbf{R}_{\mathbf{H}}) = p'$, then there may be some elements of vector $\tilde{\boldsymbol{\sigma}}$ that are zero. In this case, the corresponding rows and columns of the Hessian matrix in (46) are also zero.

We now generalize a result obtained in [17] for parallel channels were it was proved that the mutual information is concave in the power allocation for the case where the entries of the signaling vector $\boldsymbol{S}$ are assumed independent (this last assumption is actually unnecessary as shown next).

TABLE I
SUMMARY OF THE CONCAVITY TYPE OF THE MUTUAL INFORMATION
(✓ INDICATES CONCAVITY, × INDICATES NON-CONCAVITY, AND − INDICATES THAT IT DOES NOT APPLY)

Cases	Scalar Power, snr, $\mathbf{P} = \sqrt{\mathsf{snr}}\mathbf{I}$	Vector Squared singular values, $\boldsymbol{\lambda}$, $\mathbf{P} = \mathbf{U_P}\mathbf{Diag}\,(\sqrt{\boldsymbol{\lambda}})\mathbf{V_P^{\mathsf{T}}}$	Matrix Transmit covariance, $\mathbf{Q}$, $\mathbf{Q} = \mathbf{P}\mathbf{R}_{\boldsymbol{S}}\mathbf{P}^{\mathsf{T}}$
General case: $\boldsymbol{Y} = \mathbf{H}\mathbf{P}\boldsymbol{S} + \boldsymbol{Z}$	✓ [10] (Section IV-A)	× (Section IV-C)	− (Section III-D)
Channel covariance $\mathbf{R_H}$ and precoder $\mathbf{P}$ share singular/eigenvectors.	✓	✓ (Section IV-B)	− (Section III-D)
Independent parallel communication: $\mathbf{R_H} = \mathbf{U_P} = \mathbf{V_P} = \mathbf{I}$, $P_{\boldsymbol{S}} = \prod_i P_{S_i}$	✓	✓ [17]	✓ [17] (Note that $\mathbf{Q}$ is diagonal)
Low SNR regime: $\mathbf{R}_{\boldsymbol{Z}} = N_0\mathbf{I}_n$, $N_0 \gg 1$	✓	✓	✓ [16]
Gaussian signaling: $\boldsymbol{S} = \boldsymbol{S}_{\mathcal{G}}$	✓	✓	✓

Corollary 2 (Mutual Information Concavity With Respect to the Power Allocation in Parallel Channels): Particularizing Theorem 5 for the case where the channel $\mathbf{H}$, the precoder $\mathbf{P}$, and the noise covariance $\mathbf{R}_{\boldsymbol{Z}}$ are diagonal matrices, which implies that $\mathbf{U_P} = \mathbf{U_H} = \mathbf{I}_p$, it follows that the mutual information is a concave function with respect to the power allocation for parallel noninteracting channels for an arbitrary distribution of the signaling vector $\boldsymbol{S}$.

C. General Negative Results

In the previous section we have proved that the mutual information is a concave function of the squared singular values of the precoder matrix $\mathbf{P}$ for the case where the left singular vectors of the precoder $\mathbf{P}$ coincide with the eigenvectors of the channel correlation matrix, $\mathbf{R_H}$. For the general case where these vectors do not coincide, the mutual information is not a concave function of the squared singular values of the precoder. This fact is formally established through the following counterexample.

Counterexample 1 (General Nonconcavity of the Mutual Information): Consider $\boldsymbol{Y} = \mathbf{H}\mathbf{P}\boldsymbol{S} + \boldsymbol{Z}$, where all the terms are defined as in Theorem 3. It then follows that, in general, the mutual information is not a concave function with respect to the squared singular values of the precoder $\boldsymbol{\lambda}$.

Proof: See Appendix G. □

D. Concavity Results Summary

A summary of the different concavity results (positive and negative) for the mutual information as a function of the configuration of the linear vector Gaussian channel can be found in Table I.

V. APPLICATIONS

A. Multivariate Extension of Costa's Entropy Power Inequality

The entropy power of the random vector $\boldsymbol{Y} \in \mathbb{R}^n$ was first introduced by Shannon in his seminal work [18] and, since then, is defined as $N(\boldsymbol{Y}) = \exp(2h(\boldsymbol{Y})/n)/(2\pi\mathrm{e})$.

Costa proved in [10] that, provided that the random vector $\boldsymbol{Z}$ is white Gaussian distributed, then

$$N(\boldsymbol{X} + \sqrt{t}\boldsymbol{Z}) \geq (1-t)N(\boldsymbol{X}) + tN(\boldsymbol{X}+\boldsymbol{Z}) \quad (47)$$

for any $t \in [0,1]$. As Costa noted, the above entropy power inequality (EPI) is equivalent to the concavity of the entropy power function $N(\boldsymbol{X} + \sqrt{t}\boldsymbol{Z})$ with respect to the parameter t, or, formally, to $\mathsf{H}_t N(\boldsymbol{X} + \sqrt{t}\boldsymbol{Z}) \leq 0$.[3] Additionally, in his paper Costa showed that the EPI is also valid when the Gaussian vector $\boldsymbol{Z}$ is not white.

Due to its inherent interest and to the fact that the proof by Costa was rather involved, simplified proofs of his result have been subsequently given in [19]–[22]. Moreover, in [22] Rioul proved a version of Costa's EPI where the t parameter is multiplying the arbitrarily distributed random vector $\boldsymbol{X}$ (instead of $\boldsymbol{Z}$):

$$\mathsf{H}_t N(\sqrt{t}\boldsymbol{X} + \boldsymbol{Z}) \leq 0. \quad (48)$$

Observe that $\sqrt{t}$ in (48) plays the role of a scalar precoder. We next consider an extension of (48) to the case where the scalar precoder $\sqrt{t}$ is replaced by a multivariate precoder $\mathbf{P} \in \mathbb{R}^{p\times m}$ and a channel $\mathbf{H} \in \mathbb{R}^{n\times p}$ for the particular case where the precoder left singular vectors coincide with the first m' channel covariance eigenvectors.

Theorem 6 (Costa's Multivariate EPI): Consider $\boldsymbol{Y} = \mathbf{H}\mathbf{P}\boldsymbol{S} + \boldsymbol{Z}$, where all the terms are defined as in Theorem 3, for the particular case where the first m' eigenvectors of the channel covariance matrix $\mathbf{R_H}$ and the left singular vectors of the precoder $\mathbf{P} \in \mathbb{R}^{p\times m}$ coincide. It then follows that the entropy power is a concave function of $\boldsymbol{\lambda}$, $\mathsf{H}_{\boldsymbol{\lambda}} N(\boldsymbol{Y}) \leq \mathbf{0}$.

Moreover, the Hessian matrix of the entropy power is

$$\mathsf{H}_{\boldsymbol{\lambda}} N(\boldsymbol{Y}) = \frac{N(\boldsymbol{Y})}{n}\mathbf{Diag}(\tilde{\boldsymbol{\sigma}})\left(\frac{\mathbf{diag}\left(\mathbf{E}_{\mathbf{V_P^{\mathsf{T}}}S}\right)\mathbf{diag}\left(\mathbf{E}_{\mathbf{V_P^{\mathsf{T}}}S}\right)^{\mathsf{T}}}{n} - \mathsf{E}\left\{\boldsymbol{\Phi}_{\mathbf{V_P^{\mathsf{T}}}S}(\boldsymbol{Y}) \circ \boldsymbol{\Phi}_{\mathbf{V_P^{\mathsf{T}}}S}(\boldsymbol{Y})\right\}\right)\mathbf{Diag}(\tilde{\boldsymbol{\sigma}}) \quad (49)$$

where we recall that $\mathbf{diag}(\mathbf{E}_{\mathbf{V_P^{\mathsf{T}}}S})$ represents a column vector with the diagonal entries of the matrix $\mathbf{E}_{\mathbf{V_P^{\mathsf{T}}}S}$ and that $\tilde{\boldsymbol{\sigma}} = (\sigma_1\sigma_2\ldots\sigma_{m'})^{\mathsf{T}}$.

[3]The equivalence between (47) and $\mathsf{H}_t N(\boldsymbol{X} + \sqrt{t}\boldsymbol{Z}) \leq 0$ is due to the fact that the function $N(\boldsymbol{X}+\sqrt{t}\boldsymbol{Z})$ is twice differentiable almost everywhere thanks to the smoothing properties of the added Gaussian noise.

Proof: See Appendix G. □

Remark 7: For the case where $\mathbf{R_H} = \mathbf{I}_p$ and $p = m$ we recover our earlier result in [23].

Another possibility of multivariate generalization of Costa's EPI would be to study the concavity of $N(\boldsymbol{X}+\boldsymbol{Z})$ with respect to the covariance of the noise vector $\mathbf{R}_{\boldsymbol{Z}}$. This seems to be more elusive and has not been further elaborated herein.

B. Precoder Design

The concavity results presented in Theorem 5 can be used to numerically compute the optimal squared singular values $\boldsymbol{\lambda}^\star$ of the precoder $\mathbf{P} = \mathbf{U_P}\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{V_P}^\mathsf{T}$ that, under an average transmitted power constraint, maximizes the mutual information $I(\boldsymbol{S};\mathbf{HP}\boldsymbol{S}+\boldsymbol{Z})$ assuming that the right eigenvector matrix $\mathbf{V_P}$ is given and held fixed and that the optimal left eigenvector matrix of the precoder is used, i.e., $\mathbf{U_P} = \mathbf{U_H}$ where $\mathbf{U_H}$ contains the eigenvectors of the covariance matrix of the channel $\mathbf{R_H}$.[4] The details of the optimization algorithm are outside the scope of the present paper and are, thus, omitted.

APPENDIX

A. Special Matrices Used in Multivariate Differentiation

In this Appendix, we present four matrices that are often encountered when calculating Hessian matrices. The definitions of the commutation $\mathbf{K}_{q,r}$, symmetrization $\mathbf{N}_q$, and duplication $\mathbf{D}_q$ matrices have been taken from [15] and the reduction matrix $\mathbf{S}_q$ has been defined by the authors of the present work.

Given any matrix $\mathbf{A} \in \mathbb{R}^{q\times r}$, there exists a unique permutation matrix $\mathbf{K}_{q,r} \in \mathbb{R}^{qr\times qr}$ independent of $\mathbf{A}$, which is called commutation matrix, that satisfies

$$\mathsf{vec}\mathbf{A}^\mathsf{T} = \mathbf{K}_{q,r}\mathsf{vec}\mathbf{A}, \quad \text{and} \quad \mathbf{K}_{q,r}^\mathsf{T} = \mathbf{K}_{q,r}^{-1} = \mathbf{K}_{r,q}. \quad (50)$$

Thus, the entries of the commutation matrix are given by

$$[\mathbf{K}_{q,r}]_{i+(j-1)r,i'+(j'-1)q} = \delta_{i'j}\delta_{j'i} \qquad \{i',j\}\in[1,q], \quad \{i,j'\}\in[1,r]. \quad (51)$$

The main reason why we have introduced the commutation matrix is due to the property from which it obtains its name, as it enables us to commute the two matrices of a Kronecker product [15, Ch. 3, Theorem 9],

$$\mathbf{K}_{s,q}(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{B}\otimes\mathbf{A})\mathbf{K}_{t,r} \quad (52)$$

where we have considered $\mathbf{A}\in\mathbb{R}^{q\times r}$ and $\mathbf{B}\in\mathbb{R}^{s\times t}$.

We also define $\mathbf{K}_q = \mathbf{K}_{q,q}$ for the square case. An important property of the square matrix $\mathbf{K}_q$ is given next.

Lemma A.1: Let $\mathbf{A}\in\mathbb{R}^{q\times r}$ and $\mathbf{B}\in\mathbb{R}^{q\times t}$. Then,

$$[\mathbf{A}\otimes\mathbf{B}]_{i+(j-1)q,k+(l-1)t} = \mathbf{A}_{jl}\mathbf{B}_{ik} \quad (53)$$
$$[\mathbf{K}_q(\mathbf{A}\otimes\mathbf{B})]_{i+(j-1)q,k+(l-1)t} = \mathbf{A}_{il}\mathbf{B}_{jk} \quad (54)$$

with $\{i,j\}\in[1,q]$, $k\in[1,t]$, and $l\in[1,r]$.

[4]The optimality of $\mathbf{U_P} = \mathbf{U_H}$ when optimizing the mutual information follows from a similar derivation as in [24, Appendix A].

Proof: Both equalities follow straightforwardly from the definition in [25, Sec. 4.2]. In the calculation of the entries of $\mathbf{K}_q(\mathbf{A}\otimes\mathbf{B})$, the expression for the elements of $\mathbf{K}_q$ in (51) has to be used. □

When calculating Jacobian and Hessian matrices, the form $\mathbf{I}_q + \mathbf{K}_q$ is usually encountered. Hence, we define the symmetrization matrix $\mathbf{N}_q = \frac{1}{2}(\mathbf{I}_q+\mathbf{K}_q)$, which is singular and fulfills $\mathbf{N}_q = \mathbf{N}_q^\mathsf{T} = \mathbf{N}_q^2$ and $\mathbf{N}_q = \mathbf{N}_q\mathbf{K}_q = \mathbf{K}_q\mathbf{N}_q$. The name of the symmetrization matrix comes from the fact that given any square matrix $\mathbf{A}\in\mathbb{R}^{q\times q}$, then

$$\mathbf{N}_q\mathsf{vec}\mathbf{A} = \mathsf{vec}\left(\frac{\mathbf{A}+\mathbf{A}^\mathsf{T}}{2}\right). \quad (55)$$

The last important property of the symmetrization matrix is

$$\mathbf{N}_q(\mathbf{A}\otimes\mathbf{A}) = (\mathbf{A}\otimes\mathbf{A})\mathbf{N}_q \quad (56)$$

which follows from the definition of $\mathbf{N}_q$ together with (52).

Another important matrix related to the calculation of Jacobian and Hessian matrices, specially when symmetric matrices are involved, is the duplication matrix $\mathbf{D}_q$. From [15, Sec. 3.8], the duplication matrix $\mathbf{D}_q\in\mathbb{R}^{q^2\times q(q+1)/2}$ fulfills $\mathsf{vec}\mathbf{R} = \mathbf{D}_q\mathsf{vech}\mathbf{R}$, for any q-dimensional symmetric matrix $\mathbf{R}$. The duplication matrix takes its name from the fact that it duplicates the entries of $\mathsf{vech}\mathbf{R}$ which correspond to off-diagonal elements of $\mathbf{R}$ to produce the elements of $\mathsf{vec}\mathbf{R}$.

Since $\mathbf{D}_q$ has full column rank, it is possible to invert the transformation $\mathsf{vec}\mathbf{R} = \mathbf{D}_q\mathsf{vech}\mathbf{R}$ to obtain

$$\mathsf{vech}\mathbf{R} = \mathbf{D}_q^+\mathsf{vec}\mathbf{R} = \left(\mathbf{D}_q^\mathsf{T}\mathbf{D}_q\right)^{-1}\mathbf{D}_q^\mathsf{T}\mathsf{vec}\mathbf{R}. \quad (57)$$

The most important properties of the duplication matrix are [15, Ch. 3, Theorem 12]

$$\mathbf{K}_q\mathbf{D}_q = \mathbf{D}_q, \mathbf{N}_q\mathbf{D}_q = \mathbf{D}_q, \mathbf{D}_q\mathbf{D}_q^+ = \mathbf{N}_q, \mathbf{D}_q^+\mathbf{N}_q = \mathbf{D}_q^+. \quad (58)$$

The last one of the matrices introduced in this Appendix is the reduction matrix $\mathbf{S}_q\in\mathbb{R}^{q^2\times q}$. The entries of the reduction matrix are defined as

$$[\mathbf{S}_q]_{i+(j-1)q,k} = \delta_{ijk}, \quad \{i,j,k\}\in[1,q] \quad (59)$$

from which it is easy to verify that the reduction matrix fulfills $\mathbf{K}_q\mathbf{S}_q = \mathbf{S}_q$ and $\mathbf{N}_q\mathbf{S}_q = \mathbf{S}_q$. However, the most important property of the reduction matrix is that it can be used to reduce the Kronecker product of two matrices to their Schur product as it is detailed in the next lemma.

Lemma A.2: Let $\mathbf{A}\in\mathbb{R}^{q\times r}$, $\mathbf{B}\in\mathbb{R}^{q\times r}$. Then

$$\mathbf{S}_q^\mathsf{T}(\mathbf{A}\otimes\mathbf{B})\mathbf{S}_r = \mathbf{S}_q^\mathsf{T}(\mathbf{B}\otimes\mathbf{A})\mathbf{S}_r = \mathbf{A}\circ\mathbf{B}. \quad (60)$$

Proof: The proof follows easily from the expression for the elements of the Kronecker product in Lemma A.1 and the expression for the elements of the reduction matrix in (59). □

Finally, we present two basic lemmas concerning the Kronecker product and the vec operator.

Lemma A.4: Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{T}$ be four matrices such that the products $\mathbf{AB}$ and $\mathbf{FT}$ are defined. Then, $(\mathbf{A}\otimes\mathbf{F})(\mathbf{B}\otimes\mathbf{T}) = \mathbf{AB}\otimes\mathbf{FT}$.

Proof: See [15, Sec. 2.2]. □

Lemma A.4: Let $\mathbf{A}$, $\mathbf{T}$, and $\mathbf{B}$ be three matrices such that the product $\mathbf{ATB}$ is defined. Then,

$$\mathsf{vec}(\mathbf{ATB}) = \left(\mathbf{B}^{\mathsf{T}}\otimes\mathbf{A}\right)\mathsf{vec}\mathbf{T}. \tag{61}$$

Proof: See [15, Ch. 2, Theorem 2.2]. □

B. Conventions Used for Jacobian and Hessian Matrices

In this work we make extensive use of differentiation of matrix functions $\mathbf{\Psi}$ with respect to a matrix argument $\mathbf{T}$. From the many possibilities of displaying the partial derivatives $\partial^r\mathbf{\Psi}_{st}/\partial\mathbf{T}_{ij}\cdots\partial\mathbf{T}_{kl}$, we will stick to the "good notation" introduced by Magnus and Neudecker in [15, Sec. 9.4]] which is briefly reproduced next for the sake of completeness.

Definition B.1: Let $\mathbf{\Psi}$ be a differentiable $q\times t$ real matrix function of an $r\times s$ matrix of real variables $\mathbf{T}$. The Jacobian matrix of $\mathbf{\Psi}$ at $\mathbf{T}=\mathbf{T}_0$ is the $qt\times rs$ matrix

$$\mathsf{D}_{\mathbf{T}}\mathbf{\Psi}(\mathbf{T}_0) = \left.\frac{\partial\,\mathsf{vec}\,\mathbf{\Psi}(\mathbf{T})}{\partial\mathsf{vec}^{\mathsf{T}}\mathbf{T}}\right|_{\mathbf{T}=\mathbf{T}_0}. \tag{62}$$

Remark B.2: To properly deal with the case where $\mathbf{\Psi}$ is a symmetric matrix, the vec operator in the numerator in (62) has to be replaced by a vech operator to avoid obtaining repeated elements. Similarly, vech has to replace vec in the denominator in (62) for the case where $\mathbf{T}$ is a symmetric matrix. For practical purposes, it is enough to calculate the Jacobian without taking into account any symmetry properties and then add a left factor $\mathbf{D}_q^+$ to the obtained Jacobian when $\mathbf{\Psi}$ is symmetric and/or a right factor $\mathbf{D}_r$ when $\mathbf{T}$ is symmetric. This proceeding will become clearer in the examples below.

Definition B.3: Let $\mathbf{\Psi}$ be a twice differentiable $q\times t$ real matrix function of an $r\times s$ matrix of real variables $\mathbf{T}$. The Hessian matrix of $\mathbf{\Psi}$ at $\mathbf{T}=\mathbf{T}_0$ is the $qtrs\times rs$ matrix

$$\mathsf{H}_{\mathbf{T}}\mathbf{\Psi}(\mathbf{T}_0) = \mathsf{D}_{\mathbf{T}}\left(\mathsf{D}_{\mathbf{T}}^{\mathsf{T}}\mathbf{\Psi}(\mathbf{T})\right)\Big|_{\mathbf{T}=\mathbf{T}_0} \tag{63}$$

$$= \left.\frac{\partial}{\partial\mathsf{vec}^{\mathsf{T}}\mathbf{T}}\mathsf{vec}\left(\frac{\partial\,\mathsf{vec}\,\mathbf{\Psi}(\mathbf{T})}{\partial\mathsf{vec}^{\mathsf{T}}\mathbf{T}}\right)^{\mathsf{T}}\right|_{\mathbf{T}=\mathbf{T}_0}. \tag{64}$$

One can verify that the obtained Hessian matrix for the matrix function $\mathbf{\Psi}$ is the stacking of the qt Hessian matrices corresponding to each individual element of vector $\mathsf{vec}\mathbf{\Psi}$.

Remark B.4: Similarly to the Jacobian case, when $\mathbf{\Psi}$ or $\mathbf{T}$ are symmetric matrices, the vech operator has to replace the vec operator where appropriate in (64).

One of the major advantages of using the notation in [15] is that a simple chain rule can be applied for both the Jacobian and Hessian matrices, as detailed in the following lemma.

Lemma B.5 ([15, Ch. 5, Theorem 8 and Ch. 6, Theorem 9]): Let $\mathbf{\Upsilon}$ be a twice differentiable $u\times v$ real matrix function of a $q\times t$ real matrix argument. Let $\mathbf{\Psi}$ be a twice differentiable $q\times t$ real matrix function of an $r\times s$ matrix of real variables $\mathbf{T}$. Define $\mathbf{\Omega}(\mathbf{T}) = \mathbf{\Upsilon}(\mathbf{\Psi}(\mathbf{T}))$. Using $\mathbf{\Psi}_0 = \mathbf{\Psi}(\mathbf{T}_0)$, the Jacobian and Hessian matrices of $\mathbf{\Omega}(\mathbf{T})$ at $\mathbf{T}=\mathbf{T}_0$ are

$$\mathsf{D}_{\mathbf{T}}\mathbf{\Omega}(\mathbf{T}_0) = (\mathsf{D}_{\mathbf{\Psi}}\mathbf{\Upsilon}(\mathbf{\Psi}_0))(\mathsf{D}_{\mathbf{T}}\mathbf{\Psi}(\mathbf{T}_0)) \tag{65}$$

$$\mathsf{H}_{\mathbf{T}}\mathbf{\Omega}(\mathbf{T}_0) = (\mathbf{I}_{uv}\otimes\mathsf{D}_{\mathbf{T}}\mathbf{\Psi}(\mathbf{T}_0))^{\mathsf{T}}\mathsf{H}_{\mathbf{\Psi}}\mathbf{\Upsilon}(\mathbf{\Psi}_0)(\mathsf{D}_{\mathbf{T}}\mathbf{\Psi}(\mathbf{T}_0)) + (\mathsf{D}_{\mathbf{\Psi}}\mathbf{\Upsilon}(\mathbf{\Psi}_0)\otimes\mathbf{I}_{rs})\mathsf{H}_{\mathbf{T}}\mathbf{\Psi}(\mathbf{T}_0). \tag{66}$$

The notation introduced above unifies the study of scalar ($q = t = 1$), vector ($t = 1$), and matrix functions $\mathbf{\Psi}$ of scalar ($r = s = 1$), vector ($s = 1$), or matrix arguments $\mathbf{T}$ into the study of vector functions of vector arguments through the use of the vec and vech operators. However, the idea of arranging the partial derivatives of a scalar function of a matrix argument $\psi(\mathbf{T})$ into a matrix rather than a vector is quite appealing and sometimes useful, so we will also make use of the notation described next.

Definition B.6: Let ψ be differentiable scalar function of an $r\times s$ matrix of real variables $\mathbf{T}$. The gradient of ψ at $\mathbf{T}=\mathbf{T}_0$ is the $r\times s$ matrix

$$\nabla_{\mathbf{T}}\psi(\mathbf{T}_0) = \left.\frac{\partial\psi}{\partial\mathbf{T}}\right|_{\mathbf{T}=\mathbf{T}_0}. \tag{67}$$

It is easy to verify that $\mathsf{D}_{\mathbf{T}}\psi(\mathbf{T}_0) = \mathsf{vec}^{\mathsf{T}}\nabla_{\mathbf{T}}\psi(\mathbf{T}_0)$.

We now give expressions for the most common Jacobian and Hessian matrices encountered during our developments.

Lemma B.7: Consider $\mathbf{A}\in\mathbb{R}^{q\times r}$, $\mathbf{T}\in\mathbb{R}^{r\times s}$, $\mathbf{B}\in\mathbb{R}^{s\times t}$, $\mathbf{R}\in\mathbb{S}_+^s$, and $\mathbf{f}\in\mathbb{R}^{r\times 1}$, such that $\mathbf{f}$ is a function of $\mathbf{T}$. Then, the following holds:

1) If $\mathbf{\Psi} = \mathbf{ATB}$, then $\mathsf{D}_{\mathbf{T}}\mathbf{\Psi} = (\mathbf{B}^{\mathsf{T}}\otimes\mathbf{A})$. If, in addition, $\mathbf{B}$ is a function of $\mathbf{T}$, then we have $\mathsf{D}_{\mathbf{T}}\mathbf{\Psi} = (\mathbf{B}^{\mathsf{T}}\otimes\mathbf{A}) + (\mathbf{I}_t\otimes\mathbf{AT})\mathsf{D}_{\mathbf{T}}\mathbf{B}$.
2) If $\mathbf{\Psi} = \mathbf{Af}$, then $\mathsf{D}_{\mathbf{T}}\mathbf{\Psi} = \mathbf{A}\mathsf{D}_{\mathbf{T}}\mathbf{f}$.
3) If $\mathbf{\Psi} = \mathbf{ATA}^{\mathsf{T}}$, with $\mathbf{T}$ being a symmetric matrix, then $\mathsf{D}_{\mathbf{T}}\mathbf{\Psi} = \mathbf{D}_q^+(\mathbf{A}\otimes\mathbf{A})\mathbf{D}_r$.
4) If $\mathbf{\Psi} = \mathbf{T}^{-1}$, then $\mathsf{D}_{\mathbf{T}}\mathbf{\Psi} = -(\mathbf{T}^{\mathsf{T}}\otimes\mathbf{T})^{-1}$, where $\mathbf{T}$ is a square invertible matrix.
5) If $\mathbf{\Psi} = \mathbf{ATRT}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}$, then $\mathsf{D}_{\mathbf{T}}\mathbf{\Psi} = 2\mathbf{D}_q^+(\mathbf{ATR}\otimes\mathbf{A})$.

Proof: See [15, Ch. 9]. □

C. Differential Properties of $P_{\boldsymbol{Y}}(\boldsymbol{y})$, $P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}\,|\,\boldsymbol{s})$, and $\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\}$

In this Appendix we present the lemmas which are used in Appendixes E and F for the proofs of Theorems 1 and 2.

In the proofs of the following lemmas, we interchange the order of differentiation and expectation, which can be justified following similar steps as in [1, Appendix B], where it was assumed that the signaling had finite second-order moments.[5]

Lemma C.1: Let $\boldsymbol{Y} = \boldsymbol{X} + \mathbf{C}\boldsymbol{N}$, where $\boldsymbol{X}$ is arbitrarily distributed and $\boldsymbol{N}$ is a zero-mean Gaussian random variable with covariance matrix $\mathbf{R}_{\boldsymbol{N}}$ and independent of $\boldsymbol{X}$. Then, the probability density function $P_{\boldsymbol{Y}}(\boldsymbol{y})$ satisfies

$$\nabla_{\mathbf{C}}P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathsf{H}_{\boldsymbol{y}}P_{\boldsymbol{Y}}(\boldsymbol{y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}}. \tag{68}$$

Proof: First, we recall that $P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathsf{E}\{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})\}$. Thus, $\nabla_{\mathbf{C}}P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathsf{E}\{\nabla_{\mathbf{C}}P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})\}$. The computation of

[5]From recent results in [8] it is apparent that the finiteness assumption on the second order moments can be dropped.

the inner the gradient $\nabla_{\mathbf{C}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})$ can be performed by replacing $\mathbf{G}\boldsymbol{s}$ by $\boldsymbol{x}$ and $\mathbf{R}_{\boldsymbol{Z}}$ by $\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}}$ in (2), together with

$$\begin{aligned}&\nabla_{\mathbf{C}}\mathbf{a}^{\mathsf{T}}\left(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}}\right)^{-1}\mathbf{a}\\&\quad=-2(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1}\mathbf{a}\mathbf{a}^{\mathsf{T}}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1}\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \quad (69)\\&\nabla_{\mathbf{C}}\mathsf{det}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})\\&\quad=2\mathsf{det}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})(\mathbf{C}\mathbf{R}_{\boldsymbol{N}}\mathbf{C}^{\mathsf{T}})^{-1}\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \quad (70)\end{aligned}$$

where $\mathbf{a}$ is a fixed vector of the appropriate dimension and where we have used [15, Ch. 9, Sec. 9, Exercise 3] and the chain rule in Lemma B.5 in (69) and, [15, Ch. 9, Sec. 10, Exercise 4] in (70). With these expressions at hand, the gradient $\nabla_{\mathbf{C}} P_{\boldsymbol{Y}}(\boldsymbol{y})$ can be written as

$$\begin{aligned}\nabla_{\mathbf{C}} P_{\boldsymbol{Y}}(\boldsymbol{y}) &= \mathsf{E}\left\{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})\right.\\&\quad\left.\cdot\left(\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\boldsymbol{X})(\boldsymbol{y}-\boldsymbol{X})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}-\mathbf{R}_{\boldsymbol{Z}}^{-1}\right)\right\}\mathbf{C}\mathbf{R}_{\boldsymbol{N}}. \quad (71)\end{aligned}$$

To complete the proof, we need to calculate the Hessian matrix, $\mathsf{H}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y})$. First consider the following two Jacobians

$$\mathsf{D}_{\boldsymbol{y}}(\boldsymbol{y}-\boldsymbol{x})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\boldsymbol{x}) = 2(\boldsymbol{y}-\boldsymbol{x})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1} \quad (72)$$

$$\mathsf{D}_{\boldsymbol{y}}\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\boldsymbol{x}) = \mathbf{R}_{\boldsymbol{Z}}^{-1}, \quad (73)$$

which follow directly from [15, Ch. 9, Table 3] and [15, Ch. 9, Sec. 12]. Now, from (72), we can first obtain the Jacobian row vector $\mathsf{D}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y})$ as

$$\mathsf{D}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y}) = -\mathsf{E}\left\{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})(\boldsymbol{y}-\boldsymbol{X})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right\}. \quad (74)$$

Recalling the expression in (73) and that $\mathsf{H}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathsf{D}_{\boldsymbol{y}}(\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}}(\boldsymbol{y}))$ the Hessian matrix becomes

$$\begin{aligned}\mathsf{H}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y}) &= \mathsf{E}\left\{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})\right.\\&\quad\left.\times\left(\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\boldsymbol{X})(\boldsymbol{y}-\boldsymbol{X})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}-\mathbf{R}_{\boldsymbol{Z}}^{-1}\right)\right\}. \quad (75)\end{aligned}$$

By inspection from (71) and (75) the result follows. □

Lemma C.2: Let $\boldsymbol{Y} = \mathbf{G}\boldsymbol{S} + \boldsymbol{Z}$, where $\boldsymbol{S}$ is arbitrarily distributed and $\boldsymbol{Z}$ is a zero-mean Gaussian random variable with covariance matrix $\mathbf{R}_{\boldsymbol{Z}}$ and independent of $\boldsymbol{S}$. Then, the probability density function $P_{\boldsymbol{Y}}(\boldsymbol{y})$ satisfies

$$\nabla_{\mathbf{G}} P_{\boldsymbol{Y}}(\boldsymbol{y}) = -\mathsf{E}\left\{\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\boldsymbol{S}^{\mathsf{T}}\right\}. \quad (76)$$

Proof: First we write

$$\mathsf{D}_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{s}) = -P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{s})(\boldsymbol{y}-\mathbf{G}\boldsymbol{s})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1} \quad (77)$$

where we have used (72). Now, we simply need to notice that

$$\begin{aligned}\nabla_{\mathbf{G}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{s}) &= P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{s})\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\mathbf{G}\boldsymbol{s})\boldsymbol{s}^{\mathsf{T}} \quad (78)\\&= -\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{s})\boldsymbol{s}^{\mathsf{T}} \quad (79)\end{aligned}$$

where $\nabla_{\mathbf{G}}(\boldsymbol{y}-\mathbf{G}\boldsymbol{s})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\mathbf{G}\boldsymbol{s}) = -2\mathbf{R}_{\boldsymbol{Z}}^{-1}(\boldsymbol{y}-\mathbf{G}\boldsymbol{s})\boldsymbol{s}^{\mathsf{T}}$, which follows from [15, Ch. 9, Table 4]. Finally, the result follows from $\nabla_{\mathbf{G}} P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathsf{E}\{\nabla_{\mathbf{G}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\}$. □

Lemma C.3: Let $\boldsymbol{Y} = \mathbf{G}\boldsymbol{S} + \boldsymbol{Z}$, where $\boldsymbol{S}$ is arbitrarily distributed and $\boldsymbol{Z}$ is a zero-mean Gaussian random variable with covariance matrix $\mathbf{R}_{\boldsymbol{Z}}$ and independent of $\boldsymbol{S}$. Then, the conditional expectation $\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\}$ satisfies

$$\mathsf{D}_{\boldsymbol{y}}\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\} = \boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}\mathbf{R}_{\boldsymbol{Z}}^{-1}. \quad (80)$$

Proof:

$$\begin{aligned}&\mathsf{D}_{\boldsymbol{y}}\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\}\\&\quad=\mathsf{D}_{\boldsymbol{y}}\mathsf{E}\left\{\boldsymbol{S}\frac{P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right\} \quad (81)\\&\quad=\mathsf{E}\left\{\boldsymbol{S}\frac{P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{D}_{\boldsymbol{y}}P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})-P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\mathsf{D}_{\boldsymbol{y}}P_{\boldsymbol{Y}}(\boldsymbol{y})}{P_{\boldsymbol{Y}}(\boldsymbol{y})^2}\right\}\\&\quad\quad (82)\\&\quad=\mathsf{E}\left\{\boldsymbol{S}\frac{-P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})(\boldsymbol{y}-\mathbf{G}\boldsymbol{S})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right\}\\&\quad\quad+\mathsf{E}\left\{\boldsymbol{S}\frac{P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})(\boldsymbol{y}-\mathbf{G}\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right\} \quad (83)\\&\quad=(\mathsf{E}\{\boldsymbol{S}\boldsymbol{S}^{\mathsf{T}}\,|\,\boldsymbol{y}\}-\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\}\mathsf{E}\{\boldsymbol{S}^{\mathsf{T}}\,|\,\boldsymbol{y}\})\mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1} \quad (84)\end{aligned}$$

where, in (83) we have used the expression in (77) for $\mathsf{D}_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})$ and also that, from (74),

$$\begin{aligned}\mathsf{D}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y}) &= -\mathsf{E}\left\{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})(\boldsymbol{y}-\boldsymbol{X})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right\} \quad (85)\\&= -P_{\boldsymbol{Y}}(\boldsymbol{y})(\boldsymbol{y}-\mathbf{G}\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}. \quad (86)\end{aligned}$$

Now, expanding the definition in (5) for the conditional MMSE matrix $\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})$, the result in the lemma follows. □

Lemma C.4: Let $\boldsymbol{Y} = \mathbf{G}\boldsymbol{S} + \boldsymbol{Z}$, where $\boldsymbol{S}$ is arbitrarily distributed and $\boldsymbol{Z}$ is a zero-mean Gaussian random variable with covariance matrix $\mathbf{R}_{\boldsymbol{Z}}$ and independent of $\boldsymbol{S}$. Then, the Jacobian and Hessian of $\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y})$ satisfy

$$\mathsf{D}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y}) = (\mathsf{E}\{\boldsymbol{X}\,|\,\boldsymbol{y}\}-\boldsymbol{y})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1} \quad (87)$$

$$\mathsf{H}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathbf{R}_{\boldsymbol{Z}}^{-1}\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})\mathbf{R}_{\boldsymbol{Z}}^{-1}-\mathbf{R}_{\boldsymbol{Z}}^{-1}. \quad (88)$$

Proof: From the expression in (74) we can write

$$\begin{aligned}\mathsf{D}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y}) &= -\frac{\mathsf{E}\left\{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X})(\boldsymbol{y}-\boldsymbol{X})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right\}}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\\&\quad\quad (89)\\&= (\mathsf{E}\{\boldsymbol{X}\,|\,\boldsymbol{y}\}-\boldsymbol{y})^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}. \quad (90)\end{aligned}$$

Now, the Hessian can be computed as

$$\begin{aligned}\mathsf{H}_{\boldsymbol{y}}\mathsf{log}P_{\boldsymbol{Y}}(\boldsymbol{y}) &= \mathsf{D}_{\boldsymbol{y}}\mathbf{R}_{\boldsymbol{Z}}^{-1}(\mathsf{E}\{\boldsymbol{X}\,|\,\boldsymbol{y}\}-\boldsymbol{y}) \quad (91)\\&= \mathbf{R}_{\boldsymbol{Z}}^{-1}(\mathbf{G}\mathsf{D}_{\boldsymbol{y}}\mathsf{E}\{\boldsymbol{S}\,|\,\boldsymbol{y}\}-\mathbf{I}_n) \quad (92)\\&= \mathbf{R}_{\boldsymbol{Z}}^{-1}(\mathbf{G}\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}-\mathbf{I}_n) \quad (93)\end{aligned}$$

where (93) follows from Lemma C.3. □

Lemma C.5: Let $\boldsymbol{Y} = \mathbf{G}\boldsymbol{S} + \boldsymbol{Z}$, where $\boldsymbol{S}$ is arbitrarily distributed (with i-th element denoted by S_i) and $\boldsymbol{Z}$ is a zero-mean Gaussian random variable with covariance matrix $\mathbf{R}_{\boldsymbol{Z}}$ and independent of $\boldsymbol{S}$. Then, the conditional expectation $\mathsf{E}\{S_i \,|\, \boldsymbol{y}\}$ satisfies

$$\nabla_{\mathbf{G}} \mathsf{E}\{S_i \,|\, \boldsymbol{y}\} = \frac{1}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\Big(\mathsf{E}\{S_i \,|\, \boldsymbol{y}\}\mathsf{E}\left\{\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\boldsymbol{S}^{\mathsf{T}}\right\} - \mathsf{E}\left\{S_i \mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\boldsymbol{S}^{\mathsf{T}}\right\}\Big). \tag{94}$$

Proof:

$$\begin{aligned}
\nabla_{\mathbf{G}} &\mathsf{E}\{S_i \,|\, \boldsymbol{y}\} \\
&= \nabla_{\mathbf{G}} \mathsf{E}\left\{S_i \frac{P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right\} &(95)\\
&= \mathsf{E}\left\{S_i \frac{-\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\boldsymbol{S}^{\mathsf{T}}}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right\} \\
&\quad + \mathsf{E}\left\{S_i \frac{P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\mathsf{E}\left\{\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\boldsymbol{S}^{\mathsf{T}}\right\}}{P_{\boldsymbol{Y}}(\boldsymbol{y})^2}\right\} &(96)\\
&= \frac{1}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\Big(-\mathsf{E}\left\{S_i \mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\boldsymbol{S}^{\mathsf{T}}\right\} \\
&\quad + \mathsf{E}\left\{S_i \frac{P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right\}\mathsf{E}\left\{\mathsf{D}_{\boldsymbol{y}}^{\mathsf{T}} P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})\boldsymbol{S}^{\mathsf{T}}\right\}\Big) &(97)
\end{aligned}$$

where (96) follows from Lemma C.2 and from (79). □

D. Proofs of Theorems 1, 2, 3, and 4

Proof of Theorem 1: Since $\mathbf{J}_{\boldsymbol{Y}}$ is a symmetric matrix, its Jacobian is

$$\begin{aligned}
\mathsf{D}_{\mathbf{C}}\mathbf{J}_{\boldsymbol{Y}} &= \mathsf{D}_{\mathbf{C}}\mathsf{vech}\mathbf{J}_{\boldsymbol{Y}} &(98)\\
&= \mathsf{D}_{\mathbf{C}}\mathbf{D}_n^{+}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}} &(99)\\
&= \mathbf{D}_n^{+}\mathsf{D}_{\mathbf{C}}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}} &(100)\\
&= \mathbf{D}_n^{+}(-2\mathbf{N}_n\mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}) &(101)\\
&= -2\mathbf{D}_n^{+}\mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}) \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n) &(102)
\end{aligned}$$

where (99) follows from (57) and (100) follows from Lemma B.7.2. The expression for $\mathsf{D}_{\mathbf{C}}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}}$ is derived in Appendix E, which yields (101) and (102) follows from Lemma A.3. □

Proof of Theorem 2: The proof is analogous to that of Theorem 1 with the appropriate notation adaptation. The calculation of $\mathsf{D}_{\mathbf{G}}\mathsf{vec}\mathbf{E}_{\boldsymbol{S}}$ can be found in Appendix F. □

Proof of Theorem 3: The Jacobians $\mathsf{D}_{\mathbf{P}}\mathbf{E}_{\boldsymbol{S}}$ and $\mathsf{D}_{\mathbf{H}}\mathbf{E}_{\boldsymbol{S}}$ follow from the Jacobian $\mathsf{D}_{\mathbf{G}}\mathbf{E}_{\boldsymbol{S}}$ calculated in Theorem 2 applying the chain rules $\mathsf{D}_{\mathbf{P}}\mathbf{E}_{\boldsymbol{S}} = \mathsf{D}_{\mathbf{G}}\mathbf{E}_{\boldsymbol{S}} \cdot \mathsf{D}_{\mathbf{P}}\mathbf{G}$ and $\mathsf{D}_{\mathbf{H}}\mathbf{E}_{\boldsymbol{S}} = \mathsf{D}_{\mathbf{G}}\mathbf{E}_{\boldsymbol{S}} \cdot \mathsf{D}_{\mathbf{H}}\mathbf{G}$, where $\mathbf{G} = \mathbf{HP}$ and, from Lemma B.7.1, we have that $\mathsf{D}_{\mathbf{P}}\mathbf{G} = \mathbf{I}_m \otimes \mathbf{H}$ and $\mathsf{D}_{\mathbf{H}}\mathbf{G} = \mathbf{P}^{\mathsf{T}} \otimes \mathbf{I}_n$.

Similarly, the Jacobian $\mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}}$ can be calculated by applying $\mathsf{D}_{\mathbf{C}}\mathbf{J}_{\boldsymbol{Y}} = \mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}} \cdot \mathsf{D}_{\mathbf{C}}\mathbf{R}_{\boldsymbol{Z}}$, where $\mathsf{D}_{\mathbf{C}}\mathbf{R}_{\boldsymbol{Z}} = 2\mathbf{D}_n^{+}(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n)$ as in Lemma B.7.5. Recalling that, in this case, the matrix $\mathbf{C}$ is a dummy variable that is used only to obtain $\mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}}$ through the chain rule, the factor $(\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \mathbf{I}_n)$ can be eliminated from both sides of the equation.

Finally, the Jacobian $\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{J}_{\boldsymbol{Y}}$ follows from the chain rule $\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{J}_{\boldsymbol{Y}} = \mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}} \cdot \mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{R}_{\boldsymbol{Z}}$, with $\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{R}_{\boldsymbol{Z}} = \mathbf{D}_n^{+}(\mathbf{C} \otimes \mathbf{C})\mathbf{D}_{n'}$ (see Lemma B.7.3), and using $\mathbf{D}_n^{+}(\mathbf{A} \otimes \mathbf{A})\mathbf{D}_n\mathbf{D}_n^{+} = \mathbf{D}_n^{+}(\mathbf{A} \otimes \mathbf{A})\mathbf{N}_n = \mathbf{D}_n^{+}\mathbf{N}_n(\mathbf{A} \otimes \mathbf{A}) = \mathbf{D}_n^{+}(\mathbf{A} \otimes \mathbf{A})$. □

Proof of Theorem 4: The developments leading to the expressions for the Hessian matrices $\mathsf{H}_{\mathbf{P}}h(\boldsymbol{Y})$, $\mathsf{H}_{\mathbf{H}}h(\boldsymbol{Y})$, and $\mathsf{H}_{\mathbf{C}}h(\boldsymbol{Y})$ follow a very similar pattern. Consequently, we will present only one of them here.

Consider the Hessian $\mathsf{H}_{\mathbf{P}}h(\boldsymbol{Y})$, from the expression for the Jacobian $\mathsf{D}_{\mathbf{P}}h(\boldsymbol{Y})$ in (25) it follows that

$$\begin{aligned}
\mathsf{H}_{\mathbf{P}}h(\boldsymbol{Y}) &= \mathsf{D}_{\mathbf{P}}\mathsf{vec}\left(\mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{P}\mathbf{E}_{\boldsymbol{S}}\right) &(103)\\
&= \left(\mathbf{E}_{\boldsymbol{S}} \otimes \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\right) \\
&\quad + \left(\mathbf{I}_m \otimes \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}\mathbf{P}\right)\mathbf{D}_m\mathsf{D}_{\mathbf{P}}\mathbf{E}_{\boldsymbol{S}} &(104)
\end{aligned}$$

where in (104) we have used Lemma B.7.1 adding the matrix $\mathbf{D}_m$ because $\mathbf{E}_{\boldsymbol{S}}$ is a symmetric matrix. The final expression for $\mathsf{H}_{\mathbf{P}}h(\boldsymbol{Y})$ is obtained by plugging in (104) the expression for $\mathsf{D}_{\mathbf{P}}\mathbf{E}_{\boldsymbol{S}}$ obtained in Theorem 3.

From (28), the expression for $\mathsf{H}_{\mathbf{R}_{\boldsymbol{Z}}}h(\boldsymbol{Y})$ becomes

$$\begin{aligned}
\mathsf{H}_{\mathbf{R}_{\boldsymbol{Z}}}h(\boldsymbol{Y}) &= \frac{1}{2}\mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{D}_n^{\mathsf{T}}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}} &(105)\\
&= \frac{1}{2}\mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{D}_n^{\mathsf{T}}\mathbf{D}_n\mathsf{vech}\mathbf{J}_{\boldsymbol{Y}} &(106)\\
&= \frac{1}{2}\mathbf{D}_n^{\mathsf{T}}\mathbf{D}_n\mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}} &(107)
\end{aligned}$$

where, in last equality, we have used Lemma B.7.2. The result now follows from the expression for $\mathsf{D}_{\mathbf{R}_{\boldsymbol{Z}}}\mathbf{J}_{\boldsymbol{Y}}$ given in Theorem 3.

Finally, the Hessian matrix $\mathsf{H}_{\mathbf{R}_{\boldsymbol{N}}}h(\boldsymbol{Y})$ can be computed from its Jacobian $\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}h(\boldsymbol{Y})$ in (29) as

$$\begin{aligned}
\mathsf{H}_{\mathbf{R}_{\boldsymbol{N}}}h(\boldsymbol{Y}) &= \frac{1}{2}\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{D}_{n'}^{\mathsf{T}}\mathsf{vec}(\mathbf{C}^{\mathsf{T}}\mathbf{J}_{\boldsymbol{Y}}\mathbf{C}) &(108)\\
&= \frac{1}{2}\mathbf{D}_{n'}^{\mathsf{T}}(\mathbf{C}^{\mathsf{T}} \otimes \mathbf{C}^{\mathsf{T}})\mathbf{D}_n\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{J}_{\boldsymbol{Y}} &(109)
\end{aligned}$$

where we have used Lemmas A.4 and B.7.2. The expression for $\mathsf{D}_{\mathbf{R}_{\boldsymbol{N}}}\mathbf{J}_{\boldsymbol{Y}}$ can be found in Theorem 3. □

E. Calculation of $\mathsf{D}_{\mathbf{C}}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}}$

Consider the expression for the entries of the Jacobian of $\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}}$: $[\mathsf{D}_{\mathbf{C}}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}}]_{i+(j-1)n,k+(l-1)n} = \mathsf{D}_{\mathbf{C}_{kl}}[\mathbf{J}_{\boldsymbol{Y}}]_{ij}$, such that $\{i,j,k\} \in [1,n]$ and $l \in [1,n']$, which will be used throughout this proof. From (8), the entries of the Fisher information matrix are

$$[\mathbf{J}_{\boldsymbol{Y}}]_{ij} = -\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j}\mathrm{d}\boldsymbol{y}. \tag{110}$$

We now differentiate the expression above with respect to the entries of the matrix $\mathbf{C}$ and we get

$$\begin{aligned}
\mathsf{D}_{\mathbf{C}_{kl}}[\mathbf{J}_{\boldsymbol{Y}}]_{ij} = &-\int \frac{\partial P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial \mathbf{C}_{kl}}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j}\mathrm{d}\boldsymbol{y} \\
&-\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^2}{\partial y_i \partial y_j}\left(\frac{1}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\frac{\partial P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial \mathbf{C}_{kl}}\right)\mathrm{d}\boldsymbol{y}
\end{aligned} \tag{111}$$

where the interchange of the order of integration and differentiation can be justified from the Dominated Convergence Theorem following similar steps as in [1, Appendix B]. Now, using Lemma C.1, the entries of the Jacobian matrix are

$$\begin{aligned}\mathsf{D}_{\mathbf{C}_{kl}}[\mathbf{J}_{\boldsymbol{Y}}]_{ij} &= -\int [\mathsf{H}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{kl}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j}\mathrm{d}\boldsymbol{y} \\ &\quad -\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^2}{\partial y_i \partial y_j}\left(\frac{1}{P_{\boldsymbol{Y}}(\boldsymbol{y})}[\mathsf{H}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{kl}\right)\mathrm{d}\boldsymbol{y}. \end{aligned} \tag{112}$$

Expanding the expression for $[\mathsf{H}_{\boldsymbol{y}} P_{\boldsymbol{Y}}(\boldsymbol{y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{kl}$ we get

$$\begin{aligned}\mathsf{D}_{\mathbf{C}_{kl}}[\mathbf{J}_{\boldsymbol{Y}}]_{ij} &= -\sum_{r=1}^{n}[\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{rl}\left(\int \frac{\partial^2 P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_k \partial y_r}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j}\mathrm{d}\boldsymbol{y}\right. \\ &\quad \left. + \int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^2}{\partial y_i \partial y_j}\left(\frac{1}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\frac{\partial^2 P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_k \partial y_r}\right)\mathrm{d}\boldsymbol{y}\right). \end{aligned} \tag{113}$$

Integrating by parts the first term in (113) twice, we obtain[6]

$$\int \frac{\partial^2 P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_k \partial y_r}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j}\mathrm{d}\boldsymbol{y} = \int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^4 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j \partial y_k \partial y_r}\mathrm{d}\boldsymbol{y} \tag{114}$$

and applying the scalar version of the logarithm identity in (7), the second term in the right hand side of (113) becomes

$$\begin{aligned}&\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^2}{\partial y_i \partial y_j}\left(\frac{1}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\frac{\partial^2 P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_k \partial y_r}\right)\mathrm{d}\boldsymbol{y} \\ &= \int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^4 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j \partial y_k \partial y_r}\mathrm{d}\boldsymbol{y} \\ &\quad + \int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_j \partial y_r}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_k}\mathrm{d}\boldsymbol{y} \\ &\quad + \int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_r}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_j \partial y_k}\mathrm{d}\boldsymbol{y} \\ &\quad + \int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^3 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j \partial y_k}\frac{\partial \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_r}\mathrm{d}\boldsymbol{y} \\ &\quad + \int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^3 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j \partial y_r}\frac{\partial \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_k}\mathrm{d}\boldsymbol{y}. \end{aligned} \tag{115}$$

Integrating by parts the last term in (115), we have

$$\begin{aligned}&\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^3 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j \partial y_r}\frac{\partial \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_k}\mathrm{d}\boldsymbol{y} \\ &= \int \frac{\partial P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_k}\frac{\partial^3 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j \partial y_r}\mathrm{d}\boldsymbol{y} \qquad (116) \\ &= -\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial^4 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_j \partial y_k \partial y_r}\mathrm{d}\boldsymbol{y}. \qquad (117)\end{aligned}$$

[6]From the expression $\int u\mathrm{d}v = uv - \int v\mathrm{d}u$, the identity in (114) (and other similar expressions in the following) is obtained by proving that the uv term vanishes. The detailed proof is omitted for the sake of space. A similar derivation for the scalar case can be found in [26].

Plugging (114), (115), and (117) into (113), we finally obtain

$$\begin{aligned}\mathsf{D}_{\mathbf{C}_{kl}}[\mathbf{J}_{\boldsymbol{Y}}]_{ij} &= \\ &-\int P_{\boldsymbol{Y}}(\boldsymbol{y})\left(\sum_{r=1}^{n}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_j \partial y_r}[\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{rl}\right)\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_k}\mathrm{d}\boldsymbol{y} \\ &-\int P_{\boldsymbol{Y}}(\boldsymbol{y})\left(\sum_{r=1}^{n}\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_i \partial y_r}[\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{rl}\right)\frac{\partial^2 \log P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial y_j \partial y_k}\mathrm{d}\boldsymbol{y}.\end{aligned}$$

Now, recalling that $\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathsf{H}_{\boldsymbol{y}}\log P_{\boldsymbol{Y}}(\boldsymbol{y})$ and identifying the elements of the two matrices $\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}}$ and $\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{y})$ with the terms in last equation, we obtain

$$\begin{aligned}\mathsf{D}_{\mathbf{C}_{kl}}[\mathbf{J}_{\boldsymbol{Y}}]_{ij} = &-\mathsf{E}\{[\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{jl}[\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})]_{ik}\} \\ &-\mathsf{E}\{[\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}}]_{il}[\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})]_{jk}\}.\end{aligned} \tag{118}$$

Finally, as $[\mathsf{D}_{\mathbf{C}}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}}]_{i+(j-1)n,k+(l-1)n} = \mathsf{D}_{\mathbf{C}_{kl}}[\mathbf{J}_{\boldsymbol{Y}}]_{ij}$ and applying Lemma A.1, it can be shown that

$$\begin{aligned}\mathsf{D}_{\mathbf{C}}\mathsf{vec}\mathbf{J}_{\boldsymbol{Y}} &= -\mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\} \\ &\quad - \mathbf{K}_n\mathsf{E}\{(\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y}))\} \qquad (119) \\ &= -2\mathbf{N}_n\mathsf{E}\{\boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\mathbf{C}\mathbf{R}_{\boldsymbol{N}} \otimes \boldsymbol{\Gamma}_{\boldsymbol{Y}}(\boldsymbol{Y})\}. \qquad (120)\end{aligned}$$

F. Calculation of $\mathsf{D}_{\mathbf{G}}\mathsf{vec}\mathbf{E}_{\boldsymbol{S}}$

Consider the expression for the entries of the matrix $\mathbf{E}_{\boldsymbol{S}}$, $[\mathbf{E}_{\boldsymbol{S}}]_{ij} = \mathsf{E}\{S_iS_j\} - \mathsf{E}\{\mathsf{E}\{S_i\,|\,\boldsymbol{Y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{Y}\}\}$, from which it follows that

$$\begin{aligned}\mathsf{D}_{\mathbf{G}_{kl}}[\mathbf{E}_{\boldsymbol{S}}]_{ij} = &-\int \frac{\partial P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial \mathbf{G}_{kl}}\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y} \\ &-\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial \mathsf{E}\{S_i\,|\,\boldsymbol{y}\}}{\partial \mathbf{G}_{kl}}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y} \\ &-\int P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\frac{\partial \mathsf{E}\{S_j\,|\,\boldsymbol{y}\}}{\partial \mathbf{G}_{kl}}\mathrm{d}\boldsymbol{y}\end{aligned} \tag{121}$$

where, throughout this proof $\{i, j, l\} \in [1, m]$ and $k \in [1, n]$ and where, as in Appendix E, the justification of the interchange of the order of derivation and integration and two other interchanges below follow similar steps as in [1, Appendix B].

Observe that the second and third terms in (121) have the same structure and, thus, we will deal with them jointly. The first term in (121) can be rewritten as

$$\begin{aligned}&-\int \frac{\partial P_{\boldsymbol{Y}}(\boldsymbol{y})}{\partial \mathbf{G}_{kl}}\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y} \\ &\quad = \int \mathsf{E}\left\{S_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}\end{aligned} \tag{122}$$

where we have used Lemma C.2. Using Lemma C.5, the second term in (121) can be computed as

$$\begin{aligned}&-\int P_{\boldsymbol{Y}}(\boldsymbol{y})\frac{\partial \mathsf{E}\{S_i\,|\,\boldsymbol{y}\}}{\partial \mathbf{G}_{kl}}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y} \\ &\quad = -\int \mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\left\{S_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y} \\ &\quad\quad + \int \mathsf{E}\left\{S_iS_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}.\end{aligned} \tag{123}$$

Plugging (122) and (123) into (121) we can write

$$\begin{aligned}\mathsf{D}_{\mathbf{G}_{kl}}[\mathbf{E}_{\boldsymbol{S}}]_{ij} = &-\int \mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\left\{S_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}\\ &+\int \mathsf{E}\left\{S_iS_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}\\ &+\int \mathsf{E}\left\{S_jS_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}.\end{aligned} \tag{124}$$

The first term in (124) can be reformulated as

$$\begin{aligned}-&\int \mathsf{E}\left\{S_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}\\ &= -\int \frac{\partial P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_l\,|\,\boldsymbol{y}\}}{\partial y_k}\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y} && (125)\\ &= \int P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_l\,|\,\boldsymbol{y}\}\frac{\partial \mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}}{\partial y_k}\mathrm{d}\boldsymbol{y} && (126)\end{aligned}$$

where in the last step we have integrated by parts. Making use of Lemma C.3 yields

$$\begin{aligned}&\int P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_l\,|\,\boldsymbol{y}\}\frac{\partial \mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}}{\partial y_k}\mathrm{d}\boldsymbol{y}\\ &\quad= \int P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_l\,|\,\boldsymbol{y}\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\\ &\qquad\cdot\left[\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right]_{ik}\mathrm{d}\boldsymbol{y}\\ &\qquad+\int P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_l\,|\,\boldsymbol{y}\}\mathsf{E}\{S_i\,|\,\boldsymbol{y}\}\\ &\qquad\cdot\left[\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right]_{jk}\mathrm{d}\boldsymbol{y}.\end{aligned} \tag{127}$$

We now proceed to the computation of the second and third terms in (124). We have

$$\begin{aligned}&\int \mathsf{E}\left\{S_iS_l\frac{\partial P_{\boldsymbol{Y}|\boldsymbol{S}}(\boldsymbol{y}|\boldsymbol{S})}{\partial y_k}\right\}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}\\ &\quad= \int \frac{\partial P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_iS_l\,|\,\boldsymbol{y}\}}{\partial y_k}\mathsf{E}\{S_j\,|\,\boldsymbol{y}\}\mathrm{d}\boldsymbol{y} && (128)\\ &\quad= -\int P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_iS_l\,|\,\boldsymbol{y}\}\frac{\partial \mathsf{E}\{S_j\,|\,\boldsymbol{y}\}}{\partial y_k}\mathrm{d}\boldsymbol{y} && (129)\\ &\quad= -\int P_{\boldsymbol{Y}}(\boldsymbol{y})\mathsf{E}\{S_iS_l\,|\,\boldsymbol{y}\}\left[\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right]_{jk}\mathrm{d}\boldsymbol{y} && (130)\end{aligned}$$

where in (130) we have applied Lemma C.3.

Plugging (127) and (130) into (124), we finally have

$$\begin{aligned}\mathsf{D}_{\mathbf{G}_{kl}}[\mathbf{E}_{\boldsymbol{S}}]_{ij} = &-\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})]_{jl}\left[\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right]_{ik}\right\}\\ &-\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})]_{il}\left[\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\mathbf{G}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\right]_{jk}\right\}.\end{aligned} \tag{131}$$

From $\mathsf{D}_{\mathbf{G}_{kl}}[\mathbf{E}_{\boldsymbol{S}}]_{ij} = [\mathsf{D}_{\mathbf{G}}\mathsf{vec}\mathbf{E}_{\boldsymbol{S}}]_{i+(j-1)n,k+(l-1)m}$ and applying Lemma A.1, the desired result follows similarly as in (120) in Appendix E.

G. Proofs of Concavity Results

Proof of Corollary 1: First, we consider the result in [1, Corollary 1], $\mathsf{D}_{\mathsf{snr}}I(\boldsymbol{S};\boldsymbol{Y}) = \mathsf{Tr}(\mathbf{R}_{\mathbf{H}}\mathbf{E}_{\boldsymbol{S}})$. Choosing $\mathbf{P} = \sqrt{\mathsf{snr}}\mathbf{I}_p$ and applying Theorem 4 and Lemma B.5 we obtain

$$\begin{aligned}&\mathsf{H}_{\mathsf{snr}}I(\boldsymbol{S};\boldsymbol{Y})\\ &\quad= \frac{1}{2}\mathsf{D}_{\mathbf{E}_{\boldsymbol{S}}}\mathsf{Tr}(\mathbf{R}_{\mathbf{H}}\mathbf{E}_{\boldsymbol{S}})\cdot\mathsf{D}_{\mathbf{P}}\mathbf{E}_{\boldsymbol{S}}\cdot\mathsf{D}_{\mathsf{snr}}\mathbf{P} && (132)\\ &\quad= \frac{1}{2}\mathsf{vec}^{\mathsf{T}}(\mathbf{R}_{\mathbf{H}})\mathbf{D}_p(-2\mathbf{D}_p^{+}\mathsf{E}\{\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\otimes\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}\\ &\qquad\cdot(\mathbf{I}_p\otimes\sqrt{\mathsf{snr}}\mathbf{R}_{\mathbf{H}}))\frac{1}{2\sqrt{\mathsf{snr}}}\mathsf{vec}\mathbf{I}_p && (133)\\ &\quad= -\frac{1}{2}\mathsf{vec}^{\mathsf{T}}(\mathbf{R}_{\mathbf{H}})\mathsf{E}\{\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\otimes\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\}\mathsf{vec}\mathbf{R}_{\mathbf{H}} && (134)\end{aligned}$$

where we have used Lemma A.4 and the fact that, for symmetric matrices, $\mathsf{vec}^{\mathsf{T}}(\mathbf{R}_{\mathbf{H}})\mathbf{N}_p = \mathsf{vec}^{\mathsf{T}}\mathbf{R}_{\mathbf{H}}$ as in (55).

From (134), it readily follows that the mutual information is a concave function of the snr parameter because, from Lemma H.3 we have that $\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\otimes\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{y})\geq\mathbf{0}$, $\forall\boldsymbol{y}$, and, consequently, $\mathsf{H}_{\mathsf{snr}}I(\boldsymbol{S};\boldsymbol{Y})\leq 0$. Finally, applying again Lemma A.4 and $\mathsf{vec}^{\mathsf{T}}\mathbf{A}\mathsf{vec}\mathbf{B} = \mathsf{Tr}(\mathbf{A}^{\mathsf{T}}\mathbf{B})$, the desired expression follows. □

Proof of Theorem 5: To simplify the notation, in this proof we consider $m\geq p$, which implies that $m' = p$ and thus $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma}$. The derivation for $m < p$ follows similar steps.

The Hessian of the mutual information $\mathsf{H}_{\boldsymbol{\lambda}}I(\boldsymbol{S};\boldsymbol{Y})$ can be obtained from the Hessian chain rule in Lemma B.5 as

$$\begin{aligned}\mathsf{H}_{\boldsymbol{\lambda}}I(\boldsymbol{S};\boldsymbol{Y}) = &\mathsf{D}_{\boldsymbol{\lambda}}^{\mathsf{T}}\mathbf{P}\mathsf{H}_{\mathbf{P}}I(\boldsymbol{S};\boldsymbol{Y})\mathsf{D}_{\boldsymbol{\lambda}}\mathbf{P}\\ &+(\mathsf{D}_{\mathbf{P}}I(\boldsymbol{S};\boldsymbol{Y})\otimes\mathbf{I}_p)\mathsf{H}_{\boldsymbol{\lambda}}\mathbf{P}.\end{aligned} \tag{135}$$

Now we need to calculate $\mathsf{D}_{\boldsymbol{\lambda}}\mathbf{P}$ and $\mathsf{H}_{\boldsymbol{\lambda}}\mathbf{P}$. We have

$$\begin{aligned}\mathsf{D}_{\boldsymbol{\lambda}}\mathbf{P} &= \mathsf{D}_{\boldsymbol{\lambda}}\left(\mathbf{U}_{\mathbf{H}}\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\right) && (136)\\ &= (\mathbf{V}_{\mathbf{P}}\otimes\mathbf{U}_{\mathbf{H}})\mathsf{D}_{\boldsymbol{\lambda}}\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}) && (137)\\ &= \frac{1}{2}(\mathbf{V}_{\mathbf{P}}\otimes\mathbf{U}_{\mathbf{H}})\mathbf{S}_p(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1} && (138)\end{aligned}$$

where, in (137), we have used Lemmas A.4 and B.7.2 and where the last step follows from

$$[\mathsf{D}_{\boldsymbol{\lambda}}\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}})]_{i+(j-1)p,k} = \frac{\partial(\sqrt{\lambda_i}\delta_{ij})}{\partial\lambda_k} = \frac{\delta_{ijk}}{2\sqrt{\lambda_k}} \tag{139}$$

with $\{i,j,k\}\in[1,p]$ and from the definition of the reduction matrix $\mathbf{S}_p$ in (59).

Following steps similar to the derivation of $\mathsf{D}_{\boldsymbol{\lambda}}\mathbf{P}$, the Hessian matrix $\mathsf{H}_{\boldsymbol{\lambda}}\mathbf{P}$ is obtained according to

$$\mathsf{H}_{\boldsymbol{\lambda}}\mathbf{P} = -\frac{1}{4}((\mathbf{V}_{\mathbf{P}}\otimes\mathbf{U}_{\mathbf{H}})\mathbf{S}_p\otimes\mathbf{I}_p)\mathbf{S}_p(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-3}. \tag{140}$$

Plugging (138) and (140) in (135) together with the expressions for the Jacobian matrix $\mathsf{D}_{\mathbf{P}} I(\boldsymbol{S};\boldsymbol{Y})$ and the Hessian matrix $\mathsf{H}_{\mathbf{P}} I(\boldsymbol{S};\boldsymbol{Y})$ given in Remark 3 of Lemma 1 and in Remark 4 of Theorem 4, respectively, and operating we obtain

$$\begin{aligned}
&\mathsf{H}_{\boldsymbol{\lambda}} I(\boldsymbol{S};\boldsymbol{Y}) \\
&= \frac{1}{4}(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1}\mathbf{S}_p^{\mathsf{T}} \Big(\mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}} \otimes \mathbf{Diag}(\boldsymbol{\sigma}) \\
&\quad - 2(\mathbf{I}_p \otimes \mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}}))\mathbf{N}_p \\
&\quad \cdot \mathsf{E}\left\{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\mathbf{V}_{\mathbf{P}} \otimes \mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\mathbf{V}_{\mathbf{P}}\right\} \\
&\quad \cdot (\mathbf{I}_p \otimes \mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}}))\Big)\,\mathbf{S}_p(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1} \\
&\quad - \frac{1}{4}\left(\mathsf{vec}^{\mathsf{T}}\left(\mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}})\mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}\right)\mathbf{S}_p \otimes \mathbf{I}_p\right) \\
&\quad \cdot \mathbf{S}_p(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-3}
\end{aligned} \tag{141}$$

where it can be noted that the dependence of $\mathsf{H}_{\boldsymbol{\lambda}} I(\boldsymbol{S};\boldsymbol{Y})$ on $\mathbf{U}_{\mathbf{H}}$ has disappeared.

Now, applying Lemma A.2, the first term in (141) becomes

$$\begin{aligned}
&(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1}\mathbf{S}_p^{\mathsf{T}}(\mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}} \otimes \mathbf{Diag}(\boldsymbol{\sigma}))\mathbf{S}_p(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1} \\
&\quad = (\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1}(\mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}} \circ \mathbf{Diag}(\boldsymbol{\sigma}))(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1} \qquad (142) \\
&\quad = \mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}} \circ \mathbf{Diag}(\boldsymbol{\sigma}\circ(1/\boldsymbol{\lambda})) \qquad (143)
\end{aligned}$$

whereas the third term in (141) can be expressed as

$$\begin{aligned}
&\left(\mathsf{vec}^{\mathsf{T}}\left(\mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}})\mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}\right)\mathbf{S}_p \otimes \mathbf{I}_m\right)\mathbf{S}_p(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-3} \\
&\quad = \mathbf{Diag}\left(\mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}})\mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}\right)(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-3} \qquad (144) \\
&\quad = \mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}} \circ \mathbf{Diag}(\boldsymbol{\sigma}\circ(1/\boldsymbol{\lambda})) \qquad (145)
\end{aligned}$$

where in (144) we have used that, for any square matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, $[\mathsf{vec}^{\mathsf{T}}(\mathbf{A})\mathbf{S}_p]_k = \sum_{i,j=1}^{p}\mathbf{A}_{ij}\delta_{ijk} = \mathbf{A}_{kk}$ and $[(\mathbf{diag}(\mathbf{A})^{\mathsf{T}} \otimes \mathbf{I}_p)\mathbf{S}_p]_{kl} = \sum_{i,j}^{p}\mathbf{A}_{jj}\delta_{ki}\delta_{ijl} = \mathbf{A}_{ll}\delta_{kl}$.

Now, from (143) and (145) we see that the first and third terms in (141) cancel out and, recalling that $\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{\Phi}_{\boldsymbol{S}}(\boldsymbol{Y})\mathbf{V}_{\mathbf{P}} = \boldsymbol{\Phi}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{y})$, the expression for $\mathsf{H}_{\boldsymbol{\lambda}} I(\boldsymbol{S};\boldsymbol{Y})$ simplifies to

$$\begin{aligned}
&\mathsf{H}_{\boldsymbol{\lambda}} I(\boldsymbol{S};\boldsymbol{Y}) \\
&= -\frac{1}{4}(\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1} \\
&\quad \cdot \mathsf{E}\Big\{\boldsymbol{\Phi}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{Y}) \circ \mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}})\ \boldsymbol{\Phi}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{Y})\mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}}) \\
&\quad + \mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}})\boldsymbol{\Phi}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{Y}) \circ \boldsymbol{\Phi}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{Y})\mathbf{Diag}(\boldsymbol{\sigma}\circ\sqrt{\boldsymbol{\lambda}})\Big\} \\
&\quad \cdot (\mathbf{Diag}(\sqrt{\boldsymbol{\lambda}}))^{-1}.
\end{aligned} \tag{146}$$

Now, from simple inspection of the expression in (146) and recalling the properties of the Schur product, the desired result follows. □

Proof of Counterexample 1: We present a two-dimensional counterexample. Assume that the noise is white $\mathbf{R}_{\boldsymbol{Z}} = \mathbf{I}_2$ and consider the following channel and precoder

$$\mathbf{H}_{ce} = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}, \quad \mathbf{P}_{ce} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} \tag{147}$$

where $\beta \in (0,1]$ and assume that the distribution for the signal vector $\boldsymbol{S}$ has two equally likely mass points,

$$\boldsymbol{s}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{s}^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{148}$$

Accordingly, we define the noiseless received vector as $\boldsymbol{r}^{(k)} = \mathbf{H}_{ce}\mathbf{P}_{ce}\boldsymbol{s}^{(k)}$, for $k = \{1,2\}$, which yields

$$\boldsymbol{r}^{(1)} = \sqrt{\lambda_1}\begin{pmatrix} 1 \\ \beta \end{pmatrix}, \quad \boldsymbol{r}^{(2)} = \sqrt{\lambda_2}\begin{pmatrix} \beta \\ 1 \end{pmatrix}. \tag{149}$$

We now define the mutual information for this counterexample as $I_{ce}(\lambda_1,\lambda_2,\beta) = I(\boldsymbol{S};\mathbf{H}_{ce}\mathbf{P}_{ce}\boldsymbol{S}+\boldsymbol{Z})$. We use the fact that, as $\mathbf{R}_{\boldsymbol{Z}} = \mathbf{I}_2$, the mutual information can be expressed as an increasing function f of the squared distance of the two only possible received vectors $\mathrm{d}^2(\lambda_1,\lambda_2,\beta) = \|\boldsymbol{r}^{(1)} - \boldsymbol{r}^{(2)}\|^2$, which is denoted by $I_{ce}(\lambda_1,\lambda_2,\beta) = f(\mathrm{d}^2(\lambda_1,\lambda_2,\beta))$.

For a fixed value of β, we want to study the concavity of $I_{ce}(\lambda_1,\lambda_2,\beta)$ with respect to (λ_1,λ_2). In order to do so, we restrict ourselves to the study of concavity along straight lines of the type $\lambda_1 + \lambda_2 = \rho$, with $\rho > 0$, which is sufficient to disprove the concavity.

Operating with the received vectors, we obtain

$$\mathrm{d}^2(\rho,0,\beta) = \mathrm{d}^2(0,\rho,\beta) = \rho(1+\beta^2) > \rho \tag{150}$$

$$\mathrm{d}^2(\rho/2,\rho/2,\beta) = \rho(1-\beta)^2 < \rho \tag{151}$$

from which it readily follows that $I_{ce}(\rho,0,\beta) = I_{ce}(0,\rho,\beta) > I_{ce}(\rho/2,\rho/2,\beta)$, which contradicts the concavity hypothesis. □

Proof of Theorem 6: To simplify the notation, in this proof we consider $m \geq p$, which implies that $m' = p$ and thus $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma}$. Moreover, we assume that the elements of $\boldsymbol{\sigma}$ are sorted in decreasing order. The derivation for the case where $m < p$ or the elements of $\boldsymbol{\sigma}$ are not sorted is similar.

First, let us prove (49). From the definition of the entropy power and applying the chain rule for Hessians in Lemma B.5 we obtain

$$\mathsf{H}_{\boldsymbol{\lambda}} N(\boldsymbol{Y}) = \frac{2N(\boldsymbol{Y})}{n}\left(\frac{2\mathsf{D}_{\boldsymbol{\lambda}}^{\mathsf{T}} h(\boldsymbol{Y})\mathsf{D}_{\boldsymbol{\lambda}} h(\boldsymbol{Y})}{n} + \mathsf{H}_{\boldsymbol{\lambda}} h(\boldsymbol{Y})\right). \tag{152}$$

Now, from [5, eq. (61)] we can write that $\mathsf{D}_{\boldsymbol{\lambda}}^{\mathsf{T}} h(\boldsymbol{Y}) = (1/2)\mathbf{Diag}(\boldsymbol{\sigma})\mathbf{diag}(\mathbf{E}_{\mathbf{V}_{\mathbf{P}}^{\mathsf{T}}\boldsymbol{S}})$. Incorporating the expression for $\mathsf{H}_{\boldsymbol{\lambda}} h(\boldsymbol{Y})$ calculated in Theorem 5, the result in (49) follows.

To prove the negative semidefiniteness of the expression in (152), we first define the positive semidefinite matrix $\mathbf{A}(\boldsymbol{y}) \in \mathbb{S}_{+}^{p'}$, which is obtained by selecting the first $p' = \mathsf{rank}(\mathbf{R}_{\boldsymbol{H}})$ columns and rows of the positive semidefinite matrix $\mathbf{Diag}(\sqrt{\boldsymbol{\sigma}})\boldsymbol{\Phi}_{\boldsymbol{V}_{\mathrm{P}}^{\mathsf{T}}\boldsymbol{S}}(\boldsymbol{y})\mathbf{Diag}(\sqrt{\boldsymbol{\sigma}})$. With this definition, it is now easy to see that the expression

$$\frac{\mathsf{E}\{\mathbf{diag}(\mathbf{A}(\boldsymbol{y}))\}\mathsf{E}\{\mathbf{diag}(\mathbf{A}(\boldsymbol{y}))^{\mathsf{T}}\}}{n} - \mathsf{E}\{\mathbf{A}(\boldsymbol{y}) \circ \mathbf{A}(\boldsymbol{y})\} \tag{153}$$

coincides (up to the factor $2N(\boldsymbol{Y})/n$) with the first p' rows and columns of the Hessian matrix $\mathsf{H}_{\boldsymbol{\lambda}} N(\boldsymbol{Y})$ in (152). Recalling that the remaining elements of the Hessian matrix $\mathsf{H}_{\boldsymbol{\lambda}} N(\boldsymbol{Y})$ are zero due to the presence of the matrix $\mathbf{Diag}(\boldsymbol{\sigma})$, in order to prove the negative semidefiniteness of $\mathsf{H}_{\boldsymbol{\lambda}} N(\boldsymbol{Y})$ it is sufficient to show that (153) is negative semidefinite.

Now, we apply Proposition H.9 to $\mathbf{A}(\boldsymbol{y})$, yielding

$$\mathbf{A}(\boldsymbol{y}) \circ \mathbf{A}(\boldsymbol{y}) \geq \frac{\mathbf{diag}(\mathbf{A}(\boldsymbol{y}))\mathbf{diag}(\mathbf{A}(\boldsymbol{y}))^{\mathsf{T}}}{p'}. \tag{154}$$

Taking the expectation in both sides of (154), we have

$$\begin{aligned}
&\mathsf{E}\{\mathbf{A}(\boldsymbol{Y}) \circ \mathbf{A}(\boldsymbol{Y})\} \\
&\quad \geq \frac{\mathsf{E}\{\mathbf{diag}(\mathbf{A}(\boldsymbol{Y}))\mathbf{diag}(\mathbf{A}(\boldsymbol{Y}))^{\mathsf{T}}\}}{p'}, \\
&\quad \geq \frac{\mathsf{E}\{\mathbf{diag}(\mathbf{A}(\boldsymbol{Y}))\}\mathsf{E}\{\mathbf{diag}(\mathbf{A}(\boldsymbol{Y}))^{\mathsf{T}}\}}{p'}. \\
&\quad = \frac{\mathbf{diag}(\mathsf{E}\{\mathbf{A}(\boldsymbol{Y})\})\mathbf{diag}(\mathsf{E}\{\mathbf{A}(\boldsymbol{Y})\})^{\mathsf{T}}}{p'} \\
&\quad \geq \frac{\mathbf{diag}(\mathsf{E}\{\mathbf{A}(\boldsymbol{Y})\})\mathbf{diag}(\mathsf{E}\{\mathbf{A}(\boldsymbol{Y})\})^{\mathsf{T}}}{n}
\end{aligned}$$

where in last inequality we have used that $p' = \mathsf{rank}(\mathbf{R}_{\mathbf{H}}) \leq \min\{n, p\} \leq n$, as $\mathbf{R}_{\mathbf{H}} = \mathbf{H}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{Z}}^{-1}\mathbf{H}$ and $\mathbf{H} \in \mathbb{R}^{n \times p}$. □

H. Matrix Algebra Results

In this Appendix we present a number of lemmas and propositions that are used throughout this paper.

Lemma H.1 (Bhatia [27, Lemma 1.3.6]): Let $\mathbf{R} \in \mathbb{S}_{+}^{s}$ be a positive semidefinite matrix, $\mathbf{R} \geq \mathbf{0}$. Then

$$\begin{bmatrix} \mathbf{R} & \mathbf{R} \\ \mathbf{R} & \mathbf{R} \end{bmatrix} \geq \mathbf{0}.$$

Proof: Since $\mathbf{R} \geq \mathbf{0}$, consider $\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathsf{T}}$ and write

$$\begin{bmatrix} \mathbf{R} & \mathbf{R} \\ \mathbf{R} & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A} \end{bmatrix} [\mathbf{A}^{\mathsf{T}} \quad \mathbf{A}^{\mathsf{T}}].$$ □

Lemma H.2 (Bhatia [27, Exercise 1.3.10]): Let $\mathbf{R} \in \mathbb{S}_{++}^{s}$ be a positive definite matrix, $\mathbf{R} > \mathbf{0}$. Then

$$\begin{bmatrix} \mathbf{R} & \mathbf{I}_s \\ \mathbf{I}_s & \mathbf{R}^{-1} \end{bmatrix} \geq \mathbf{0}. \tag{155}$$

Proof: Consider again $\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathsf{T}}$, then we have $\mathbf{R}^{-1} = \mathbf{A}^{-\mathsf{T}}\mathbf{A}^{-1}$. Now, simply write (155) as

$$\begin{bmatrix} \mathbf{R} & \mathbf{I}_s \\ \mathbf{I}_s & \mathbf{R}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathbf{I}_s & \mathbf{I}_s \\ \mathbf{I}_s & \mathbf{I}_s \end{bmatrix} \begin{bmatrix} \mathbf{A}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \end{bmatrix}$$

which, from Sylvester's law of inertia for congruent matrices [27, Ch. 1, Sec. 2] and Lemma H.1, is positive semidefinite. □

Lemma H.3: If the matrices $\mathbf{R}$ and $\mathbf{T}$ are positive (semi)definite, then so is the product $\mathbf{R} \otimes \mathbf{T}$. In other words, the class of positive (semi)definite matrices is closed under the Kronecker product.

Proof: See [28, Fact 7.4.15]. □

Corollary H.4 (Schur Theorem): The class of positive (semi)definite matrices is also closed under the Schur matrix product, $\mathbf{R} \circ \mathbf{T}$.

Proof: The proof follows from Lemma H.3 by noting that the Schur product $\mathbf{R} \circ \mathbf{T}$ is a principal submatrix of the Kronecker product $\mathbf{R} \otimes \mathbf{T}$ as in [28, Prp. 7.3.1] and that any principal submatrix of a positive (semi)definite matrix is also positive (semi)definite, [28, Prop. 8.2.6 and 8.2.7]. Alternatively, see [29, Theorem 7.5.3] or [25, Theorem 5.2.1] for a completely different proof. □

Lemma H.5 (Schur complement): Let the matrices $\mathbf{R} \in \mathbb{S}_{++}^{s}$ and $\mathbf{T} \in \mathbb{S}_{++}^{q}$ be positive definite, $\mathbf{R} > \mathbf{0}$ and $\mathbf{T} > \mathbf{0}$, and not necessarily of the same dimension. Then the following statements are equivalent

1) $\begin{bmatrix} \mathbf{R} & \mathbf{A} \\ \mathbf{A}^{\mathsf{T}} & \mathbf{T} \end{bmatrix} \geq 0$;
2) $\mathbf{T} \geq \mathbf{A}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{A}$;
3) $\mathbf{R} \geq \mathbf{A}\mathbf{T}^{-1}\mathbf{A}^{\mathsf{T}}$;

where $\mathbf{A} \in \mathbb{R}^{s \times r}$ is any arbitrary matrix.

Proof: See [29, Theorem 7.7.6] and the second exercise following it or [28, Prp. 8.2.3]. □

With the above lemmas at hand, we are now ready to prove the following proposition.

Proposition H.6: Consider two positive definite matrices $\mathbf{R} \in \mathbb{S}_{++}^{s}$ and $\mathbf{T} \in \mathbb{S}_{++}^{s}$ of the same dimension. Then it follows that

$$\mathbf{R} \circ \mathbf{T}^{-1} \geq \mathbf{Diag}(\mathbf{R})(\mathbf{R} \circ \mathbf{T})^{-1}\mathbf{Diag}(\mathbf{R}). \tag{156}$$

Proof: From Lemmas H.1, H.2, and H.4, it follows that

$$\begin{bmatrix} \mathbf{R} & \mathbf{R} \\ \mathbf{R} & \mathbf{R} \end{bmatrix} \circ \begin{bmatrix} \mathbf{T} & \mathbf{I}_s \\ \mathbf{I}_s & \mathbf{T}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{R} \circ \mathbf{T} & \mathbf{Diag}(\mathbf{R}) \\ \mathbf{Diag}(\mathbf{R}) & \mathbf{R} \circ \mathbf{T}^{-1} \end{bmatrix} \geq 0. \tag{157}$$

Now, from Lemma H.5, the result follows directly. □

Corollary H.7: Let $\mathbf{R} \in \mathbb{S}_{++}^{s}$ be a positive definite matrix. Then

$$\mathbf{diag}(\mathbf{R})^{\mathsf{T}}(\mathbf{R} \circ \mathbf{R})^{-1}\mathbf{diag}(\mathbf{R}) \leq s. \tag{158}$$

Proof: Particularizing the result in Proposition H.6 with $\mathbf{T} = \mathbf{R}$ and pre- and post-multiplying it by $\mathbf{1}^{\mathsf{T}}$ and $\mathbf{1}$ we obtain $\mathbf{1}^{\mathsf{T}}(\mathbf{R} \circ \mathbf{R}^{-1})\mathbf{1} \geq \mathbf{1}^{\mathsf{T}}\mathbf{Diag}(\mathbf{R})(\mathbf{R} \circ \mathbf{R})^{-1}\mathbf{Diag}(\mathbf{R})\mathbf{1}$. The result in (158) now follows straightforwardly from the fact $\mathbf{1}^{\mathsf{T}}(\mathbf{R} \circ \mathbf{R}^{-\mathsf{T}})\mathbf{1} = s$, [30] (see also [28, Fact 7.6.10], [25, Lemma 5.4. 2(a)]). Note that $\mathbf{R}$ is symmetric and thus $\mathbf{R}^{\mathsf{T}} = \mathbf{R}$ and $\mathbf{R}^{-\mathsf{T}} = \mathbf{R}^{-1}$. □

Remark H.8: Note that the proof of Corollary H.7 is based on the result of Proposition H.6 in (156). An alternative proof could follow from a different inequality by Styan in [31]

$$\mathbf{R} \circ \mathbf{R}^{-1} + \mathbf{I}_s \geq 2(\mathbf{R} \circ \mathbf{R})^{-1}$$

where, in this case, $\mathbf{R}$ is constrained to have ones in its main diagonal, i.e., $\mathbf{R} \circ \mathbf{I}_s = \mathbf{I}_s$.

Proposition H.9: Consider now the positive semidefinite matrix $\mathbf{R} \in \mathbb{S}_{+}^{s}$. Then

$$\mathbf{R} \circ \mathbf{R} \geq \frac{\mathbf{diag}(\mathbf{R})\mathbf{diag}(\mathbf{R})^{\mathsf{T}}}{s}.$$

Proof: For the case where $\mathbf{R} \in \mathbb{S}^s_{++}$ is positive definite, from (158) in Corollary H.7 and Lemma H.5, it follows that

$$\begin{bmatrix} \mathbf{R} \circ \mathbf{R} & \mathbf{diag}(\mathbf{R}) \\ \mathbf{diag}(\mathbf{R})^{\mathsf{T}} & s \end{bmatrix} \geq 0.$$

Applying again Lemma H.5, we get

$$\mathbf{R} \circ \mathbf{R} \geq \frac{\mathbf{diag}(\mathbf{R})\mathbf{diag}(\mathbf{R})^{\mathsf{T}}}{s}. \quad (159)$$

Now, assume that $\mathbf{R} \in \mathbb{S}^s_{+}$ is positive semidefinite. We thus define $\epsilon > 0$ and consider the positive definite matrix $\mathbf{R} + \epsilon \mathbf{I}_s$. From (159), we know that

$$(\mathbf{R} + \epsilon \mathbf{I}_s) \circ (\mathbf{R} + \epsilon \mathbf{I}_s) \geq \frac{\mathbf{diag}(\mathbf{R} + \epsilon \mathbf{I}_s)\mathbf{diag}(\mathbf{R} + \epsilon \mathbf{I}_s)^{\mathsf{T}}}{s}.$$

Taking the limit as ϵ tends to 0, the validity of (159) for positive semidefinite matrices follows from continuity. □

REFERENCES

[1] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, 2006.

[2] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. Control*, vol. 2, pp. 101–112, 1959.

[3] T. E. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.*, vol. 19, pp. 215–220, Jul. 1970.

[4] T. T. Kadota, M. Zakai, and J. Ziv, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 17, pp. 368–371, Jul. 1971.

[5] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[6] M. Zakai, "On mutual information, likelihood ratios, and estimation error for the additive Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3017–3024, Sep. 2005.

[7] D. P. Palomar and S. Verdú, "Representation of mutual information via input estimates," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 453–470, Feb. 2007.

[8] D. Guo, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, 2008, submitted for publication.

[9] D. Guo, S. Shamai, and S. Verdú, "Estimation of non-Gaussian random variables in Gaussian noise: Properties of the MMSE," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT'08)*, 2008, pp. 1083–1087.

[10] M. H. M. Costa, "A new entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 751–760, 1985.

[11] R. Tandon and S. Ulukus, "Dependence balance based outer bounds for Gaussian networks with cooperation and feedback," *IEEE Trans. Inf. Theory*, 2008, submitted for publication.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[13] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, pp. 1501–1518, 1991.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Hills, NJ: Prentice-Hall, 1993.

[15] J. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*, 3rd ed. New York: Wiley, 2007.

[16] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[17] A. Lozano, A. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[18] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J*, vol. 27, no. 3, pp. 379–423, 1948.

[19] A. Dembo, "Simple proof of the concavity of the entropy power with respect to added Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 887–888, 1989.

[20] C. Villani, "A short proof of the "concavity of entropy power"," *IEEE Trans. Inf. Theory* , vol. 46, no. 4, pp. 1695–1696, 2000.

[21] D. Guo, S. Shamai, and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT'06)*, Jul. 2006.

[22] O. Rioul, Information theoretic proofs of entropy power inequalities 2007, arXiv:0704.1751v1 [cs.IT].

[23] M. Payaró and D. P. Palomar, "A multivariate generalization of Costa's entropy power inequality," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT'08)*, Toronto, Canada, Jul. 2008, pp. 1088–1092.

[24] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[25] R. A. Horn and C. R. Johnson, *Topics on Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[26] G. Toscani, "Entropy production and the rate of convergence to equilibrium for the Fokker-Planck equation," *Quart. Appl. Math.*, vol. 57, no. 3, pp. 521–541, Sep. 1999.

[27] R. Bhatia, *Positive Definite Matrices*. Princeton, NJ: Princeton Univ. Press, 2007.

[28] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ: Princeton Univ. Press, 2005.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[30] C. R. Johnson and H. M. Shapiro, "Mathematical aspects of the relative gain array $A \circ A^{-T}$," *SIAM J. Algebr. Discrete Methods*, vol. 7, p. 627, 1986.

[31] G. P. H. Styan, "Hadamard products and multivariate statistical analysis," *Linear Algebra Appl.*, vol. 6, pp. 217–240, 1973.

Miquel Payaró (S'03–M'08) was born in Barcelona, Spain, in 1979. He received the degree in electrical engineering and the Ph.D. degree from the Universitat Politecnica de Catalunya (UPC), Barcelona, in 2002 and 2007, respectively. During his Ph.D. studies, he held a research appointment with the University of New South Wales, Sydney, Australia. He is a Physics undergraduate student with the Universitat de Barcelona.

He received a Predoctoral Grant from the Generalitat de Catalunya and from CTTCs Predoctoral fellowships program. From October 2002 to February 2007, he was involved in several research projects in the field of wireless communications, such as MARQUIS included in the European program EUREKA-MEDEA+ or NEWCOM financed by the Information Society Technologies program of the European Commission. From February 2007 to December 2008, he held a Postdoctoral position with the Electrical Engineering Department, Hong Kong University of Science and Technology, Hong Kong. Since January 2009, he has been a Research Associate with the Centre Tecnològic de Telecomunicacions de Catalunya.

Daniel P. Palomar (S'99–M'03–SM'08) received the electrical engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

Since 2006, he has been an Assistant Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong. He has held several research appointments, namely, with King's College London (KCL), London, U.K.; Technical University of Catalonia (UPC), Barcelona; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza," Rome, Italy; and Princeton University, Princeton, NJ.

Dr. Palomar is an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE 2010 special issue on "Convex Optimization for Signal Processing," was a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 special issue on "Game Theory in Communication Systems," as well as the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 special issue on "Optimization of MIMO Transceivers for Realistic Communication Networks." He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM). He received a 2004/2006 Fulbright Research Fellowship; the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society; the 2002/2003 best Ph.D. prize in information technologies and communications by the UPC; the 2002/2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation; and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.