Conditional Mean Estimation in Gaussian Noise: A Meta Derivative Identity With Applications

Alex Dytso, *Member, IEEE*, H. Vincent Poor, *Life Fellow, IEEE*, and Shlomo Shamai (Shitz), *Life Fellow, IEEE*

***Abstract*— Consider a channel $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ where $\mathbf{X}$ is an n-dimensional random vector, and $\mathbf{N}$ is a multivariate Gaussian vector with a full-rank covariance matrix $\mathsf{K}_\mathbf{N}$. The object under consideration in this paper is the conditional mean of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$, that is $\mathbf{y} \mapsto \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$. Several identities in the literature connect $\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$ to other quantities such as the conditional variance, score functions, and higher-order conditional moments. The objective of this paper is to provide a unifying view of these identities. In the first part of the paper, a general derivative identity for the conditional mean estimator is derived. Specifically, for the Markov chain $\mathbf{U} \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y}$, it is shown that the Jacobian matrix of $\mathbb{E}[\mathbf{U}|\mathbf{Y} = \mathbf{y}]$ is given by $\mathsf{K}_\mathbf{N}^{-1}\mathsf{Cov}(\mathbf{X}, \mathbf{U}|\mathbf{Y} = \mathbf{y})$ where $\mathsf{Cov}(\mathbf{X}, \mathbf{U}|\mathbf{Y} = \mathbf{y})$ is the conditional covariance. In the second part of the paper, via various choices of the random vector $\mathbf{U}$, the new identity is used to recover and generalize many of the known identities and derive some new identities. First, a simple proof of the Hatsel and Nolte identity for the conditional variance is shown. Second, a simple proof of the recursive identity due to Jaffer is provided. The Jaffer identity is then further explored, and several equivalent statements are derived, such as an identity for the higher-order conditional expectation (i.e., $\mathbb{E}[\mathbf{X}^k|\mathbf{Y}]$) in terms of the derivatives of the conditional expectation. Third, a new fundamental connection between the conditional cumulants and the conditional expectation is demonstrated. In particular, in the univariate case, it is shown that the k-th derivative of the conditional expectation is proportional to the $(k+1)$-th conditional cumulant. A similar expression is derived in the multivariate case.**

***Index Terms*— Vector Gaussian noise, conditional mean estimator, conditional cumulant, minimum mean squared error.**

Manuscript received 2 May 2021; revised 28 April 2022; accepted 6 October 2022. Date of publication 19 October 2022; date of current version 16 February 2023. This work was supported in part by the U.S. National Science Foundation under Grant CCF-1908308 and in part by the United States–Israel Binational Science Foundation under Grant BSF-2018710. An earlier version of this paper was presented in part at the 2020 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT44484.2020.9174071] and in part at the 2021 IEEE Information Theory Workshop [DOI: 10.1109/ITW46852.2021.9457595]. *(Corresponding author: Alex Dytso.)*

Alex Dytso is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: alex.dytso@njit.edu).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Shlomo Shamai (Shitz) is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 3200003, Israel (e-mail: sshlomo@ee.technion.ac.il).

Communicated by M. Lops, Associate Editor for Detection and Estimation.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3216012.

Digital Object Identifier 10.1109/TIT.2022.3216012

I. Introduction

CONSIDER a model given by the following input-output relationship:

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

where $\mathbf{N} \in \mathbb{R}^n$ is a zero mean, normally distributed with the covariance matrix $\mathsf{K}_\mathbf{N}$, and independent of $\mathbf{X} \in \mathbb{R}^n$. Throughout the paper $\mathsf{K}_\mathbf{N}$ is assumed to be a positive definite matrix, and we make no assumptions about the probability distribution of $\mathbf{X}$. In the case of $n = 1$, we denote $\mathsf{K}_\mathbf{N} = \sigma^2$. Also throughout the paper deterministic scalar quantities are denoted by lowercase letters, scalar random variables are denoted by uppercase letters, vectors are denoted by bold lowercase letters, random vectors by bold uppercase letters, and matrices by bold uppercase sans serif letters (e.g., x, X, $\mathbf{x}$, $\mathbf{X}$, $\mathbf{X}$).

In this work, we are interested in studying properties of the conditional mean estimator of the input $\mathbf{X}$ given the output $\mathbf{Y}$ according to (1), that is

$$\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] = \int \mathbf{x}\, \mathrm{d}P_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}),\ \mathbf{y} \in \mathbb{R}^n. \tag{2}$$

The conditional expectation is of interest in view of the wide range of applications of the conditional expectation in fields such as estimation theory and information theory. For example, the conditional expectation is known to be the unique optimal estimator under a very large family of loss functions, namely Bregman divergences [3]. In this work, we will view the conditional mean estimator as a function of channel realizations, that is $\mathbf{y} \mapsto \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$, and will be interested in characterizing analytical properties of the conditional expectation. Specifically, we focus on characterizing various derivative identities involving conditional expectations and will show a few applications of these identities.

There are several derivative identities in the literature that relate the conditional mean estimator to other quantities such as the score function and the conditional variance. Such identities are often used in information theory to give way to estimation theoretic arguments (e.g., the I-MMSE relationship [4], [5]). In estimation theory such identities are often used to design new estimation procedures (e.g., empirical Bayes [6], [7]) or establish connections to detection theory [8]. Perhaps the most well-known such identity is

$$\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] = \mathbf{y} + \mathsf{K}_\mathbf{N}\nabla_\mathbf{y} \log f_\mathbf{Y}(\mathbf{y}),\ \mathbf{y} \in \mathbb{R}^n \tag{3}$$

0018-9448 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

where $f_{\mathbf{Y}}(\mathbf{y})$ is the probability density function (pdf) of $\mathbf{Y}$. We note that the quantity $\nabla_{\mathbf{y}} \log f_{\mathbf{Y}}(\mathbf{y}) = \frac{\nabla_{\mathbf{y}} f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}$ is commonly known as the score function. The scalar version of the identity in (3) has been derived by Robbins in [9] where he credits Maurice Tweedie for the derivation. The vector version of the identity in (3) was derived by Esposito in [10]. Therefore, throughout this paper, we refer to the identity in (3) as Tweedie-Robbins-Esposito identity or TRE for short.

The observation that, via the TRE identity, the conditional expectation can be represented in terms of only the marginal distribution of the output $\mathbf{Y}$ has led to the development of the empirical Bayes procedure [9]; the interested reader is referred to [7] for an overview of this procedure. In addition to developing statistical procedures the TRE identity in (3) can considerably simplify the computation of $\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$ itself as we do not need to derive the conditional distribution $P_{\mathbf{X}|\mathbf{Y}}$ and only need to compute $f_{\mathbf{Y}}(\mathbf{y})$ and the gradient of $f_{\mathbf{Y}}(\mathbf{y})$. For an example of such an application, the interested reader is referred to [11] where the TRE identity was used to compute $\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$ for the case where $\mathbf{X}$ is uniform on a sphere in $\mathbb{R}^n$. In information theory, the TRE identity has also been used in the proofs of the scalar and vector versions of the I-MMSE relationship in [12] and [5], respectively.

In addition, to the TRE identity, first and higher-order derivatives of the conditional expectation also find a wide variety of applications. The first-order derivatives and gradients of the conditional expectation, which were first characterized by Hatsell and Nolte in [13], have been used to derive a vector version of the I-MMSE relationship in [5], prove Lipschitz continuity of the minimum mean squared error (MMSE) in [14], study sparse mixtures in [15], and derive converse bounds in network information theory [16], [17]. Higher-order derivatives of the conditional expectation have recently been used in [18] in the context of finding the best polynomial approximation of the conditional expectation.

The *first goal* of this work is to show that many of the known identities in the literature can be derived systematically from a single unifying derivative identity. The *second goal* is to show that the new identity leads to generalizations of the previously known identities and can be used to discover new identities.

A. Contributions and Paper Outline

The contribution and the outline of the paper are as follows:

- In Section II, Theorem 1 presents a new identity for the Jacobian of the conditional mean. Throughout the paper, this identity will be used for systematic proofs of old and new identities.
- In Section III, Proposition 1 presents a simple proof of a vector version of the Hatsel-Nolte identity, which relates the Jacobian of the conditional expectation to the conditional variance.
- In Section IV, we study recursive derivative identities for the conditional expectation and show:
  - In Section IV-A, Proposition 2 shows that the main identity in Theorem 1 can lead to a simple proof of a recursive identity due to Jaffer. To the best of our knowledge, Jaffer's identity is not well-known and until now has had no applications;
  - In Section IV-B, Proposition 3 provides an alternative integral version of Jaffer's identity. This new integral identity is shown to be very powerful and leads to simple proofs of old results and several new results. In particular, we have the following three new results. First, in Proposition 3, the integral version of Jaffer's identity leads to an expression for all higher-order derivatives of the conditional expectation in terms of Bell polynomials. Second, in Proposition 3, the integral version of Jaffer's identity leads to a representation of higher-order conditional expectations (i.e., $\mathbb{E}[X^k|Y]$) in terms of the derivatives of the conditional expectation. Third, in Proposition 4, the integral version of Jaffer's identity is used to generalize the TRE identity to higher-order conditional expectation. This generalized TRE identity maintains the property that $\mathbb{E}[X^k|Y]$ depends on the joint distribution only through the marginal of Y.
  - In Section IV-C, Proposition 5 and Proposition 6 provide two vector generalizations of Jaffer's identity.
- In Section V, we establish several new fundamental connections between the conditional expectations and the conditional cumulants and show:
  - In Section V-A, Proposition 7, for the univariate case, shows that the k-th derivative of the conditional expectation is proportional to the $(k+1)$-th conditional cumulant. Interestingly, Proposition 7, in combination with the TRE identity, shows that the conditional cumulants depend on the joint distribution only through the marginal of Y. Moreover, the combination of the TRE identity and Proposition 7 is used to study the properties of $P_{X|Y=y}$. In particular, it is shown that while $P_{X|Y=y}$ is sub-Gaussian, it can be strictly sub-Gaussian only on a set of measure zero. Finally, Theorem 2 establishes a new derivative identity that connects the higher-order derivative of the conditional expectation and the partial derivatives of the cumulant generating function; and
  - In Section V-B, the univariate results of Section V-A are generalized to the multivariate case. Specifically, Theorem 3 and Theorem 8 establish connections between the partial derivatives of the conditional cumulant generating function, the partial derivatives of the conditional cumulants, and the partial derivatives of the conditional expectation.
  - In Section V-C, Theorem 4 establishes a power series expansion of the conditional expecation.
- In Section VI, we study identities relating the conditional expectation to quantities such as the conditional distribution $P_{\mathbf{X}|\mathbf{Y}}$, the pdf of $\mathbf{Y}$, and the information density of the pair $(\mathbf{X}, \mathbf{Y})$ and show:
  - In Section VI-A, for the multivariate case, Proposition 10 and Proposition 12 establish the gradient

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

of the information density and the Hessian of the information density. In addition, for the univariate case, in Proposition 13, every k-th order derivative of the information density is determined.
  - In Section VI-B, the inverse version of the TRE identity is discussed and is used to show that the conditional expectation uniquely determines the input distribution provided the noise's covariance matrix is full rank.
  - In Section VI-C, the identities for the information density are used to find two new alternative representations of the MMSE.
- Section VII concludes the paper. All of the identities are summarized in Table I.

B. Notation

The set of all positive integers is denoted by $\mathbb{N}$, $[n]$ is the set of integers $\{1,\ldots,n\}$, and $\mathbb{R}^n$ is the set of all n-dimensional real-valued vectors. All logarithms in the paper are to the base e. All vectors in the paper are column vectors.

For random vectors $\mathbf{U}\in\mathbb{R}^m$, $\mathbf{X}\in\mathbb{R}^n$ and $\mathbf{Y}\in\mathbb{R}^k$ we define the *conditional variance matrix* and the *conditional cross-covariance matrix* as follows:

$$\mathsf{Var}(\mathbf{X}|\mathbf{Y})=\mathbb{E}\left[\mathbf{X}\mathbf{X}^{\mathsf{T}}|\mathbf{Y}\right]-\mathbb{E}\left[\mathbf{X}|\mathbf{Y}\right]\mathbb{E}\left[\mathbf{X}^{\mathsf{T}}|\mathbf{Y}\right], \quad (4)$$

$$\mathsf{Cov}(\mathbf{X},\mathbf{U}|\mathbf{Y})=\mathbb{E}\left[\mathbf{X}\mathbf{U}^{\mathsf{T}}|\mathbf{Y}\right]-\mathbb{E}\left[\mathbf{X}|\mathbf{Y}\right]\mathbb{E}\left[\mathbf{U}^{\mathsf{T}}|\mathbf{Y}\right]. \quad (5)$$

The *MMSE matrix* is denoted by

$$\mathsf{MMSE}(\mathbf{X}|\mathbf{Y})=\mathbb{E}\left[(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])^{\mathsf{T}}\right]. \quad (6)$$

As usual, we refer to the trace of the MMSE matrix as the MMSE and denote it by

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y})=\mathbb{E}\left[\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|^2\right]. \quad (7)$$

The *standard basis vectors* for $\mathbb{R}^n$ are denoted by $\mathbf{e}_i$, $i\in[n]$. For a matrix A, we use $[\mathsf{A}]_{ij}$ to denote the entry of row i and column j. The *Euclidian norm* of a vector $\mathbf{x}\in\mathbb{R}^n$ in this paper is denoted by $\|\mathbf{x}\|$. The *inner product* between vectors $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u}\cdot\mathbf{v}$.

The *gradient* of a function $f:\mathbb{R}^n\to\mathbb{R}$ is denoted by

$$\nabla_{\mathbf{x}}f(\mathbf{x})=\left[\frac{\partial f(\mathbf{x})}{\partial x_1},\ldots,\frac{\partial f(\mathbf{x})}{\partial x_n}\right]^{\mathsf{T}}\in\mathbb{R}^n. \quad (8)$$

The *Jacobian matrix* of a function $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ is denoted by

$$\mathsf{J}_{\mathbf{x}}\mathbf{f}(\mathbf{x})=\begin{pmatrix}\frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_1}\\ \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f_1(\mathbf{x})}{\partial x_n} & \frac{\partial f_2(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n}\end{pmatrix}\in\mathbb{R}^{n\times m}. \quad (9)$$

The *Hessian matrix* of a function $f:\mathbb{R}^n\to\mathbb{R}$ is denoted by

$$\mathsf{D}^2_{\mathbf{x}}\mathbf{f}(\mathbf{x})=\begin{pmatrix}\frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_n}\\ \frac{\partial^2 f_1(\mathbf{x})}{\partial x_2\partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2\partial x_n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial^2 f_1(\mathbf{x})}{\partial x_n\partial x_1} & \frac{\partial^2 f_2(\mathbf{x})}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f_m(\mathbf{x})}{\partial^2 x_n}\end{pmatrix}\in\mathbb{R}^{n\times n}. \quad (10)$$

The (n,k)-th *partial Bell polynomial* is denoted by $\mathsf{B}_{n,k}(x_1,\ldots,x_{n-k+1})$ and the n-th *complete Bell polynomial* is denoted by

$$\mathsf{B}_n(x_1,\ldots,x_n)=\sum_{k=1}^{n}\mathsf{B}_{n,k}(x_1,\ldots,x_{n-k+1}). \quad (11)$$

Finally, the pdf of a zero mean Gaussian random vector with a covariance matrix K is denoted by $\phi_{\mathsf{K}}(\cdot)$.

II. A New Identity for the Conditional Expectation

The first main result of this paper is the following general identity.

Theorem 1: Suppose that random vectors $\mathbf{U}\in\mathbb{R}^m,\mathbf{X}\in\mathbb{R}^n$, and $\mathbf{Y}\in\mathbb{R}^n$ satisfy the following conditions:

$$\mathbf{U}\leftrightarrow\mathbf{X}\leftrightarrow\mathbf{Y} \text{ form a Markov chain, in that order;} \quad (12)$$

$$\mathbb{E}\left[\|\mathbf{U}\|\|\mathbf{X}\|\,|\mathbf{Y}=\mathbf{y}\right]<\infty,\mathbf{y}\in\mathbb{R}^n;\text{and} \quad (13)$$

$$\mathbb{E}\left[\|\mathbf{U}\|\,|\mathbf{Y}=\mathbf{y}\right]<\infty,\mathbf{y}\in\mathbb{R}^n. \quad (14)$$

Then,

$$\mathsf{J}_{\mathbf{y}}\mathbb{E}\left[\mathbf{U}|\mathbf{Y}=\mathbf{y}\right]=\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Cov}(\mathbf{X},\mathbf{U}|\mathbf{Y}=\mathbf{y}),\ \mathbf{y}\in\mathbb{R}^n. \quad (15)$$

Proof: See Appendix A. ∎

It is not difficult to see that conditions in Theorem 1 are rather mild. For example, by using Bayes' formula, we have that

$$\mathbb{E}\left[\|\mathbf{U}\|\,|\mathbf{Y}=\mathbf{y}\right]=\frac{\mathbb{E}\left[\|\mathbf{U}\|\,\phi_{\mathsf{K}_{\mathbf{N}}}(\mathbf{y}-\mathbf{X})\right]}{f_{\mathbf{Y}}(\mathbf{y})}. \quad (16)$$

Consequently, since $f_{\mathbf{Y}}(\mathbf{y})>0$, $\mathbb{E}\left[\|\mathbf{U}\|\,|\mathbf{Y}=\mathbf{y}\right]<\infty$ if and only if $\mathbb{E}\left[\|\mathbf{U}\|\,\phi_{\mathsf{K}_{\mathbf{N}}}(\mathbf{y}-\mathbf{X})\right]<\infty$. Therefore, in order to violate the conditions in Theorem 1 one needs to find $\|\mathbf{U}\|$ that goes to infinity faster than a Gaussian density. In particular, by setting $\mathbf{U}=\mathbf{X}$, the above discussion shows that $\mathbb{E}\left[\mathbf{X}|\mathbf{Y}=\mathbf{y}\right]$ always exists.

In the rest of the paper, it is shown that many of the known identities in the literature can be derived systematically from the identity in Theorem 1. Moreover, we use this new identity to derive several generalizations of the previously known identities and discover some new identities. Specifically, this will be done by evaluating Theorem 1 with different choices of $\mathbf{U}$ such as $\mathbf{U}=\mathbf{X}$, $U=1_{\mathcal{A}}(\mathbf{X})$, $\mathbf{U}=(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{k-1}\mathbf{X}$, $k\in\mathbb{N}$ and $\mathbf{U}=\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}$, $\mathbf{t}\in\mathbb{R}^n$. For all these choices the conditional expectations will be finite.

III. The Variance Identity of Hatsell and Nolte

Our first application is to use (15) to recover a variance identity shown by Hatsell and Nolte in [13]. By setting $\mathbf{U}=\mathbf{X}$ in (15), we arrive at the following identity.

Proposition 1: For $\mathbf{y}\in\mathbb{R}^n$

$$\mathsf{J}_{\mathbf{y}}\mathbb{E}\left[\mathbf{X}|\mathbf{Y}=\mathbf{y}\right]=\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y}). \quad (17)$$

The identity in (17) has been first derived by Hatsell and Nolte in [13] for the case of $\mathsf{K}_{\mathbf{N}}=\mathsf{I}$. The general version in (17) was first derived in [5]. In terms of applications, in [5], the identity in (17) was used, together with the TRE identity in (3), to give a proof of the vector version of the I-MMSE

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

relationship; in [14], the scalar version of the identity in (17), was used to show that the minimum mean squared error is Lipschitz continuous with respect to the Wasserstein distance; and in [15], the identity in (17) was used to show log-convexity of the function akin to the log-likelihood ratio.

The identity of Hatsell and Nolte in (17) can be used to make various statements about the minimum and maximum 'slope' of $\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]$.

Corollary 1: For every $\mathbf{y} \in \mathbb{R}^n$

$$0 \leq \mathsf{Tr}\left(\mathbf{J}_{\mathbf{y}}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\right). \tag{18}$$

In addition, if $\|\mathbf{X}\| \leq R$, then

$$\mathsf{Tr}\left(\mathbf{J}_{\mathbf{y}}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\right) \leq R^2\mathsf{Tr}\left(\mathbf{K}_{\mathbf{N}}^{-1}\right). \tag{19}$$

Proof: The proof of the lower bound follows by using (17) together with the properties that both variance are positive definite matrices and that the trace of the product of two positive definite matrices is non-negative. To show the upper bound in (19), we use (17) together with the Cauchy-Schwarz inequality

$$\mathsf{Tr}\left(\mathbf{J}_{\mathbf{y}}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\right) = \mathsf{Tr}\left(\mathbf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y})\mathbf{K}_{\mathbf{N}}^{-1}\right) \tag{20}$$

$$\leq R^2\mathsf{Tr}\left(\mathbf{K}_{\mathbf{N}}^{-1}\right). \tag{21}$$

■

Several new applications of (17) will be given in subsequent sections. For example, in Section VI the identity in (17) will be used to show the concavity of the information density.

IV. Recursive Identities for Higher-Order Conditional Moments

In this section, we study recursive identities for higher-order conditional moments. The treatment of the scalar and vector case will be done separately. We begin by first showing a simple proof of the recursive identity due to Jaffer.

A. Jaffer's Identity

In [19], Jaffer has shown the following identity, which now easily follows from our main result in Theorem 1.

Proposition 2: For $y \in \mathbb{R}$ and $k \in \mathbb{N} \cup \{0\}$

$$\mathbb{E}\left[X^{k+1}|Y=y\right] = \sigma^2\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}\left[X^k|Y=y\right] + \mathbb{E}\left[X^k|Y=y\right]\mathbb{E}[X|Y=y]. \tag{22}$$

Proof: Letting $U = X^k$ in Theorem 1, we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}[X^k|Y=y] = \frac{1}{\sigma^2}\mathrm{Cov}(X, X^k|Y=y) \tag{23}$$

$$= \frac{1}{\sigma^2}\left(\mathbb{E}[X^{k+1}|Y=y] - \mathbb{E}[X^k|Y=y]\mathbb{E}[X|Y=y]\right). \tag{24}$$

This concludes the proof. ■

To the best of our knowledge, Jaffer's identity in (22) has had no applications and is not well-known. In what follows, we develop several alternative representations and generalizations of Jaffer's identity and also show the utility of Jaffar's identity. Specifically, we will first derive an alternative but equivalent integral version of Jaffer's identity and show how this new identity can be used to prove the uniqueness of the conditional mean estimator. In Section V, this identity will be used to show a new fundamental connection between conditional cumulants and conditional expectations.

B. A New Perspective on Jaffer's Identity

Next, we show that Jaffer's identity has an alternative integral version. This new integral representation leads to several interesting consequences. The following lemma will be useful in deriving this new representation.

Lemma 1: Let $f_k : \mathbb{R} \to \mathbb{R}$ be a sequence of functions with $k \in \mathbb{N} \cup \{0\}$ such that

$$f_k(x) = \frac{\mathrm{d}}{\mathrm{d}x}f_{k-1}(x) + f_{k-1}(x)f_1(x), k = 1, 2, \ldots, \tag{25}$$

with $f_0 \equiv 1$.

Then, the following statements hold:

- The solution to (25) is given by

$$f_k(x) = \mathrm{e}^{-\int_0^x f_1(t)\mathrm{d}t}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t}, \tag{26}$$

$$= \mathsf{B}_k\left(f_1^{(0)}(x), \ldots, f_1^{(k-1)}(x)\right). \tag{27}$$

- The derivatives of f_1 are given by

$$f_1^{(k)}(x) = \sum_{m=1}^{k+1} c_m \mathsf{B}_{k+1,m}\left(f_1(x), \ldots, f_{k-m+2}(x)\right), \tag{28}$$

where $c_m = (-1)^{m-1}(m-1)!$.

Proof: See Appendix B. ■

Using Lemma 1, we can now present an alternative integral version of Jaffer's identity. In addition, we also provide an expression for all higher-order derivatives of the conditional expectation.

Proposition 3: For $y \in \mathbb{R}$ and $k \in \mathbb{N}$

$$\mathbb{E}\left[X^k|Y=y\right] = \sigma^{2k}\mathrm{e}^{-\frac{1}{\sigma^2}\int_0^y \mathbb{E}[X|Y=t]\mathrm{d}t}\frac{\mathrm{d}^k}{\mathrm{d}y^k}\mathrm{e}^{\frac{1}{\sigma^2}\int_0^y \mathbb{E}[X|Y=t]\mathrm{d}t} \tag{29}$$

$$= \sigma^{2k}\mathsf{B}_k\left(\mathbb{E}^{(0)}\left[\frac{X}{\sigma^2}|Y=y\right], \ldots, \mathbb{E}^{(k-1)}\left[\frac{X}{\sigma^2}|Y=y\right]\right), \tag{30}$$

where $\mathbb{E}^{(k)}[X|Y=y] = \frac{\mathrm{d}^k}{\mathrm{d}y^k}\mathbb{E}[X|Y=y]$. Moreover,

$$\frac{\mathrm{d}^k}{\mathrm{d}y^k}\mathbb{E}[X|Y=y] = \sigma^2\sum_{m=1}^{k+1} c_m \cdot \mathsf{B}_{k+1,m}\left(\mathbb{E}\left[\frac{X}{\sigma^2}|Y=y\right], \ldots, \mathbb{E}\left[\left(\frac{X}{\sigma^2}\right)^{k-m+2}|Y=y\right]\right), \tag{31}$$

where $c_m = (-1)^{m-1}(m-1)!$.

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

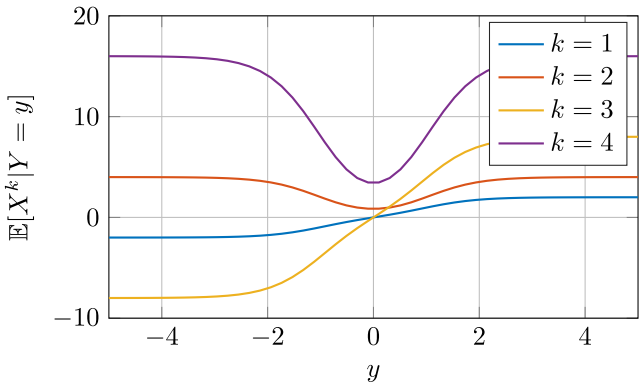

Fig. 1. Plot of $\mathbb{E}[X^k|Y=y]$ vs. y for $k=1,2,3$ and 4.

Proof: First, observe that Jaffer's identity in (22) can be re-written as

$$\mathbb{E}\left[\left(\frac{X}{\sigma^2}\right)^k|Y=y\right]=\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}\left[\left(\frac{X}{\sigma^2}\right)^{k-1}|Y=y\right]+\mathbb{E}\left[\left(\frac{X}{\sigma^2}\right)^{k-1}|Y=y\right]\mathbb{E}\left[\left(\frac{X}{\sigma^2}\right)|Y=y\right]. \quad (32)$$

Hence, if we take $f_k(y)=\mathbb{E}\left[\left(\frac{X}{\sigma^2}\right)^k|Y=y\right]$, then the Jaffer's identity is of the same the form as the recurrence relationship in (25). In view of this observation, the proof now follows by applying the results in Lemma 1. ∎

Example: The second and third Bell polynomials are given by

$$\mathsf{B}_2(x_1,x_2)=x_1^2+x_2, \quad (33)$$

$$\mathsf{B}_3(x_1,x_2,x_3)=x_1^2+3x_1x_2+x_3. \quad (34)$$

Therefore, using (30), we have that

$$\mathbb{E}\left[X^2|Y=y\right]=\mathbb{E}^2[X|Y=y]+\sigma^2\mathbb{E}^{(1)}[X|Y=y], \quad (35)$$

and

$$\mathbb{E}\left[X^3|Y=y\right]=\sigma^2\mathbb{E}^2[X|Y=y]+3\sigma^2\mathbb{E}[X|Y=y]\mathbb{E}^{(1)}[X|Y=y]+\sigma^4\mathbb{E}^{(2)}[X|Y=y]. \quad (36)$$

As an illustration of this procedure the k-th order conditional expectation for $\sigma^2=1$ of a random variable X uniformly distributed on $\{-2,0,2\}$ is plotted in Fig. 1.

An important feature of the integral version of Jaffer's identity in (29) is that every higher-order conditional moment is determined by the first-order conditional moment.

Another identity, equivalent to that in (29) and which allows expressing higher-order conditional moments in terms of the pdf of Y, is shown next.

Proposition 4: For any $k\in\mathbb{N}$ and $y\in\mathbb{R}$

$$\mathbb{E}[X^k|Y=y]=\sigma^{2k}\frac{\frac{\mathrm{d}^k}{\mathrm{d}y^k}\left(f_Y(y)\phi^{-1}_{\sigma^2}(y)\right)}{f_Y(y)\phi^{-1}_{\sigma^2}(y)}. \quad (37)$$

Alternatively, let $t\mapsto H_{e_m}(t)$, $m\in\mathbb{N}\cup\{0\}$ be a *probabilistic Hermite polynomial*; then

$$\mathbb{E}[X^k|Y=y]=\sigma^{2k}\frac{\sum_{m=0}^k\binom{k}{m}f_Y^{(k-m)}(y)\frac{(-i)^m}{\sigma^m}H_{e_m}\left(i\frac{y}{\sigma}\right)}{f_Y(y)}, \quad (38)$$

where $i=\sqrt{-1}$.

Proof: First, observe that using the scalar TRE identity in (3) we have that

$$\int_0^y\frac{\mathbb{E}\left[X|Y=t\right]}{\sigma^2}\mathrm{d}t=\int_0^y\left(\frac{t}{\sigma^2}+\frac{\mathrm{d}}{\mathrm{d}t}\log(f_Y(t))\right)\mathrm{d}t \quad (39)$$

$$=\frac{y^2}{2\sigma^2}+\log(f_Y(y))-\log(f_Y(0)). \quad (40)$$

Inserting (40) into (29) leads to (37). The proof of (38) follows by applying the generalized product rule to (37) and the following derivative [20, eq. 19.13.3]:

$$\frac{\mathrm{d}^m}{\mathrm{d}y^m}\mathrm{e}^{\frac{x^2}{2\sigma}}=\frac{(-i)^m}{\sigma^m}H_{e_m}\left(i\frac{y}{\sigma}\right)\mathrm{e}^{\frac{x^2}{2\sigma^2}}. \quad (41)$$

∎

The identity in (37) can be thought of as a generalization of the TRE identity in (3) to the higher-order moments. Indeed for $k=1$, we recover the TRE identity. Similarly to the TRE identity, the important feature of the identity in (37) is that $\mathbb{E}[X^k|Y]$ depends on the joint distribution $P_{X,Y}$ only through the marginal pdf of Y.

C. Vector Generalizations of Jaffer's Identity

Given the fact that there is no unique generalization of higher-order moments to the vector case, several vector generalizations of the identity in (22) are possible. Next, we present two such generalizations.

The first generalization of (22) is in terms of the powers of a matrix.

Proposition 5: For $k\in\mathbb{N}$ and $\mathbf{y}\in\mathbb{R}^n$

$$\mathbb{E}\left[(\mathbf{X}\mathbf{X}^\mathsf{T})^k|\mathbf{Y}=\mathbf{y}\right]=\mathbf{K_N}\mathbf{J_y}\mathbb{E}\left[(\mathbf{X}\mathbf{X}^\mathsf{T})^{k-1}\mathbf{X}|\mathbf{Y}=\mathbf{y}\right]+\mathbb{E}\left[\mathbf{X}|\mathbf{Y}=\mathbf{y}\right]\mathbb{E}\left[\mathbf{X}^\mathsf{T}(\mathbf{X}\mathbf{X}^\mathsf{T})^{k-1}|\mathbf{Y}=\mathbf{y}\right]. \quad (42)$$

Proof: The proof follows by evaluating (15) with $\mathbf{U}=(\mathbf{X}\mathbf{X}^\mathsf{T})^{k-1}\mathbf{X}$ and noting that $\mathbf{U}^\mathsf{T}=\mathbf{X}^\mathsf{T}(\mathbf{X}\mathbf{X}^\mathsf{T})^{k-1}$. ∎

The second generalization of (22) allows for different exponents across elements of $\mathbf{X}$.

Proposition 6: For every $m\in[n]$, $v_i\in\mathbb{N}\cup\{0\}$, $i\in[n]$ and $\mathbf{y}\in\mathbb{R}^n$

$$\frac{\mathrm{d}}{\mathrm{d}y_m}\mathbb{E}\left[\prod_{i=1}^n\left(\mathbf{e}_i^\mathsf{T}\mathbf{K}_\mathbf{N}^{-1}\mathbf{X}\right)^{v_i}|\mathbf{Y}=\mathbf{y}\right]=\mathbb{E}\left[\prod_{i=1:i\neq m}^n\left(\mathbf{e}_i^\mathsf{T}\mathbf{K}_\mathbf{N}^{-1}\mathbf{X}\right)^{v_i}\left(\mathbf{e}_m^\mathsf{T}\mathbf{K}_\mathbf{N}^{-1}\mathbf{X}\right)^{v_m+1}|\mathbf{Y}=\mathbf{y}\right]-\mathbb{E}\left[\prod_{i=1}^n\left(\mathbf{e}_i^\mathsf{T}\mathbf{K}_\mathbf{N}^{-1}\mathbf{X}\right)^{v_i}|\mathbf{Y}=\mathbf{y}\right]\mathbb{E}\left[\mathbf{e}_m^\mathsf{T}\mathbf{K}_\mathbf{N}^{-1}\mathbf{X}|\mathbf{Y}=\mathbf{y}\right]. \quad (43)$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

Proof: The proof follows by evaluating (15) with $U = \prod_{i=1}^{n} \left(\mathrm{e}_i^{\mathsf{T}} \mathbf{K}_{\mathbf{N}}^{-1} \mathbf{X}\right)^{v_i}$. ∎

In the case when $\mathbf{K_N}$ is a diagonal matrix with $[\mathbf{K_N}]_{ii} = \sigma_{ii}^2$ the identity (43) reduces to

$$\mathbb{E}\left[X_m^{v_m+1} \prod_{i=1:i\neq m}^{n} X_i^{v_i} | \mathbf{Y} = \mathbf{y}\right] = \sigma_{mm}^2 \frac{\mathrm{d}}{\mathrm{d}y_m} \mathbb{E}\left[\prod_{i=1}^{n} X_i^{v_i} | \mathbf{Y} = \mathbf{y}\right] + \mathbb{E}\left[\prod_{i=1}^{n} X_i^{v_i} | \mathbf{Y} = \mathbf{y}\right] \mathbb{E}[X_m | \mathbf{Y} = \mathbf{y}]\,. \tag{44}$$

V. Identities for the Conditional Cumulants

This section establishes a new connection between the conditional cumulants and the conditional expectation. For ease of exposition, we first focus on the univariate case and then generalize the results to the multivariate case.

A. The Univariate Case

Consider the *conditional cumulant generating function*

$$K_X(t|Y=y) = \log\left(\mathbb{E}[\mathrm{e}^{tX}|Y=y]\right),\ y \in \mathbb{R}, t \in \mathbb{R}. \tag{45}$$

The k-th conditional cumulant is given by

$$\kappa_{X|Y=y}(k) = \frac{\mathrm{d}^k}{\mathrm{d}t^k} K_X(t|Y=y)\Big|_{t=0},\ k \in \mathbb{N},\ t \in \mathbb{R}. \tag{46}$$

Remark 1: The conditional moment generating (i.e., $\mathbb{E}[\mathrm{e}^{tX}|Y=y]$) is well-defined in view of (16). An alternative way to argue this is to use the fact all $t \in \mathbb{R}$, the conditional distribution $P_{X|Y=y}$ is sub-Gaussian [21].

It is well-known that the cumulants and the moments of a random variable U have a one-to-one correspondence with the inverse relationship given by

$$\kappa_U(k) = \sum_{m=1}^{k} c_m \mathsf{B}_{k,m}\left(\mu_1, \ldots, \mu_{k-m+1}\right), \tag{47}$$

where $\mu_m = \mathbb{E}[U^m]$ [22, Example 11.4]. This expression together with the integral version of Jaffer's identity in (31) leads to the following simple relationship between the conditional expectation and the conditional cumulants.

Proposition 7: For $y \in \mathbb{R}$ and $k \in \mathbb{N} \cup \{0\}$

$$\sigma^{2k} \frac{\mathrm{d}^k}{\mathrm{d}y^k} \mathbb{E}[X|Y=y] = \kappa_{X|Y=y}(k+1). \tag{48}$$

Proof: First, let $X_y \sim P_{X|Y=y}$ and let $U = \frac{X_y}{\sigma^2}$. Second, by using the scaling property of cumulants, we have that

$$\kappa_U(k) = \frac{1}{\sigma^{2k}} \kappa_{X|Y=y}(k). \tag{49}$$

Next, by using (47), for $k \geq 1$ we have that

$$\frac{1}{\sigma^{2k}} \kappa_{X|Y=y}(k) = \sum_{m=1}^{k} c_m \mathsf{B}_{k,m}\left(\mathbb{E}\left[\frac{X_y}{\sigma^2}\right], \ldots, \mathbb{E}\left[\left(\frac{X_y}{\sigma^2}\right)^{k-m+1}\right]\right) \tag{50}$$

$$= \frac{1}{\sigma^2} \frac{\mathrm{d}^{k-1}}{\mathrm{d}y^{k-1}} \mathbb{E}[X_y]\,, \tag{51}$$

where the last step follows by using (31). This concludes the proof. ∎

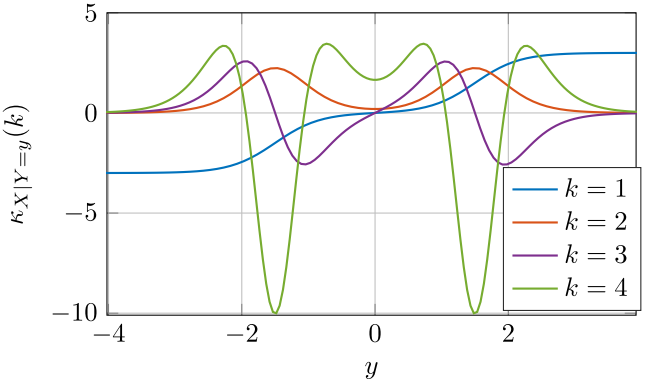

Fig. 2. Plot of $\kappa_{X|Y=y}(k)$ vs. y for $k = 1, 2, 3$ and 4.

From Proposition 7, we make the following two observations:

- For $k \in \mathbb{N}$

$$\sigma^2 \frac{\mathrm{d}}{\mathrm{d}y} \kappa_{X|Y=y}(k) = \kappa_{X|Y=y}(k+1);\ \text{and} \tag{52}$$

- By using the TRE identity, we arrive at a new the representation of cumulants in terms of only f_Y

$$\kappa_{X|Y=y}(1) = y + \sigma^2 \frac{\mathrm{d}}{\mathrm{d}y} \log f_Y(y), \tag{53}$$

$$\kappa_{X|Y=y}(2) = \sigma^2 + \sigma^4 \frac{\mathrm{d}^2}{\mathrm{d}y^2} \log f_Y(y), \tag{54}$$

$$\kappa_{X|Y=y}(k) = \sigma^{2k} \frac{\mathrm{d}^k}{\mathrm{d}y^k} \log f_Y(y), k \geq 3. \tag{55}$$

  In other words, the conditional cumulants depend on the joint distribution P_{XY} only through the marginal distribution P_Y.

Example: In the case when X is standard Gaussian the conditional expectation $\mathbb{E}[X|Y=y]$ is a linear function of y. Therefore, by using (48), we have that

$$\kappa_{X|Y=y}(1) = \frac{1}{1+\sigma^2} y, \tag{56}$$

$$\kappa_{X|Y=y}(2) = \frac{\sigma^2}{1+\sigma^2}, \tag{57}$$

$$\kappa_{X|Y=y}(k) = 0,\ k \geq 3. \tag{58}$$

Note that this is as expected since $P_{X|Y}$ is Gaussian, and for the Gaussian distribution only the first and the second cumulants are non-zero.

Example: Consider an example of a random variable $X \in \{-3, 0, 3\}$ with a uniform distribution. Fig. 2 shows plots of $\kappa_{X|Y=y}$ vs. y for several values of k.

One more example of the expression for the conditional cumulants will be given in Section V-B for the case of X distributed uniformly on $\{-R, R\}$.

We next show a small application of the new identity in (48).

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

Example: A random variable U with mean $\mu = \mathbb{E}[U]$ is said to have a *sub-Gaussian* distribution if there exists a γ^2 such that

$$\mathbb{E}\left[\mathrm{e}^{\lambda(U-\mu)}\right] \leq \mathrm{e}^{\frac{\lambda^2\gamma^2}{2}},\ \lambda \in \mathbb{R}. \tag{59}$$

The quantity γ^2 is known as *a proxy variance*. The distribution is said to be *strictly* sub-Gaussian if $\mathrm{Var}(U) = \gamma$. In [21], it was shown that for an arbitrary distribution of X, the conditional distribution $P_{X|Y=y}$ is sub-Gaussian for all y. We now use the identity in (55) to answer if $P_{X|Y=y}$ is strictly sub-Gaussian. As shown in [23, Prop. 4.3] a necessary condition for strict sub-Gaussianity requires that the third cumulant is zero. Therefore, a necessary condition for $P_{X|Y=y}$ to be strictly sub-Gaussian is

$$0 = \kappa_{X|Y=y}(3),\ \forall y \in \mathbb{R}. \tag{60}$$

This certainly holds if X is Gaussian. Moreover, as shown in Fig. 2, for a non-Gaussian example, the above can be zero for some values of y. However, this set of y's must necessarily be of Lebesgue measure zero for a non-Gaussian X. To see this, suppose that $0 = \kappa_{X|Y=y}(3), y \in \mathcal{S}$ where $\mathcal{S}$ is a set of positive Lebesgue measure. Then, since $\kappa_{X|Y=y}(k)$ is real-analytic (see Lemma 2 below in Section V-C), by the identity theorem for real-analytic functions [24], it follows that $0 = \kappa_{X|Y=y}(3)$ for all $y \in \mathbb{R}$. Next, using this and the identity in (55), we arrive at a differential equation

$$0 = \kappa_{X|Y=y}(3) = \sigma^6 \frac{\mathrm{d}^3}{\mathrm{d}y^3} \log f_Y(y), y \in \mathbb{R}. \tag{61}$$

The solution to (61) states that $\log f_Y(y)$ must be a quadratic function and implies that Y must be Gaussian. This further implies that X is Gaussian and contradicts our assumption that X is non-Gaussian. In conclusion, *for a non-Gaussian X, the distribution $P_{X|Y=y}$ can only be strictly sub-Gaussian on a set of Lebesque measure zero.*

We conclude this section by showing that Theorem 1 can be used to get a more general identity than that in (48).

Theorem 2: For $y \in \mathbb{R}$ and $k \in \mathbb{N} \cup \{0\}$

$$\frac{\mathrm{d}^{k+1}}{\mathrm{d}t^{k+1}} K_X(t|Y=y) = \sigma^{2(k+1)} \frac{\mathrm{d}^{k+1}}{\mathrm{d}y^{k+1}} K_X(t|Y=y) + \sigma^{2k} \frac{\mathrm{d}^k}{\mathrm{d}y^k} \mathbb{E}[X|Y=y]. \tag{62}$$

Proof: First, consider the case of $k = 0$. By setting $U = \mathrm{e}^{tX}$, $t \in \mathbb{R}$ in Theorem 1, we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}y} K_X(t|Y=y) = \frac{\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}[\mathrm{e}^{tX}|Y=y]}{\mathbb{E}[\mathrm{e}^{tX}|Y=y]} \tag{63}$$

$$= \frac{1}{\sigma^2} \frac{\mathbb{E}[X\mathrm{e}^{tX}|Y=y] - \mathbb{E}[\mathrm{e}^{tX}|Y=y]\mathbb{E}[X|Y=y]}{\mathbb{E}[\mathrm{e}^{tX}|Y=y]} \tag{64}$$

$$= \frac{1}{\sigma^2} \frac{\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\mathrm{e}^{tX}|Y=y] - \mathbb{E}[\mathrm{e}^{tX}|Y=y]\mathbb{E}[X|Y=y]}{\mathbb{E}[\mathrm{e}^{tX}|Y=y]} \tag{65}$$

$$= \frac{1}{\sigma^2} \left(\frac{\mathrm{d}}{\mathrm{d}t} \log(\mathbb{E}[\mathrm{e}^{tX}|Y=y]) - \mathbb{E}[X|Y=y] \right) \tag{66}$$

$$= \frac{1}{\sigma^2} \left(\frac{\mathrm{d}}{\mathrm{d}t} K_X(t|Y=y) - \mathbb{E}[X|Y=y] \right). \tag{67}$$

The rest of the proof follows by using (67) together with a simple induction. ∎

Alternatively, we could have used Theorem 2 to show the derivative identity in (48). This approach has the benefit of avoiding the use of Bell polynomials. This alternative view will be taken in the next section to derive a multivariate generalization of the identity in (48).

B. The Multivariate Case

Consider the *multivariate conditional cumulant generating function*

$$K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y}) = \log\left(\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]\right),\ \mathbf{y} \in \mathbb{R}^n, \mathbf{t} \in \mathbb{R}^n. \tag{68}$$

The conditional cumulants are now defined as

$$\kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1, \ldots, s_j) = \frac{\partial^j}{\partial t_{s_1} \ldots \partial t_{s_j}} K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})\big|_{\mathbf{t}=\mathbf{0}}, \tag{69}$$

where $j \in \mathbb{N}$ and $s_1, \ldots, s_j \in [n]$. Note that the cumulants are the same for all permutations of the sequence $s_1, \ldots, s_j$. The above definitions follow the conventions of [25].

It is instructive to consider the following example.

Example: Let $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathsf{I})$. Then, $\mathbf{X}|\mathbf{Y} = \mathbf{y} \sim \mathcal{N}\left(\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}], \mathbf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y})\right)$ where

$$\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}] = (\mathsf{I} + \mathbf{K_N})^{-1}\mathbf{y}, \mathbf{y} \in \mathbb{R}^n, \tag{70}$$

$$\mathbf{\Sigma} = \mathbf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y}) = \mathbf{K_N}(\mathsf{I} + \mathbf{K_N})^{-1},\ \mathbf{y} \in \mathbb{R}^n, \tag{71}$$

with the moment generating function given by

$$\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}] = \mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]}\mathrm{e}^{\frac{\mathbf{t}^\mathsf{T}\mathbf{\Sigma}\mathbf{t}}{2}}, \mathbf{t} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n, \tag{72}$$

and the cumulant generating function given by

$$K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y}) = \mathbf{t}^\mathsf{T}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}] + \frac{\mathbf{t}^\mathsf{T}\mathbf{\Sigma}\mathbf{t}}{2}, \mathbf{t} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n. \tag{73}$$

Using the definition in (69), the cumulants are calculated to be

$$\begin{aligned} &j = 1:\ \kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1) = \mathbb{E}[X_{s_1}|\mathbf{Y}=\mathbf{y}],\ s_1 \in [n], \\ &j = 2:\ \kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1, s_2) = [\mathbf{\Sigma}]_{s_1,s_2},\ s_1, s_2 \in [n], \\ &j \geq 3:\ \kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1, \ldots, s_j) = 0,\ s_1, \ldots, s_j \in [n]. \end{aligned}$$

We first show a multivariate generalization of Theorem 2, which follows by letting $U = \mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}$ in Theorem 1.

Theorem 3: Let $\mathbf{k}_i^\mathsf{T} = \mathbf{e}_i^\mathsf{T}\mathbf{K_N}^{-1}$ (i.e., the i-th row of $\mathbf{K_N}^{-1}$). Then, for $j \in \mathbb{N}$ and $\mathbf{y} \in \mathbb{R}^n$

$$\frac{\partial^j K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})}{\partial y_{s_1} \ldots \partial y_{s_j}} = \mathbf{k}_{s_1}^\mathsf{T} \nabla_{\mathbf{t}} \mathbf{k}_{s_2}^\mathsf{T} \nabla_{\mathbf{t}} \ldots \mathbf{k}_{s_j}^\mathsf{T} \nabla_{\mathbf{t}} K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y}) - \mathbf{k}_{s_j}^\mathsf{T} \frac{\partial^{j-1}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]}{\partial y_{s_1} \ldots \partial y_{s_{j-1}}}. \tag{74}$$

Proof: See Appendix E. ∎

Remark 2: The first term on the right side of (74) can be equivalently written as

$$\mathbf{k}_{s_1}^\mathsf{T} \nabla_{\mathbf{t}} \mathbf{k}_{s_2}^\mathsf{T} \nabla_{\mathbf{t}} \ldots \mathbf{k}_{s_j}^\mathsf{T} \nabla_{\mathbf{t}} K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

$$= \sum_{p_1=1}^{n} \cdots \sum_{p_j=1}^{n} \prod_{i=1}^{j} k_{s_i,p_i} \frac{\partial^j K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})}{\partial t_{p_1} \ldots \partial t_{p_j}}, \tag{75}$$

where k_{s_j,p_j} is p_j's entry of $\mathbf{k}_{s_j}$. This representation will be useful in the next proof, which relates the conditional cumulants and the conditional moments.

The next result generalizes the derivative identity in (48) to the multivariate case.

Proposition 8: Let $\mathbf{k}_m^{\mathsf{T}} = \mathbf{e}_m^{\mathsf{T}} \mathbf{K}_{\mathbf{N}}^{-1}$ and $k_{m,i} = \mathbf{e}_m^{\mathsf{T}} \mathbf{K}_{\mathbf{N}}^{-1} \mathbf{e}_i$. Then, for $j \in \mathbb{N}$ and $\mathbf{y} \in \mathbb{R}^n$

$$\mathbf{k}_{s_j}^{\mathsf{T}} \frac{\partial^{j-1} \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]}{\partial y_{s_1} \ldots \partial y_{s_{j-1}}} = \sum_{p_1=1}^{n} \cdots \sum_{p_j=1}^{n} \prod_{i=1}^{j} k_{s_i,p_i} \kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(p_1, \ldots, p_j). \tag{76}$$

Proof: The proof follows by showing that for every j we have that $\frac{\partial^j K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})}{\partial y_{s_1} \ldots \partial y_{s_j}}|_{\mathbf{t}=\mathbf{0}} = 0$. See Appendix F for the details. ∎

In the case when $\mathbf{K}_{\mathbf{N}}$ is a diagonal matrix the above simplifies to

$$\kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1, \ldots, s_j) = \left(\prod_{i=1}^{j-1} \sigma_{s_i,s_i}^2 \right) \frac{\partial^{j-1} \mathbb{E}[X_{s_j}|\mathbf{Y}=\mathbf{y}]}{\partial y_{s_1} \ldots \partial y_{s_{j-1}}}, \tag{77}$$

where we let $\prod_{i=1}^{0} \sigma_{s_i,s_i}^2 = 1$.

Closed-form expressions for the conditional expectation are rare in the univariate case and even more so in the multivariate case. In particular, not many examples of $\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]$ are known when $\mathbf{X}$ has a non-product distribution. The next example computes the first two cumulants for one of the rare cases when we do have a closed-form expression for the conditional expectation.

Example: Consider a case when $\mathbf{X}$ is distributed uniformly on $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = R\}$ (i.e., $(n-1)$-sphere of radius R) and let $\mathbf{K}_{\mathbf{N}} = \mathsf{I}$. This distribution has several applications in information theory and estimation theory. For example, in information theory, this distribution is the capacity-achieving distribution for an amplitude-constrained channel [11]. In estimation theory, this distribution has been shown to be the least favorable distribution for the problem of estimating a bounded normal mean [26]. The conditional expectation for this distribution is given by [11]

$$\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}] = \frac{R\mathbf{y}}{\|\mathbf{y}\|} h_{\frac{n}{2}}(R\|\mathbf{y}\|), \, \mathbf{y} \in \mathbb{R}^n, \tag{78}$$

where $h_\nu(t) = \frac{I_\nu(t)}{I_{\nu-1}(t)}$ and where $I_\nu(\cdot)$ is the modified Bessel function of the first kind of order ν. Next, using Proposition 8, we characterize the first two conditional cumulants of this distribution. For $j=1$, $\kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1) = \mathbb{E}[X_{s_2}|\mathbf{Y}=\mathbf{y}]$, and for $j=2$ the expression for $\kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1,s_2)$ is given in (79), shown at the bottom of the page. The proofs for $j=2$ can be found in Appendix D. For the case of $n=1$, the distribution under consideration becomes uniform on $\{-R,R\}$ and as expected the expression in (79) reduces to $k_{X|Y=y}(2) = \left(\frac{R}{\cosh(Ry)}\right)^2$, which is the derivative of $\mathbb{E}[X|Y=y] = R\tanh(Ry)$.

C. Power Series Expansion of the Conditional Expectation

In this section, we find the power series expansion of the conditional expectation in terms of the conditional cumulants. The fact that a power series expansion exists follows from the next result.

Lemma 2: The functions $y \mapsto \mathbb{E}[X^k|Y=y], k \in \mathbb{N}$ and $y \mapsto \kappa_{X|Y=y}(k), k \in \mathbb{N}$ are real-analytic.

Proof: Note that by the TRE identity in (3)

$$\mathbb{E}[X|Y=y] = y + \sigma^2 \frac{\frac{\mathrm{d}}{\mathrm{d}y} f_Y(y)}{f_Y(y)}. \tag{80}$$

Hence, since the ratios and sums of analytic functions are analytic, $\mathbb{E}[X|Y=y]$ is real-analytic provided that $f_Y(y)$ is real-analytic. The analyticity of f_Y is a known consequence of convolution with Gaussian measures (see e.g., [27, p. 242]). Since $\mathbb{E}[X|Y=y]$ is real-analytic, the identities in (30) and (48) imply that $\mathbb{E}[X^k|Y=y]$ and $\kappa_{X|Y=y}(k)$ are also real-analytic. ∎

Before studying the Taylor series of the conditional expectation, it is instructive to consider the following example.

*Example:*For X uniformly distributed on $\{-1,1\}$, the conditional expectation is given by

$$\mathbb{E}[X|Y=y] = \tanh\left(\frac{y}{\sigma^2}\right), y \in \mathbb{R}.$$

By using the Taylor series of $\tanh(\cdot)$ around zero [20, Eq. 4.5.64], we have that

$$\mathbb{E}[X|Y=y] = \sum_{k=1}^{\infty} \frac{2^{2k}(2^{2k}-1)b_{2k}}{(2k)!} \left(\frac{y}{\sigma^2}\right)^{2k-1}, \, |y| < \frac{\sigma^2\pi}{2},$$

where b_n is the n-th Bernoulli number.

The key observation here is that even in this simple case, the Taylor expansion has a finite radius of convergence. Therefore, in general, we cannot expect to get a power series representation of $\mathbb{E}[X|Y=y]$ that converges for all $\mathbb{R}$ (i.e., the power series with an infinite radius of convergence).

We now show two bounds on the absolute value of the conditional cumulants. Together with the identity in (48), these bounds produce bounds on the rate of growth of the conditional expectation. In addition, these bounds will be used to characterize the Taylor series expansion of the conditional expectation.

Proposition 9: For $y \in \mathbb{R}$ and $k \in \mathbb{N}$

$$|\kappa_{X|Y=y}(k)| \le 2^{k-1} k^k \mathbb{E}[|X|^k|Y=y] \tag{81}$$

$$\kappa_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s_1,s_2) = \begin{cases} \frac{Ry_{s_2}y_{s_1}}{\|\mathbf{y}\|} h_{\frac{n}{2}}(R\|\mathbf{y}\|) + \frac{R^2 y_{s_2} y_{s_1}}{\|\mathbf{y}\|^2} \left(1 - \frac{n-1}{R\|\mathbf{y}\|} h_{\frac{n}{2}}(R\|\mathbf{y}\|) - \left(h_{\frac{n}{2}}(R\|\mathbf{y}\|)\right)^2\right), & s_1 \neq s_2 \\ R\frac{\|\mathbf{y}\| - \frac{y_{s_2}^2}{\|\mathbf{y}\|}}{\|\mathbf{y}\|^2} h_{\frac{n}{2}}(R\|\mathbf{y}\|) + \frac{R^2 y_{s_2}^2}{\|\mathbf{y}\|^2} \left(1 - \frac{n-1}{R\|\mathbf{y}\|} h_{\frac{n}{2}}(R\|\mathbf{y}\|) - \left(h_{\frac{n}{2}}(R\|\mathbf{y}\|)\right)^2\right), & s_1 = s_2 \end{cases}. \tag{79}$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

$$\leq a_k|y|^k + b_k, \tag{82}$$

where

$$a_k = k^k 2^{k-1}(2^{\max(\frac{k}{2}-1,1)}+2), \tag{83}$$

$$b_k = k^k(2^{\max(\frac{k}{2}-1,1)+k}\mathbb{E}^{\frac{k}{2}}[X^2] + \mathbb{E}[|X|^k]). \tag{84}$$

Proof: See Appendix C. ∎

The next result provides a power series representation for the conditional expectation. In addition, it also finds a lower bound on the radius of converges in the case when X is bounded.

Theorem 4: Fix some $a \in \mathbb{R}$. Then, for every X there exists some $r_{\sigma,a} > 0$ such that

$$\mathbb{E}[X|Y=y] = \sum_{k=0}^{\infty} \frac{\kappa_{X|Y=a}(k+1)}{k!\sigma^{2k}}(y-a)^k,\ |y-a| < r_{\sigma,a}. \tag{85}$$

In addition, if $|X| \leq A$, then $r_{\sigma,a} \geq \frac{\sigma^2}{2Ae}$.

Proof: From Lemma 2, we have that for every $a \in \mathbb{R}$ there exits an $r_{\sigma,a} > 0$ such that $\mathbb{E}[X|Y=y]$ has a power series representation on $(a - r_{\sigma,a}, a + r_{\sigma,a})$. Moreover, by using (48), this power series is given by

$$\mathbb{E}[X|Y=y] = \sum_{k=0}^{\infty} \frac{\mathbb{E}^{(k)}[X|Y=a]}{k!}(y-a)^k \tag{86}$$

$$= \sum_{k=0}^{\infty} \frac{\kappa_{X|Y=a}(k+1)}{k!\sigma^{2k}}(y-a)^k. \tag{87}$$

Finally, the radius of convergence for $|X| \leq A$ can be found as follows by using the root test:

$$r_{\sigma,a} = \limsup_{n\to\infty} \left| \frac{\kappa_{X|Y=a}(k+1)}{k!\sigma^{2k}} \right|^{-\frac{1}{k}} \tag{88}$$

$$\geq \limsup_{n\to\infty} \left| \frac{2^k(k+1)^{k+1}\mathbb{E}[|X|^{k+1}|Y=a]}{k!\sigma^{2k}} \right|^{-\frac{1}{k}} \tag{89}$$

$$\geq \limsup_{n\to\infty} \left| \frac{2^k(k+1)^{k+1}A^{k+1}}{k!\sigma^{2k}} \right|^{-\frac{1}{k}} \tag{90}$$

$$= \frac{\sigma^2}{2Ae}, \tag{91}$$

where (89) follows from the bound in (81); (90) follows by using $\mathbb{E}[|X|^{k+1}|Y] \leq A^{k+1}$; and (91) follows from the limit $\limsup_{n\to\infty} \left| \frac{(k+1)^{k+1}}{k!} \right|^{-\frac{1}{k}} = \frac{1}{e}$. ∎

Remark 3: It is important to note that since the conditional cumulants can be expressed in terms of f_Y and derivatives of f_Y (see (55)), the power series can also be expressed in terms of f_Y only.

The approximation of the conditional expectation with polynomials of degree k has been recently considered in [18]. Using results from Bernstein's approximation theory, the authors of [18] were able to quantify the average approximation error as a function of the degree k.

VI. Identities for Distributions, the Information Density and the MMSE

In this section, we study identities between the conditional expectation and quantiles such the conditional distribution $P_{\mathbf{X}|\mathbf{Y}}$, the pdf of $\mathbf{Y}$ and the information density of the pair $(\mathbf{X}, \mathbf{Y})$. As a small application, we show how such identities can be used to find lower bounds on the MMSE.

A. An Alternative View and Generalization of the TRE Identity, and Higher-Order Derivative of the Information Density

Let *the information density* be defined as

$$\iota_{P_{\mathbf{XY}}}(\mathbf{x};\mathbf{y}) = \log \frac{\mathrm{d}P_{\mathbf{XY}}}{\mathrm{d}(P_{\mathbf{X}} \otimes P_{\mathbf{Y}})}(\mathbf{x},\mathbf{y}),\ \mathbf{x},\mathbf{y} \in \mathbb{R}^n. \tag{92}$$

In this section, we are interested in characterizing derivatives of the information density with respect to the variable $\mathbf{y}$.

We start this section by observing the following alternative version of the TRE identity, which establishes the gradient of information density.

Proposition 10: For $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$

$$\nabla_{\mathbf{y}} \iota_{P_{\mathbf{XY}}}(\mathbf{x};\mathbf{y}) = \mathbf{K}_{\mathbf{N}}^{-1}(\mathbf{x} - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]). \tag{93}$$

Proof: Fix some $\mathbf{x}$ and $\mathbf{y}$. Then,

$$\nabla_{\mathbf{y}} \iota(\mathbf{x};\mathbf{y}) = \nabla_{\mathbf{y}} \log \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} \tag{94}$$

$$= \nabla_{\mathbf{y}} \log \frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} \tag{95}$$

$$= -\mathbf{K}_{\mathbf{N}}^{-1}(\mathbf{y}-\mathbf{x}) - \nabla_{\mathbf{y}} \log f_{\mathbf{Y}}(\mathbf{y}) \tag{96}$$

$$= -\mathbf{K}_{\mathbf{N}}^{-1}(\mathbf{y}-\mathbf{x}) - \mathbf{K}_{\mathbf{N}}^{-1}(\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}] - \mathbf{y}) \tag{97}$$

$$= \mathbf{K}_{\mathbf{N}}^{-1}(\mathbf{x} - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]), \tag{98}$$

where in (96) we have used $\nabla_{\mathbf{y}} \log \phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{x}) = -\mathbf{K}_{\mathbf{N}}^{-1}(\mathbf{y}-\mathbf{x})$; and (97) follows by using the TRE identity in (3). This concludes the proof. ∎

Using Theorem 1 and setting $U = 1_{\mathcal{A}}(\mathbf{X})$, $\mathcal{A} \subseteq \mathbb{R}^n$, the TRE identity can be generalized as follows.

Proposition 11: Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a measurable set such that $\mathbb{P}[\mathbf{X} \in \mathcal{A}] > 0$. Then, for $\mathbf{y} \in \mathbb{R}^n$

$$\nabla_{\mathbf{y}} \log\left(\mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}]\right) = \mathbf{K}_{\mathbf{N}}^{-1}(\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}, \mathbf{X} \in \mathcal{A}] - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]). \tag{99}$$

Proof: Let $U = 1_{\mathcal{A}}(\mathbf{X})$ in Theorem 1. This choice of $\mathbf{U}$ implies the following:

$$\mathbb{E}[U|\mathbf{Y}=\mathbf{y}] = \mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}], \tag{100}$$

$$\mathbb{E}[\mathbf{X}U|\mathbf{Y}=\mathbf{y}] = \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}, \mathbf{X} \in \mathcal{A}]\mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}]. \tag{101}$$

Combining (100) and (101) with the identity in (15) we arrive at

$$\frac{\nabla_{\mathbf{y}} \mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}]}{\mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}]} = \mathbf{K}_{\mathbf{N}}^{-1}(\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}, \mathbf{X} \in \mathcal{A}] - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]), \tag{102}$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

where we have used that $\mathbb{P}[\mathbf{X} \in \mathcal{A}] > 0$. The proof is concluded by observing that $\nabla_{\mathbf{y}} \log(\mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y} = \mathbf{y}]) = \frac{\nabla_{\mathbf{y}} \mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}]}{\mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}]}$. ∎

To see that (99) is a generalization of (93) suppose that $\mathbf{X}$ is a discrete random vector. Then, by setting $\mathcal{A} = \{\mathbf{x}\}$ where $\mathbf{x}$ is a point of support of $\mathbf{X}$, the identity in (99) reduces to

$$\begin{aligned}\nabla_{\mathbf{y}} \log\left(\mathbb{P}[\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}]\right) &= \nabla_{\mathbf{y}} \imath_{P_{\mathbf{XY}}}(\mathbf{x};\mathbf{y}) \quad (103)\\ &= \mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{x} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]), \quad (104)\end{aligned}$$

where we have used that $\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}] = \mathbf{x}$.

As an application of Proposition 10 and Proposition 11, we now characterizing the Hessian of the information density and the Hessian of the log of the posterior distribution. Again, the key ingredient in the proof will be the identity in Theorem 1.

Proposition 12: For $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n$:

- (Hessian of the Information Density)

$$\mathsf{D}_{\mathbf{y}}^2 \imath_{P_{\mathbf{XY}}}(\mathbf{x};\mathbf{y}) = -\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Var}(\mathbf{X}|\mathbf{Y} = \mathbf{y})\mathsf{K}_{\mathbf{N}}^{-1}. \quad (105)$$

- (General Case) Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a measurable set such that $\mathbb{P}[\mathbf{X} \in \mathcal{A}] > 0$. Then,

$$\mathsf{D}_{\mathbf{y}}^2 \log\left(\mathbb{P}[\mathbf{X} \in \mathcal{A}|\mathbf{Y}=\mathbf{y}]\right) = \mathsf{K}_{\mathbf{N}}^{-1}\Big(\mathsf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y}, \mathbf{X} \in \mathcal{A}) - \mathsf{Var}(\mathbf{X}|\mathbf{Y} = \mathbf{y})\Big)\mathsf{K}_{\mathbf{N}}^{-1}. \quad (106)$$

Proof: See Appendix G. ∎

By choosing a specific family of sets $\mathcal{A}$, we next illustrate an example of how the identity in (105) can be used to study other quantities such the conditional cumulative distribution function (cdf).

Example: Consider a family of sets given by $\mathcal{A} = (-\infty, t]$, $t > 0$. With this choice, we have that $\mathbb{P}[X \in \mathcal{A}|Y = y]$ is equal to the conditional cumulative distribution function $F(X \le t|Y = y)$ and by using (99), we arrive at

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} \log\left(F(X \le t|Y = y)\right) = \frac{\mathsf{Var}\left(X|Y = y, X \le t\right) - \mathsf{Var}\left(X|Y = y\right)}{\sigma^4}. \quad (107)$$

Now, consider the case of $X \sim \mathcal{N}(0,1)$. Then, $X|Y = y \sim \mathcal{N}(\frac{y}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2})$. Moreover, since $X|\{X \le t\}$ is a truncated Gaussian [28], it can be shown that

$$\mathsf{Var}\left(X|Y = y, X \le t\right) = \frac{\sigma^2}{1+\sigma^2}\left(1 - \frac{\beta(y)\phi(\beta(y))}{\Phi(\beta(y))} - \left(\frac{\phi(\beta(y))}{\Phi(\beta(y))}\right)^2\right), \quad (108)$$

where

$$\beta(y) = \frac{t - \frac{y}{1+\sigma^2}}{\sqrt{\frac{\sigma^2}{1+\sigma^2}}}. \quad (109)$$

Consequently, from (106) we have that

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} \log\left(F(X \le t|Y = y)\right) = -\frac{1}{\sigma^2(1+\sigma^2)}\left(\frac{\beta(y)\phi(\beta(y))}{\Phi(\beta(y))} + \left(\frac{\phi(\beta(y))}{\Phi(\beta(y))}\right)^2\right). \quad (110)$$

From Proposition 12 we have the following corollary.

Corollary 2: The mapping $(\mathbf{x}, \mathbf{y}) \to \mathsf{D}_{\mathbf{y}}^2 \imath_{P_{\mathbf{XY}}}(\mathbf{x};\mathbf{y})$ is only a function of the variable $\mathbf{y}$. Moreover, $\mathbf{y} \to \imath_{P_{\mathbf{XY}}}(\cdot;\mathbf{y})$ is a concave function.

We were not able to locate an explicit statement of the above result in the literature. The equivalent statement was produced as an intermediate step in a proof used in [15, p. 2229] where convexity of the function akin to the log-likelihood ratio was shown.

So far we have characterized the first and the second order derivatives of the information density and discussed generalizations of these identities. The next result provides an expression for all the higher-order derivatives in the scalar case.

Proposition 13: For $k \ge 2$ and $(x, y) \in \mathbb{R} \times \mathbb{R}$

$$\frac{\mathrm{d}^k}{\mathrm{d}y^k} \imath_{P_{XY}}(x;y) = -\frac{1}{\sigma^{2k}} \kappa_{X|Y=y}(k). \quad (111)$$

Proof:

$$\begin{aligned}\frac{\mathrm{d}^k}{\mathrm{d}y^k} \imath_{P_{XY}}(x;y) &= \frac{\mathrm{d}^{k-1}}{\mathrm{d}y^{k-1}} \frac{1}{\sigma^2}(x - \mathbb{E}[X|Y = y]) \quad (112)\\ &= -\frac{1}{\sigma^{2k}} \kappa_{X|Y=y}(k), \quad (113)\end{aligned}$$

where in (112) we have used (93); and in (113) we have used (48). ∎

B. An Inverse TRE Identity

The TRE identity shows that the conditional expectation is completely determined by $f_{\mathbf{Y}}$. It is also possible to have an inverse statement that shows that $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ completely determines $f_{\mathbf{Y}}$.

Proposition 14: For $\mathbf{y} \in \mathbb{R}^n$

$$f_{\mathbf{Y}}(\mathbf{y}) = c \exp\left(\oint_{\mathbf{0}}^{\mathbf{y}} \mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{t} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{t}]) \cdot \mathbf{dt}\right), \quad (114)$$

where $0 < c < \infty$ is the normalization constant and is given by

$$c^{-1} = \int_{\mathbb{R}^n} \exp\left(\oint_{\mathbf{0}}^{\mathbf{y}} \mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{t} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{t}]) \cdot \mathbf{dt}\right) \mathbf{dy}. \quad (115)$$

Proof: The TRE expression in (3) can be rewritten as

$$\nabla_{\mathbf{y}} \log(f_{\mathbf{Y}}(\mathbf{y})) = \mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{y} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]). \quad (116)$$

Using the fundamental theorem of calculus for the line integral, we have that

$$\log(f_{\mathbf{Y}}(\mathbf{y})) - \log(f_{\mathbf{Y}}(\mathbf{0})) = \oint_{\mathbf{0}}^{\mathbf{y}} \mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{t} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{t}]) \cdot \mathbf{dt}, \quad (117)$$

or equivalently

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}}(\mathbf{0}) \exp\left(\oint_{\mathbf{0}}^{\mathbf{y}} \mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{t} - \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{t}]) \cdot \mathbf{dt}\right). \quad (118)$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

After integrating both sides of (118), we arrive at

$$f_{\mathbf{Y}}(\mathbf{0}) = \frac{1}{\int_{\mathbb{R}^n} \exp\left(\oint_{\mathbf{0}}^{\mathbf{y}} \mathbf{K_N^{-1}}(\mathbf{t} - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{t}]) \cdot \mathbf{dt}\right) \mathbf{dy}}. \tag{119}$$

Letting $c = f_{\mathbf{Y}}(\mathbf{0})$ and noting that

$$0 < f_{\mathbf{Y}}(\mathbf{0}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det^{\frac{1}{2}}(\mathbf{K_N})} \mathbb{E}\left[\mathrm{e}^{-\frac{\mathbf{x}^{\mathsf{T}}\mathbf{K_N^{-1}}\mathbf{x}}{2}}\right] < \infty \tag{120}$$

concludes the proof. ∎

We now use the representation of $f_{\mathbf{Y}}$ in Proposition 14 to show the following important result on the uniquness of the conditional expectation.

Corollary 3: The conditional expectation $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ is a bijective operator of $P_{\mathbf{X}}$ or $f_{\mathbf{Y}}$. In other words, for $\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{N}_1$ and $\mathbf{Y}_2 = \mathbf{X}_2 + \mathbf{N}_2$ where $\mathbf{N}_1$ and $\mathbf{N}_2$ are Gaussian with the same covariance matrix,

$$\begin{aligned} &\mathbb{E}[\mathbf{X}_1|\mathbf{Y}_1 = \mathbf{y}] = \mathbb{E}[\mathbf{X}_2|\mathbf{Y}_2 = \mathbf{y}], \forall \mathbf{y} \in \mathbb{R}^n \\ &\quad \Longleftrightarrow P_{\mathbf{X}_1} = P_{\mathbf{X}_2} \\ &\quad \Longleftrightarrow f_{\mathbf{Y}_1}(\mathbf{y}) = f_{\mathbf{Y}_2}(\mathbf{y}),\ \forall \mathbf{y} \in \mathbb{R}^n. \end{aligned} \tag{121}$$

Proof: First, let $P_{\mathbf{X}_1} = P_{\mathbf{X}_2}$, then it is immediate that

$$\mathbb{E}[\mathbf{X}_1|\mathbf{Y}_1 = \mathbf{y}] = \mathbb{E}[\mathbf{X}_2|\mathbf{Y}_2 = \mathbf{y}],\ \forall \mathbf{y} \in \mathbb{R}^n. \tag{122}$$

Now, suppose that

$$\mathbb{E}[\mathbf{X}_1|\mathbf{Y}_1 = \mathbf{y}] = \mathbb{E}[\mathbf{X}_2|\mathbf{Y}_2 = \mathbf{y}],\ \forall \mathbf{y} \in \mathbb{R}^n. \tag{123}$$

Then, by using (114) we have that

$$f_{\mathbf{Y}_1}(\mathbf{y}) = f_{\mathbf{Y}_2}(\mathbf{y}),\ \forall \mathbf{y} \in \mathbb{R}^n. \tag{124}$$

The fact that (124) implies that $P_{\mathbf{X}_1} = P_{\mathbf{X}_2}$ follows from the standard argument that uses characteristics functions. ∎

Corollary 3 has important ramification in estimation theory. In particular, combining Corollary 3 and [14, Thm. 1] leads to a conclusion that the $\mathrm{mmse}(\mathbf{X}|\mathbf{Y})$ is a *strictly* convex function of the input distribution $P_{\mathbf{X}}$. This can further be used to argue that optimization problems of the following form have unique maximizers:

$$\max_{P_{\mathbf{X}} \in \mathcal{P}} \mathrm{mmse}(\mathbf{X}|\mathbf{Y}), \tag{125}$$

where $\mathcal{P}$ is some compact set of probability distributions. The interested reader is referred to [29] for more details.

C. Representations of the MMSE

In this section, we use properties developed for the information density to find alternative representations of the MMSE. These representations are then used to find lower bounds on the MMSE.

Proposition 15:

- (Gradient Representation)

$$\begin{aligned} &\mathbb{E}\left[\nabla_{\mathbf{Y}} \imath_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y}) \left(\nabla_{\mathbf{Y}} \imath_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})\right)^{\mathsf{T}}\right] \\ &\quad = \mathbf{K_N^{-1}}\mathsf{MMSE}(\mathbf{X}|\mathbf{Y})\mathbf{K_N^{-1}}. \end{aligned} \tag{126}$$

Consequently,

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \mathbb{E}\left[\|\mathbf{K_N} \nabla_{\mathbf{Y}} \imath_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})\|^2\right]. \tag{127}$$

- (Hessian Representation)

$$\mathbb{E}\left[\mathbf{D}^2_{\mathbf{Y}} \imath_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})\right] = -\mathbf{K_N^{-1}}\mathsf{MMSE}(\mathbf{X}|\mathbf{Y})\mathbf{K_N^{-1}}. \tag{128}$$

Consequently,

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = -\mathrm{Tr}\left(\mathbf{K_N^2}\mathbb{E}\left[\mathbf{D}^2_{\mathbf{Y}} \imath_{P_{\mathbf{XY}}}(\mathbf{X};\mathbf{Y})\right]\right). \tag{129}$$

Proof: First, observe that the MMSE matrix can be represented as

$$\mathsf{MMSE}(\mathbf{X}|\mathbf{Y}) = \mathbb{E}\left[\mathsf{Var}(\mathbf{X}|\mathbf{Y})\right]. \tag{130}$$

The proof of (127) now follows by using the expression for the conditional variance in Proposition 10, and the proof of (128) follows from the expression in Proposition 12. This concludes the proof. ∎

As was recently shown in [30], identities of the type in Proposotion 15 can be used to derive new Bayesian lower bounds on the MMSE.

VII. Conclusion and Outlook

This work has derived a general derivative identity for a conditional mean estimator. This identity has been used to recover several known derivative identities, such as the Hatsel and Nolte identity for the conditional variance and the recursive Jaffer's identity. Moreover, several new identities have been derived, the most notable of which include: a new integral version of Jaffer's identity, the identity that connects higher-order conditional moments and the derivatives of the conditional expectation via Bell polynomials, and the identity that shows that the derivatives of the conditional expectation are proportional the conditional cumulants. All of the derived identities are summarized in Table I (top of the next page).

One interesting future direction is to generalize the identities derived in this paper to more general noise settings; see for example [31] on a recent extension to the exponential family. Another interesting future direction would be to examine whether the main identity in Theorem 1 can shed light on the vector generalization of the single crossing property in [32].

Appendix A
Proof of Theorem 1

First, observe that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}y_m}\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X}) \\ &= -\frac{1}{2}\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})\frac{\mathrm{d}}{\mathrm{d}y_m}(\mathbf{y}-\mathbf{X})^{\mathsf{T}}\mathbf{K_N^{-1}}(\mathbf{y}-\mathbf{X}) \qquad (131) \\ &= \phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})\mathbf{e}_m^{\mathsf{T}}\mathbf{K_N^{-1}}(\mathbf{X}-\mathbf{y}). \qquad (132) \end{aligned}$$

Second, for the moment assume that the interchange of expectation and differentiation is allowed, and observe the following sequence of steps:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}y_m}\mathbb{E}\left[\mathbf{U}|\mathbf{Y}=\mathbf{y}\right] \\ &= \frac{\mathrm{d}}{\mathrm{d}y_m}\mathbb{E}\left[\mathbf{U}\frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}\right] \qquad (133) \\ &= \mathbb{E}\left[\mathbf{U}\frac{\mathrm{d}}{\mathrm{d}y_m}\frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}\right] \qquad (134) \end{aligned}$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

TABLE I
LIST OF IDENTITIES

	Name	Identity
Two Main Identities	TRE Identity (3)	$\mathbb{E}\left[\mathbf{X}\middle\vert\mathbf{Y}=\mathbf{y}\right]=\mathbf{y}+\mathsf{K}_{\mathbf{N}}\frac{\nabla_{\mathbf{y}}f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}$
	Identity (15)	$\mathsf{J}_{\mathbf{y}}\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]=\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Cov}(\mathbf{X},\mathbf{U}\vert\mathbf{Y}=\mathbf{y}),\quad \mathbf{U}\leftrightarrow\mathbf{X}\leftrightarrow\mathbf{Y}$
Variance	Hatsell and Nolte (17)	$\mathsf{J}_{\mathbf{y}}\mathbb{E}\left[\mathbf{X}\middle\vert\mathbf{Y}=\mathbf{y}\right]=\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Var}(\mathbf{X}\vert\mathbf{Y}=\mathbf{y})$
Recursive Identities	Jaffer's Identity (22)	$\mathbb{E}\left[X^{k+1}\middle\vert Y=y\right]=\sigma^2\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}\left[X^k\middle\vert Y=y\right]+\mathbb{E}\left[X^k\middle\vert Y=y\right]\mathbb{E}[X\vert Y=y])$
	Integral version of Jaffer's Identity (29)	$\mathbb{E}\left[X^k\middle\vert Y=y\right]=\sigma^{2k}\mathrm{e}^{-\frac{1}{\sigma^2}\int_0^y\mathbb{E}[X\vert Y=t]\mathrm{d}t}\frac{\mathrm{d}^k}{\mathrm{d}y^k}\mathrm{e}^{\frac{1}{\sigma^2}\int_0^y\mathbb{E}[X\vert Y=t]\mathrm{d}t}$
	Vector Jaffer's Identity (Version 1) (42)	$\mathbb{E}\left[(\mathbf{X}\mathbf{X}^{\mathsf{T}})^k\middle\vert\mathbf{Y}=\mathbf{y}\right]=\mathsf{K}_{\mathbf{N}}\mathsf{J}_{\mathbf{y}}\mathbb{E}\left[(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{k-1}\mathbf{X}\middle\vert\mathbf{Y}=\mathbf{y}\right]$ $+\mathbb{E}[\mathbf{X}\vert\mathbf{Y}=\mathbf{y}]\mathbb{E}\left[\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{k-1}\middle\vert\mathbf{Y}=\mathbf{y}\right]$
	Conditional Expectation via Higher-Order Derivatives (30)	$\mathbb{E}\left[X^k\middle\vert Y=y\right]=\sigma^{2k}\mathsf{B}_k\left(\mathbb{E}^{(0)}\left[\frac{X}{\sigma^2}\middle\vert Y=y\right],\ldots,\mathbb{E}^{(k-1)}\left[\frac{X}{\sigma^2}\middle\vert Y=y\right]\right)$
	Derivatives of the Conditional Expectation (31)	$\frac{\mathrm{d}^k}{\mathrm{d}y^k}\mathbb{E}[X\vert Y=y]=\sigma^2\sum_{m=1}^{k+1}c_m\mathsf{B}_{k+1,m}\left(\mathbb{E}\left[\frac{X}{\sigma^2}\middle\vert Y=y\right],\ldots,\mathbb{E}\left[\left(\frac{X}{\sigma^2}\right)^{k-m+2}\middle\vert Y=y\right]\right)$
	Generalized TRE Identity (37)	$\mathbb{E}[X^k\vert Y=y]=\sigma^{2k}\frac{\frac{\mathrm{d}^k}{\mathrm{d}y^k}\left(f_Y(y)\mathrm{e}^{\frac{y^2}{2\sigma^2}}\right)}{f_Y(y)\mathrm{e}^{\frac{y^2}{2\sigma^2}}}=\sigma^{2k}\frac{\sum_{m=0}^k\binom{k}{m}f_Y^{(k-m)}(y)\frac{(-i)^m}{\sigma^m}H_{e_m}\left(i\frac{y}{\sigma}\right)}{f_Y(y)}$
Identities for Conditional Cumulants	Conditional Cumulants and Conditional Expectation (48)	$\kappa_{X\vert Y=y}(k+1)=\sigma^{2k}\frac{\mathrm{d}^k}{\mathrm{d}y^k}\mathbb{E}[X\vert Y=y]$
	Recursion for Conditional Cumulants (52)	$\sigma^2\frac{\mathrm{d}}{\mathrm{d}y}\kappa_{X\vert Y=y}(k+1)=\kappa_{X\vert Y=y}(k+2)$
	Conditional Cumulants Generating Function and Conditional Expectation (62)	$\frac{\mathrm{d}^{k+1}}{\mathrm{d}t^{k+1}}K_X(t\vert Y=y)=\frac{\mathrm{d}^{k+1}}{\mathrm{d}y^{k+1}}K_X(t\vert Y=y)+\frac{\mathrm{d}^k}{\mathrm{d}y^k}\mathbb{E}[X\vert Y=y]$
	Multivariate Identity for Conditional Cumulants and Conditional Expectation (76)	$\mathbf{k}_{s_j}^{\mathsf{T}}\frac{\partial^{j-1}\mathbb{E}[\mathbf{X}\vert\mathbf{Y}=\mathbf{y}]}{\partial y_{s_1}\ldots\partial y_{s_{j-1}}}=\sum_{p_1=1}^n\cdots\sum_{p_j=1}^n\prod_{i=1}^j k_{s_i,p_i}\kappa_{\mathbf{X}\vert\mathbf{Y}=\mathbf{y}}(p_1,\ldots,p_j)$
Identities for Distributions and the Information Density	Inverse TRE Identity (114)	$f_{\mathbf{Y}}(\mathbf{y})=c\exp\left(\oint_{\mathbf{0}}^{\mathbf{y}}\mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{t}-\mathbb{E}[\mathbf{X}\vert\mathbf{Y}=\mathbf{t}])\cdot\mathbf{dt}\right)$
	Gradient of the Information Density (93)	$\nabla_{\mathbf{y}}\iota(\mathbf{x};\mathbf{y})=\mathsf{K}_{\mathbf{N}}^{-1}(\mathbf{x}-\mathbb{E}[\mathbf{X}\vert\mathbf{Y}=\mathbf{y}]))$
	Gradient of the Conditional Distribution (99)	$\nabla_{\mathbf{y}}\log\left(\mathbb{P}[\mathbf{X}\in\mathcal{A}\vert\mathbf{Y}=\mathbf{y}]\right)=\mathsf{K}_{\mathbf{N}}^{-1}(\mathbb{E}[\mathbf{X}\vert\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}]-\mathbb{E}[\mathbf{X}\vert\mathbf{Y}=\mathbf{y}])$
	Hessian of Information Density (105)	$\mathsf{D}_{\mathbf{y}}^2\iota(\mathbf{x};\mathbf{y})=-\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Var}(\mathbf{X}\vert\mathbf{Y}=\mathbf{y})\mathsf{K}_{\mathbf{N}}^{-1}$
	Hessian of the Conditional Distribution (105)	$\mathsf{D}_{\mathbf{y}}^2\log\left(\mathbb{P}[\mathbf{X}\in\mathcal{A}\vert\mathbf{Y}=\mathbf{y}]\right)=\mathsf{K}_{\mathbf{N}}^{-1}\left(\mathsf{Var}(\mathbf{X}\vert\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A})-\mathsf{Var}(\mathbf{X}\vert\mathbf{Y}=\mathbf{y})\right)\mathsf{K}_{\mathbf{N}}^{-1}$
	Higher-Order Derivatives of Information Density (111)	$\frac{\mathrm{d}^k}{\mathrm{d}y^k}\iota_{P_{XY}}(x;y)=-\frac{1}{\sigma^{2k}}\kappa_{X\vert Y=y}(k),k\geq 3$

$$=\mathbb{E}\left[\mathbf{U}\frac{\frac{\mathrm{d}}{\mathrm{d}y_m}\phi_{\mathsf{K}_{\mathbf{N}}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}\right]-\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]\frac{\frac{\mathrm{d}}{\mathrm{d}y_m}f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} \tag{135}$$

$$=\mathbb{E}\left[\mathbf{U}\mathbf{e}_m^{\mathsf{T}}\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{X}\middle\vert\mathbf{Y}=\mathbf{y}\right]-\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]\mathbf{e}_m^{\mathsf{T}}\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{y}$$
$$-\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]\frac{\frac{\mathrm{d}}{\mathrm{d}y_m}f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} \tag{136}$$

$$=\mathbb{E}\left[\mathbf{U}\mathbf{e}_m^{\mathsf{T}}\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{X}\middle\vert\mathbf{Y}=\mathbf{y}\right]$$
$$-\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]\mathbf{e}_m^{\mathsf{T}}\left(\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{y}+\frac{\nabla_{\mathbf{y}}f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}\right) \tag{137}$$

$$=\mathbb{E}\left[\mathbf{U}\mathbf{e}_m^{\mathsf{T}}\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{X}\middle\vert\mathbf{Y}=\mathbf{y}\right]$$
$$-\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]\mathbb{E}\left[\mathbf{e}_m^{\mathsf{T}}\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{X}\middle\vert\mathbf{Y}=\mathbf{y}\right] \tag{138}$$

$$=\mathbb{E}\left[\mathbf{U}\mathbf{X}^{\mathsf{T}}\middle\vert\mathbf{Y}=\mathbf{y}\right]\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{e}_m$$
$$-\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]\mathbb{E}\left[\mathbf{X}^{\mathsf{T}}\middle\vert\mathbf{Y}=\mathbf{y}\right]\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{e}_m \tag{139}$$

$$=\mathsf{Cov}(\mathbf{U},\mathbf{X}\vert\mathbf{Y}=\mathbf{y})\mathsf{K}_{\mathbf{N}}^{-1}\mathbf{e}_m, \tag{140}$$

where the equalities follow from: (133) using Bayes' formula; (136) using the expression in (132); and (138) using the TRE identity in (3).

Now using the definition of Jacobian in (9), we have that

$$\mathsf{J}_{\mathbf{y}}\mathbb{E}\left[\mathbf{U}\middle\vert\mathbf{Y}=\mathbf{y}\right]=\left(\mathsf{Cov}(\mathbf{U},\mathbf{X}\vert\mathbf{Y}=\mathbf{y})\mathsf{K}_{\mathbf{N}}^{-1}\right)^{\mathsf{T}} \tag{141}$$
$$=\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Cov}^{\mathsf{T}}(\mathbf{U},\mathbf{X}\vert\mathbf{Y}=\mathbf{y}) \tag{142}$$
$$=\mathsf{K}_{\mathbf{N}}^{-1}\mathsf{Cov}(\mathbf{X},\mathbf{U}\vert\mathbf{Y}=\mathbf{y}), \tag{143}$$

where we have used the symmetry of $\mathsf{K}_{\mathbf{N}}$ and the property that $\mathsf{Cov}^{\mathsf{T}}(\mathbf{U},\mathbf{X}\vert\mathbf{Y}=\mathbf{y})=\mathsf{Cov}(\mathbf{X},\mathbf{U}\vert\mathbf{Y}=\mathbf{y})$.

Therefore, to conclude the proof, we require to show that the interchange of differentiation and expectation in (134) is permitted. A sufficient condition for the interchange of differentiation and expectation is given by the Leibniz integral rule, which requires verifying that

$$\mathbb{E}\left[\left\|\mathbf{U}\frac{\mathrm{d}}{\mathrm{d}y_m}\frac{\phi_{\mathsf{K}_{\mathbf{N}}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}\right\|\right]<\infty. \tag{144}$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

To this end, note that

$$\mathbb{E}\left[\left\|\mathbf{U}\frac{\mathrm{d}}{\mathrm{d}y_m}\frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}\right\|\right] = \mathbb{E}\left[\left\|\mathbf{U}\left(\frac{\frac{\mathrm{d}}{\mathrm{d}y_m}\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}-\frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})\frac{\mathrm{d}}{\mathrm{d}y_m}f_{\mathbf{Y}}(\mathbf{y})}{f^2_{\mathbf{Y}}(\mathbf{y})}\right)\right\|\right] \tag{145}$$

$$\leq \mathbb{E}\left[\left\|\mathbf{U}\frac{\frac{\mathrm{d}}{\mathrm{d}y_m}\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}\right\|\right] + \mathbb{E}\left[\left\|\mathbf{U}\frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})\frac{\mathrm{d}}{\mathrm{d}y_m}f_{\mathbf{Y}}(\mathbf{y})}{f^2_{\mathbf{Y}}(\mathbf{y})}\right\|\right], \tag{146}$$

where the last step follows by using the triangle inequality. We now bound each term of (146) separately. The first term in (146) is bounded by

$$\mathbb{E}\left[\left\|\mathbf{U}\frac{\frac{\mathrm{d}}{\mathrm{d}y_m}\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y})}\right\|\right] = \mathbb{E}\left[\left\|\mathbf{U}\frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})\mathbf{e}_m^{\mathsf{T}}\mathbf{K}_{\mathbf{N}}^{-1}(\mathbf{X}-\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}\right\|\right] \tag{147}$$

$$= \mathbb{E}\left[\left\|\mathbf{U}\mathbf{e}_m^{\mathsf{T}}\mathbf{K}_{\mathbf{N}}^{-1}(\mathbf{X}-\mathbf{y})\right\|\,|\mathbf{Y}=\mathbf{y}\right] \tag{148}$$

$$\leq \mathbb{E}\left[\left\|\mathbf{U}\mathbf{e}_m^{\mathsf{T}}\mathbf{K}_{\mathbf{N}}^{-1}\mathbf{X}\right\|\,|\mathbf{Y}=\mathbf{y}\right] + \mathbb{E}\left[\left\|\mathbf{U}\mathbf{e}_m^{\mathsf{T}}\mathbf{K}_{\mathbf{N}}^{-1}\mathbf{y}\right\|\,|\mathbf{Y}=\mathbf{y}\right] \tag{149}$$

$$\leq \|\mathbf{e}_m^{\mathsf{T}}\mathbf{K}_{\mathbf{N}}^{-1}\|\mathbb{E}\left[\|\mathbf{U}\|\|\mathbf{X}\|\,|\mathbf{Y}=\mathbf{y}\right] + \|\mathbf{e}_m^{\mathsf{T}}\mathbf{K}_{\mathbf{N}}^{-1}\mathbf{y}\|\mathbb{E}\left[\|\mathbf{U}\|\,|\mathbf{Y}=\mathbf{y}\right]. \tag{150}$$

where (148) follows by using sing Bayes' formula; and (150) follows by using the Cauchy-Schwarz inequality.

The second term in (146) can be rewritten as

$$\mathbb{E}\left[\left\|\mathbf{U}\frac{\phi_{\mathbf{K_N}}(\mathbf{y}-\mathbf{X})\frac{\mathrm{d}}{\mathrm{d}y_m}f_{\mathbf{Y}}(\mathbf{y})}{f^2_{\mathbf{Y}}(\mathbf{y})}\right\|\right] = \mathbb{E}\left[\left\|\mathbf{U}\frac{\frac{\mathrm{d}}{\mathrm{d}y_m}f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}\right\|\,\middle|\,\mathbf{Y}=\mathbf{y}\right] \tag{151}$$

$$= \frac{|\frac{\mathrm{d}}{\mathrm{d}y_m}f_{\mathbf{Y}}(\mathbf{y})|}{f_{\mathbf{Y}}(\mathbf{y})}\mathbb{E}\left[\|\mathbf{U}\|\,\middle|\,\mathbf{Y}=\mathbf{y}\right]. \tag{152}$$

Therefore, by combining (150) and (152), the condition in (144) holds if

$$\mathbb{E}\left[\|\mathbf{U}\|\|\mathbf{X}\|\,|\mathbf{Y}=\mathbf{y}\right]<\infty, \mathbf{y}\in\mathbb{R}^n, \tag{153}$$

$$\mathbb{E}\left[\|\mathbf{U}\|\,|\mathbf{Y}=\mathbf{y}\right]<\infty, \mathbf{y}\in\mathbb{R}^n. \tag{154}$$

This concludes the proof.

Appendix B
Proof of Lemma 1

First, we multiply both side of (25) by $\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t}$, which leads to

$$f_k(x)\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}x}f_{k-1}(x)\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t} + f_{k-1}(x)f_1(x)\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t} \tag{155}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x}\left(f_{k-1}(x)\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t}\right), \tag{156}$$

where the last expression is due to the product rule. Now, let

$$h_k(x) = f_k(x)\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t}. \tag{157}$$

Then, the expression in (156) implies that

$$h_k(x) = \frac{\mathrm{d}}{\mathrm{d}x}h_{k-1}(x), \tag{158}$$

which by using the recursion $k-1$ times implies that

$$h_k(x) = \frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}}h_1(x) = \frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}}\left(f_1(x)\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t}\right). \tag{159}$$

Furthermore, by applying the chain rule in (159), we arrive at

$$h_k(x) = \frac{\mathrm{d}^k}{\mathrm{d}x^k}\mathrm{e}^{\int_0^x f_1(t)\mathrm{d}t}. \tag{160}$$

The proof of (26) is concluded by using the definition of $h_k(x)$ in (158).

To show the expression in (27) recall the Faà di Bruno formula for the higher-order chain rule [22, Thm. 11.4]: given two $k\in\mathbb{N}$ differentiable functions $g(t)$ and $\xi(t)$

$$\frac{\mathrm{d}^k\xi(g(t))}{\mathrm{d}t^k} = \sum_{m=1}^{k}\xi^{(m)}(g(t))\mathsf{B}_{k,m}\left(g^{(1)}(x),\ldots,g^{(k-m+1)}(x)\right). \tag{161}$$

Next, by letting $g(x)=\int_0^x f_1(t)\mathrm{d}t$, the expression in (26) can be re-written as

$$f_k(x) = \mathrm{e}^{-g(x)}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\mathrm{e}^{g(x)} \tag{162}$$

$$= \mathrm{e}^{-g(x)}\sum_{m=1}^{k}\mathrm{e}^{g(x)}\mathsf{B}_{k,m}\left(g^{(1)}(x),\ldots,g^{(k-m+1)}(x)\right) \tag{163}$$

$$= \mathsf{B}_k\left(g^{(1)}(x),\ldots,g^{(k)}(x)\right) \tag{164}$$

$$= \mathsf{B}_k\left(f_1^{(0)}(x),\ldots,f_1^{(k-1)}(x)\right), \tag{165}$$

where (163) follows by using the Faà di Bruno formula in (161); (164) follows from (11); and (165) follows by noting that $g^{(m)}(x)=f^{(m-1)}(x), m\in\mathbb{N}$.

To show (28), we use the inversion formula for the Bell polynomial [22, Rem. 11.3] which asserts the following: if

$$y_k = \mathsf{B}_k(t_1,\ldots,t_k), \tag{166}$$

then

$$t_k = \sum_{m=1}^{k}(-1)^{m-1}(m-1)!\mathsf{B}_{k,m}(y_1,\ldots,y_{k-m+1}). \tag{167}$$

Setting $y_k=f_k(x)$ and $t_k=f_1^{(k-1)}(x)$ leads to

$$f_1^{(k-1)}(x) = \sum_{m=1}^{k}(-1)^{m-1}(m-1)!\mathsf{B}_{k,m}(f_1(x),\ldots,f_{k-m+1}(x)). \tag{168}$$

The proof is concluded by using a change of variable from $k-1$ to k.

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

APPENDIX C
PROOF OF PROPOSITION 9

First, observe that

$$|\kappa_{X|Y=y}(k)| \leq k^k \mathbb{E}[|X-\mathbb{E}[X|Y]|^k|Y=y] \tag{169}$$
$$\leq 2^{k-1}k^k\mathbb{E}[|X|^k|Y=y], \tag{170}$$

where in (169) we have used the bound in [33, eq. (4)]; and in (170) we have used $|a+b|^k \leq 2^{k-1}(|a|^k+|b|^k), k \geq 1$.

We now show a bound on $\mathbb{E}[|X|^k|Y=y]$, which is a generalization of the bound shown in [34, Proposition 1.2]. First,

$$\begin{aligned}&\mathbb{E}[|X|^k|Y=y]\\ &=\int_{f_{Y|X}(y|x)\leq f_Y(y)}|x|^k\frac{f_{Y|X}(y|x)}{f_Y(y)}\mathrm{d}P_X(x)\\ &\quad+\int_{f_{Y|X}(y|x)>f_Y(y)}|x|^k\frac{f_{Y|X}(y|x)}{f_Y(y)}\mathrm{d}P_X(x) \qquad (171)\\ &\leq \mathbb{E}[|X|^k]+\int_{f_{Y|X}(y|x)>f_Y(y)}|x|^k\frac{f_{Y|X}(y|x)}{f_Y(y)}\mathrm{d}P_X(x). \qquad (172)\end{aligned}$$

Next, we bounds the second term in (172)

$$f_{Y|X}(y|x) > f_Y(y) \tag{173}$$
$$\Rightarrow (y-x)^2 \leq 2\sigma^2\log\left(\frac{1}{\sqrt{2\pi\sigma^2}f_Y(y)}\right) \tag{174}$$
$$\Rightarrow |y-x| \leq \sqrt{2\sigma^2\log\left(\frac{1}{\sqrt{2\pi\sigma^2}f_Y(y)}\right)} \tag{175}$$
$$\Rightarrow |x| \leq |y|+\sqrt{2\sigma^2\log\left(\frac{1}{\sqrt{2\pi\sigma^2}f_Y(y)}\right)} \tag{176}$$
$$\Rightarrow |x|^k \leq 2^{k-1}\left(|y|^k+\left(2\sigma^2\log\left(\frac{1}{\sqrt{2\pi\sigma^2}f_Y(y)}\right)\right)^{\frac{k}{2}}\right) \tag{177}$$
$$\Rightarrow |x|^k \leq 2^{k-1}\left(|y|^k+2\left(y^2+\mathbb{E}[X^2]\right)^{\frac{k}{2}}\right) \tag{178}$$
$$\Rightarrow |x|^k \leq 2^{k-1}(2^{\max(\frac{k}{2}-1,1)}+2)|y|^k + 2^{\max(\frac{k}{2}-1,1)+k}\mathbb{E}^{\frac{k}{2}}[X^2], \tag{179}$$

where (176) follows from the reverse triangle inequality; (177) follows by using the bound $|a+b|^k \leq 2^{k-1}(|a|^k+|b|^k), k \geq 1$; (178) follows by using Jensen's inequality to

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathbb{E}\left[\mathrm{e}^{-\frac{(y-X)^2}{2\sigma^2}}\right] \tag{180}$$
$$\geq \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{\mathbb{E}\left[(y-X)^2\right]}{2\sigma^2}} \tag{181}$$
$$\geq \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{y^2+\mathbb{E}[X^2]}{\sigma^2}}; \text{ and} \tag{182}$$

(179) follows form the bound $|a+b|^k \leq 2^{\max(k-1,1)}(|a|^k+|b|^k), k \geq 0$.

Combining (172) and (179) concludes the proof.

APPENDIX D
PROOF OF THE EXPRESSION IN (79)

The first ingredient of the proof is the following derivative expression:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{I_\nu(t)}{I_{\nu-1}(t)} = 1-\frac{2\nu-1}{t}\frac{I_\nu(t)}{I_{\nu-1}(t)}-\left(\frac{I_\nu(t)}{I_{\nu-1}(t)}\right)^2, \tag{183}$$

which follows form the identities $\frac{\mathrm{d}}{\mathrm{d}t}I_\nu(t) = \frac{\nu}{t}I_{\nu-1}(t)-\frac{\nu}{t}I_\nu(t)$ and $\frac{\mathrm{d}}{\mathrm{d}t}I_\nu(t) = \frac{\nu}{t}I_\nu(t)+I_{\nu+1}(t)$ [35].

Next, consider the case of $s_1 \neq s_2$. Then,

$$\begin{aligned}&\frac{\partial\mathbb{E}[X_{s_2}|\mathbf{Y}=\mathbf{y}]}{\partial y_{s_1}}\\ &= Ry_{s_2}\frac{\partial}{\partial y_{s_1}}\frac{1}{\|\mathbf{y}\|}\frac{I_{\frac{n}{2}}(R\|\mathbf{y}\|)}{I_{\frac{n}{2}-1}(R\|\mathbf{y}\|)} \qquad (184)\\ &= \frac{Ry_{s_2}y_{s_1}}{\|\mathbf{y}\|}\frac{I_{\frac{n}{2}}(R\|\mathbf{y}\|)}{I_{\frac{n}{2}-1}(R\|\mathbf{y}\|)}+\frac{Ry_{s_2}}{\|\mathbf{y}\|}\frac{\partial}{\partial y_{s_1}}\frac{I_{\frac{n}{2}}(R\|\mathbf{y}\|)}{I_{\frac{n}{2}-1}(R\|\mathbf{y}\|)} \qquad (185)\\ &= \frac{Ry_{s_2}y_{s_1}}{\|\mathbf{y}\|}\frac{I_{\frac{n}{2}}(R\|\mathbf{y}\|)}{I_{\frac{n}{2}-1}(R\|\mathbf{y}\|)}\\ &+\frac{R^2y_{s_2}y_{s_1}}{\|\mathbf{y}\|^2}\left(1-\frac{n-1}{R\|\mathbf{y}\|}\frac{I_{\frac{n}{2}}(R\|\mathbf{y}\|)}{I_{\frac{n}{2}-1}(R\|\mathbf{y}\|)}-\left(\frac{I_{\frac{n}{2}}(R\|\mathbf{y}\|)}{I_{\frac{n}{2}-1}(R\|\mathbf{y}\|)}\right)^2\right), \qquad (186)\end{aligned}$$

where the last equality follows by using the derivative expression in (183).

The proof for $s_1 = s_2$ follows along the similar lines.

APPENDIX E
PROOF OF THEOREM 3

It is sufficient to characterize only the second order partial derivatives $\frac{\partial}{\partial y_j\partial y_i}K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})$ and then apply a simple induction.

Let $U = \mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}$. Then, by using Theorem 1, we have that the gradient of the cumulant generating function can be expressed as

$$\begin{aligned}&\nabla_{\mathbf{y}}K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})\\ &= \frac{\nabla_{\mathbf{y}}\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]}{\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]} \qquad (187)\\ &= \frac{\mathbf{K}_{\mathbf{N}}^{-1}\mathbf{Cov}(\mathbf{X},U|\mathbf{Y}=\mathbf{y})}{\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]} \qquad (188)\\ &= \frac{\mathbf{K}_{\mathbf{N}}^{-1}\Big(\mathbb{E}[\mathbf{X}\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]-\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]\Big)}{\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]} \qquad (189)\\ &= \mathbf{K}_{\mathbf{N}}^{-1}\left(\nabla_{\mathbf{t}}\log\Big(\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]\Big)-\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\right) \qquad (190)\\ &= \mathbf{K}_{\mathbf{N}}^{-1}\nabla_{\mathbf{t}}K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})-\mathbf{K}_{\mathbf{N}}^{-1}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}], \qquad (191)\end{aligned}$$

where in (190) we have used that $\frac{\mathbb{E}[\mathbf{X}\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]}{\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]} = \frac{\nabla_{\mathbf{t}}\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]}{\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]} = \nabla_{\mathbf{t}}\log\Big(\mathbb{E}[\mathrm{e}^{\mathbf{t}^\mathsf{T}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]\Big)$.

Consequently, the partial derivative with respect to i is given by

$$\frac{\partial}{\partial y_i}K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y}) = \mathbf{k}_i^\mathsf{T}\nabla_{\mathbf{t}}K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})-\mathbf{k}_i^\mathsf{T}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}], \tag{192}$$

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

where we have used that $\mathbf{k}_i^{\mathsf{T}} = \mathbf{e}_i^{\mathsf{T}}\mathbf{K}_{\mathbf{N}}^{-1}$ (i.e., the i-th row of $\mathbf{K}_{\mathbf{N}}^{-1}$).

Next, by differentiating with respect to j, we arrive at

$$\frac{\partial^2}{\partial y_j \partial y_i} K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y}) = \frac{\partial}{\partial y_j}\left(\mathbf{k}_i^{\mathsf{T}}\nabla_{\mathbf{t}} K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y}) - \mathbf{k}_i^{\mathsf{T}}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\right) \tag{193}$$

$$= \mathbf{k}_i^{\mathsf{T}}\nabla_{\mathbf{t}}\mathbf{k}_j^{\mathsf{T}}\nabla_{\mathbf{t}} K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y}) - \mathbf{k}_i^{\mathsf{T}}\frac{\partial}{\partial y_j}\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}], \tag{194}$$

where (194) comes from using the partial derivative formula in (192).

Appendix F
Proof of Proposition 8

To see this note that $\frac{\partial^j K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})}{\partial y_{s_1}\ldots\partial y_{s_j}}|_{\mathbf{t}=\mathbf{0}} = 0$ for every j

$$\frac{\partial^j K_{\mathbf{X}}(\mathbf{t}|\mathbf{Y}=\mathbf{y})}{\partial y_{s_1}\ldots\partial y_{s_j}} = \frac{\partial^j \log\left(\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]\right)}{\partial y_{s_1}\ldots\partial y_{s_j}} \tag{195}$$

$$= \frac{\partial^{j-1}}{\partial y_{s_1}\ldots\partial y_{s_{j-1}}}\frac{\frac{\partial}{\partial y_{s_j}}\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]}{\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]} \tag{196}$$

$$= \frac{g(\mathbf{y})}{(\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}])^{\max(1,2(j-1))}}. \tag{197}$$

We do not attempt to find the exact expression for $g(\mathbf{y})$ and only observe that it contains terms of the form $\frac{\partial^k}{\partial y_{s_1}\ldots\partial y_{s_k}}\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]$ for some $k \geq 1$. We next show that $\frac{\partial^k}{\partial y_{s_1}\ldots\partial y_{s_k}}\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]$ evaluated at $\mathbf{t}=\mathbf{0}$ is equal to zero, which through (197) will lead to the desired conclusion. To do this we use the Taylor expression for the multivariate moment generating function given by [25, Ch. 6]

$$\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}] = 1 + \sum_{j=1}^{\infty}\sum_{\mathbf{s}\in S(j)}\frac{1}{j!}\mu(\mathbf{y};s_1,\ldots,s_j)t_{s_1}\ldots t_{s_j}, \tag{198}$$

where $\mu(\mathbf{y};s_1,\ldots,s_j) = \mathbb{E}[X_{s_1}\ldots X_{s_j}|\mathbf{Y}=\mathbf{y}]$ and $S(j) = \{1,\ldots,n\}^j$ is set of all vectors of integers with j components and all entries between 1 and n. Now we have that

$$\frac{\partial^k}{\partial y_{s_1}\ldots\partial y_{s_k}}\mathbb{E}[\mathrm{e}^{\mathbf{t}^{\mathsf{T}}\mathbf{X}}|\mathbf{Y}=\mathbf{y}]\Big|_{\mathbf{t}=\mathbf{0}} = \sum_{j=1}^{\infty}\sum_{\mathbf{s}\in S(j)}\frac{1}{j!}\frac{\partial^k}{\partial y_{s_1}\ldots\partial y_{s_k}}\mu(\mathbf{y};s_1,\ldots,s_j)t_{s_1}\ldots t_{s_j}\Big|_{\mathbf{t}=\mathbf{0}} = 0. \tag{199}$$

This concludes the proof.

Appendix G
Proof of Proposition 12

The proof of (105) follows by applying Hatsell and Nolte identity in (17) to the expressions in Proposition 10 together with the identity $\mathbf{D}_{\mathbf{x}}^2 f(\mathbf{x}) = \mathbf{J}_{\mathbf{x}}\nabla_{\mathbf{x}} f(\mathbf{x})$. To show the general case observe that

$$\mathbf{D}_{\mathbf{y}}^2 \log\left(\mathbb{P}[\mathbf{X}\in\mathcal{A}|\mathbf{Y}=\mathbf{y}]\right)$$

$$= \mathbf{J}_{\mathbf{y}}\nabla_{\mathbf{y}} \log\left(\mathbb{P}[\mathbf{X}\in\mathcal{A}|\mathbf{Y}=\mathbf{y}]\right) \tag{200}$$

$$= \mathbf{J}_{\mathbf{y}}\mathbf{K}_{\mathbf{N}}^{-1}(\mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}] - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]) \tag{201}$$

$$= \mathbf{J}_{\mathbf{y}}\left(\frac{\mathbb{E}[\mathbf{X}1_{\mathcal{A}}(\mathbf{X})|\mathbf{Y}=\mathbf{y}]}{\mathbb{P}[\mathbf{X}\in\mathcal{A}|\mathbf{Y}=\mathbf{y}]} - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\right)\mathbf{K}_{\mathbf{N}}^{-1} \tag{202}$$

$$= \Big(\frac{\mathbf{K}_{\mathbf{N}}^{-1}\mathbf{Cov}(\mathbf{X},\mathbf{X}1_{\mathcal{A}}(\mathbf{X})|\mathbf{Y}=\mathbf{y})}{\mathbb{P}[\mathbf{X}\in\mathcal{A}|\mathbf{Y}=\mathbf{y}]} - \frac{\mathbf{K}_{\mathbf{N}}^{-1}\mathbf{Cov}(\mathbf{X},1_{\mathcal{A}}(\mathbf{X})|\mathbf{Y}=\mathbf{y})\mathbb{E}^{\mathsf{T}}[\mathbf{X}1_{\mathcal{A}}(\mathbf{X})|\mathbf{Y}=\mathbf{y}]}{\mathbb{P}^2[\mathbf{X}\in\mathcal{A}|\mathbf{Y}=\mathbf{y}]} - \mathbf{K}_{\mathbf{N}}^{-1}\mathbf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y})\Big)\mathbf{K}_{\mathbf{N}}^{-1} \tag{203}$$

$$= \mathbf{K}_{\mathbf{N}}^{-1}\Big(\mathbb{E}[\mathbf{X}\mathbf{X}^{\mathsf{T}}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}] - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}]\mathbb{E}[\mathbf{X}^{\mathsf{T}}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}] - \mathbf{K}_{\mathbf{N}}^{-1}\mathbf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y})\Big) \tag{204}$$

$$= \mathbf{K}_{\mathbf{N}}^{-1}\left(\mathbf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}) - \mathbf{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y})\right)\mathbf{K}_{\mathbf{N}}^{-1}, \tag{205}$$

where (200) follows by using the identity $\mathbf{D}_{\mathbf{x}}^2 f(\mathbf{x}) = \mathbf{J}_{\mathbf{x}}\nabla_{\mathbf{x}} f(\mathbf{x})$; (201) follows by using (99); (202) follows by using the property that $\mathbf{J}_{\mathbf{x}}\mathbf{K}_{\mathbf{N}}^{-1} f(\mathbf{x}) = \mathbf{J}_{\mathbf{x}} f(\mathbf{x})\mathbf{K}_{\mathbf{N}}^{-\mathsf{T}} = \mathbf{J}_{\mathbf{x}} f(\mathbf{x})\mathbf{K}_{\mathbf{N}}^{-1}$; (203) follows by using identity in (15) with $\mathbf{U} = \mathbf{X}1_{\mathcal{A}}(\mathbf{X})$ and the quotient rule for differentiation; (204) follows by rewriting the first covariance term as

$$\frac{\mathbf{Cov}(\mathbf{X},\mathbf{X}1_{\mathcal{A}}(\mathbf{X})|\mathbf{Y}=\mathbf{y})}{\mathbb{P}[\mathbf{X}\in\mathcal{A}|\mathbf{Y}=\mathbf{y}]} = \mathbb{E}[\mathbf{X}\mathbf{X}^{\mathsf{T}}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}] - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\mathbb{E}[\mathbf{X}^{\mathsf{T}}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}], \tag{206}$$

and the second covariance term as

$$\frac{\mathbf{Cov}(\mathbf{X},1_{\mathcal{A}}(\mathbf{X})|\mathbf{Y}=\mathbf{y})\mathbb{E}^{\mathsf{T}}[\mathbf{X}1_{\mathcal{A}}(\mathbf{X})|\mathbf{Y}=\mathbf{y}]}{\mathbb{P}^2[\mathbf{X}\in\mathcal{A}|\mathbf{Y}=\mathbf{y}]} = \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}]\mathbb{E}[\mathbf{X}^{\mathsf{T}}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}] - \mathbb{E}[\mathbf{X}|\mathbf{Y}=\mathbf{y}]\mathbb{E}[\mathbf{X}^{\mathsf{T}}|\mathbf{Y}=\mathbf{y},\mathbf{X}\in\mathcal{A}]; \text{ and} \tag{207}$$

and (205) follows from the definition of conditional variance. This concludes the proof.

References

[1] A. Dytso, H. V. Poor, and S. Shamai (Shitz), "A general derivative identity for the conditional mean estimator in Gaussian noise and some applications," in *Proc. IEEE Int. Symp. Inf. Theory*. Los Angeles, CA, USA: IEEE, 2020, pp. 1183–1188.

[2] A. Dytso, H. V. Poor, and S. Shamai (Shitz), "On the distribution of the conditional mean estimator in Gaussian noise," in *Proc. IEEE Inf. Theory Workshop*, Riva del Garda, Italy, Apr. 2021, pp. 1–5.

[3] A. Banerjee, X. Guo, and H. Wang, "On the optimality of conditional expectation as a Bregman predictor," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2664–2669, Jul. 2005.

[4] D. Guo, S. Shamai, and S. Verdú, *The Interplay Between Information and Estimation Measures*. Boston, MA, USA: Now, 2013.

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

[6] B. Efron. (2005). *Local False Discovery Rates*. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.179.1376&rep=rep1&type=pdf

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.

[7] B. Efron and T. Hastie, *Computer Age Statistical Inference*, vol. 5. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[8] A. G. Jaffer and S. C. Gupta, "On relations between detection and estimation of discrete time processes," *Inf. Control*, vol. 20, no. 1, pp. 46–54, 1972.
[9] H. Robbins, "An empirical Bayes approach to statistics," in *Proc. 3rd Berkeley Symp. Math. Statist. Probab.*, 1956, pp. 388–394.
[10] R. Esposito, "On a relation between detection and estimation in decision theory," *Inf. Control*, vol. 12, no. 2, pp. 116–120, Feb. 1968.
[11] A. Dytso, M. Al, H. V. Poor, and S. Shamai (Shitz), "On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3907–3921, 2019.
[12] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[13] C. Hatsell and L. Nolte, "Some geometric properties of the likelihood ratio (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 5, pp. 616–618, Sep. 1971.
[14] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.
[15] T. T. Cai and Y. Wu, "Optimal detection of sparse mixtures against a given null distribution," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2217–2232, Apr. 2014.
[16] A. Atalik, A. Köse, and M. Gastpar, "Differential entropy of the conditional expectation under Gaussian noise," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2021, pp. 1–6.
[17] A. Atalik, A. Köse, and M. Gastpar, "The price of distributed: Rate loss in the CEO problem," in *Proc. Conf. Inf. Sci. Sys.*, 2022, pp. 125–130.
[18] W. Alghamdi and F. P. Calmon, "Polynomial approximations of conditional expectations in scalar Gaussian channels," 2021, *arXiv:2102.05970*.
[19] A. Jaffer, "A note on conditional moments of random signals in Gaussian noise (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 513–514, Jul. 1972.
[20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Table* (National Bureau of Standards Applied Mathematics Series), vol. 55. Washington, DC, USA: US Government Printing Office, 1965.
[21] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.
[22] C. A. Charalambides, *Enumerative Combinatorics*. Boca Raton, FL, USA: CRC Press, 2018.
[23] J. Arbel, O. Marchal, and H. D. Nguyen, "On strict sub-gaussianity, optimal proxy variance and symmetry for bounded random variables," 2019, *arXiv:1901.09188*.
[24] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*. Cham, Switzerland: Springer, 2002.
[25] J. E. Kolassa, *Series Approximation Methods in Statistics*, vol. 88. Cham, Switzerland: Springer, 2006.
[26] J. C. Berry, "Minimax estimation of a bounded normal mean vector," *J. Multivariate Anal.*, vol. 35, no. 1, pp. 130–139, 1990.
[27] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*. Hoboken, NJ, USA: Wiley, 2013.
[28] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. Hoboken, NJ, USA: Wiley, 1995.
[29] A. Dytso, H. V. Poor, R. Bustin, and S. Shamai (Shitz), "On the structure of the least favorable prior distributions," in *Proc. IEEE Int. Symp. Inf. Theory*. Vail, CO, USA: IEEE, Jun. 2018, pp. 1081–1085.
[30] I. Zieder, A. Dytso, and M. Cardone, "An MMSE lower bound via Poincaré inequality," in *Proc. IEEE Int. Symp. Inf. Theory*, May 2022, pp. 957–962.
[31] A. Dytso and M. Cardone, "A general derivative identity for the conditional expectation with focus on the exponential family," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2021, pp. 1–6.
[32] R. Bustin, M. Payaró, D. P. Palomar, and S. Shamai (Shitz), "On MMSE crossing properties and implications in parallel vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 818–844, Feb. 2013.
[33] A. Dubkov and A. Malakhov, "Properties and interdependence of the cumulants of a random variable," *Radiophysics Quantum Electron.*, vol. 19, no. 8, pp. 833–839, 1976.
[34] M. Fozunbal, "On regret of parametric mismatch in minimum mean square error estimation," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 1408–1412.
[35] G. N. Watson, *A Treatise Theory Bessel Functions*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

Alex Dytso (Member, IEEE) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL, USA, in 2016. From September 2016 to August 2020, he was a Post-Doctoral Associate with the Department of Electrical Engineering, Princeton University. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology (NJIT). His current research interests include multi-user information theory and estimation theory, and their applications in wireless networks.

H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer science (EECS) from Princeton University in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana–Champaign. Since 1990, he has been on the faculty at Princeton, where he is currently the Michael Henry Strater University Professor. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other universities, including most recently at Berkeley and Cambridge. His research interests include information theory, machine learning, and network science and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the recent book *Machine Learning and Wireless Communications* (Cambridge University Press, 2022). Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences and a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. He received the IEEE Alexander Graham Bell Medal in 2017.

Shlomo Shamai (Shitz) (Life Fellow, IEEE) is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is currently a Technion Distinguished Professor, and holds the William Fondiller Chair of telecommunications. He is also a URSI Fellow, a member of the Israeli Academy of Sciences and Humanities, and a Foreign Member of the U.S. National Academy of Engineering. He was a recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering, and the 2017 IEEE Richard W. Hamming Medal, and numerous technical and paper awards and recognitions of the IEEE (Donald G. Fink Prize Paper Award), Information Theory, Communications and Signal Processing Societies as well as EURASIP. He was also a co-recipient of the 2018 Third Bell Labs Prize for Shaping the Future of Information and Communications Technology. He is listed as a Highly Cited Researcher (Computer Science) for the year 2013–2018. He has served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY. He has also served twice on the Board of Governors of the Information Theory Society and an Executive Editorial Board for the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE Information Theory Society Nominations and Appointments Committee, and the IEEE Information Theory Society, Shannon Award Committee.

Authorized licensed use limited to: Princeton University. Downloaded on August 23,2023 at 17:43:05 UTC from IEEE Xplore. Restrictions apply.