On the Minimum Mean pth Error in Gaussian Noise Channels and Its Applications

Alex Dytso, Ronit Bustin, *Member, IEEE*, Daniela Tuninetti, Natasha Devroye, H. Vincent Poor, *Fellow, IEEE*, and Shlomo Shamai (Shitz), *Fellow, IEEE*

***Abstract*— The problem of estimating an arbitrary random vector from its observation corrupted by additive white Gaussian noise, where the cost function is taken to be the minimum mean pth error (MMPE), is considered. The classical minimum mean square error (MMSE) is a special case of the MMPE. Several bounds, properties, and applications of the MMPE are derived and discussed. The optimal MMPE estimator is found for Gaussian and binary input distributions. Properties of the MMPE as a function of the input distribution, signal-to-noise-ratio (SNR) and order p are derived. The "single-crossing-point property" (SCPP) which provides an upper bound on the MMSE, and which together with the mutual information-MMSE relationship is a powerful tool in deriving converse proofs in multi-user information theory, is extended to the MMPE. Moreover, a complementary bound to the SCPP is derived. As a first application of the MMPE, a bound on the conditional differential entropy in terms of the MMPE is provided, which then yields a generalization of the Ozarow–Wyner lower bound on the mutual information achieved by a discrete input on a Gaussian noise channel. As a second application, the MMPE is shown to improve on previous characterizations of the phase transition phenomenon that manifests, in the limit as the length of the capacity achieving code goes to infinity, as a discontinuity of the MMSE as a function of SNR. As a final application, the MMPE is used to show new bounds on the second derivative of mutual information, or the first derivative of the MMSE.**

***Index Terms*— I-MMSE, estimation.**

Manuscript received July 5, 2016; revised August 7, 2017; accepted November 18, 2017. Date of publication December 13, 2017; date of current version February 15, 2018. A. Dytso was supported in part by the U.S. National Science Foundation (NSF) under Grants CNS-1456793, CCF-1420575, and CCF-1422511. R. Bustin was supported in part by the Women Postdoctoral Scholarship of Israel's Council for Higher Education (VATAT) 2014-2015 and in part by Union's Horizon 2020 Research and Innovation Programme (2016) under Grant 694630. D. Tuninetti and N. Devroye were supported by the NSF under Grant CCF-1422511. H. V. Poor was supported by the NSF under Grant CNS-1456793 and Grant CCF-1420575. The work of S. Shamai (Shitz) was supported in part by the Israel Science Foundation, in part by the S. and N. Grand Research Fund, and in part by the Union's Horizon 2020 Research and Innovation Programme under Grant 694630. This paper was presented in part at the 2016 IEEE Information Theory Workshop and in part at the 2016 IEEE International Symposium on Information Theory.

A. Dytso and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: adytso@princeton.edu; poor@princeton.edu).

R. Bustin and S. Shamai (Shitz) are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: ronitbustin@post.tau.ac.il; sshlomo@ee.technion.ac.il).

D. Tuninetti and N. Devroye are with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: danielat@uic.edu; devroye@uic.edu).

Communicated by M. K. Varanasi, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2017.2782786

I. Introduction

In the Bayesian setting the Minimum Mean Square Error (MMSE) of estimating a random variable X from an observation Y is understood as a cost function[1] with a quadratic loss function (i.e., L_2 norm):

$$\mathrm{mmse}(X \mid Y) = \mathbb{E}\left[|X - \mathbb{E}[X \mid Y]|^2\right]. \quad (1)$$

Another commonly used cost function is the L_1 norm with loss function given by the absolute value of the error (i.e., the difference between the variable of interest and its estimate). In general, cost functions with non-quadratic loss functions are not well understood and have been considered only for special cases, such as under the assumption of Gaussian statistics.

The interplay between estimation theoretic and information theoretic measures has been very fruitful; for example the so called mutual information-MMSE (I-MMSE) relationship [3], that relates the derivative of the mutual information with respect to the Signal-to-Noise-Ratio (SNR) to the MMSE, has found numerous applications throughout information theory [4]. The goal of this work is to show that the study of estimation problems with non-quadratic loss functions can also offer new insights into classical information theoretic problems. *The program of this paper is thus to develop the necessary theory for a class of loss functions, and then apply the developed tools to information theoretic problems.*

A. Past Work

The popularity of the MMSE stems from its analytical tractability, which is rooted in the fact that the MMSE is defined through the L_2 norm in (1). The L_2 norm, in turn, allows applications of the well understood Hilbert space theory [5]. In information theoretic applications the L_2 norm is used, for example, to define an average input power constraint. The connection between the power constraint and the L_2 norm leads to a continuous analog of Fano's inequality that relates the conditional differential entropy and the MMSE [6, Th. 8.6.6].

Recently, in view of the I-MMSE relationship [3], the MMSE (in an Additive White Gaussian Noise (AWGN) channel) has received considerable attention. For example, in [7] the I-MMSE relationship was used to give a simple alternative proof of the Entropy Power Inequality (EPI) [8]. Moreover, the

[1] Another common term used is a risk function.

so called 'Single-Crossing-Point Property' (SCPP) [9], [10] that bounds the MMSE for all SNR values *above* a certain value at which the MMSE is known, together with the I-MMSE relationship, offers an alternative, unifying framework for deriving information theoretic converses, such as: [9] to provide an alternative proof of the converse for the Gaussian broadcast channel (BC) and show a special case of the EPI; in [11] to provide a simple proof for the information combining problem and a converse for the BC with confidential messages; in [10], by using various extensions of the SCPP, to prove a special case of the vector EPI, a converse for the capacity region of the parallel degraded BC under per-antenna power constraints and under an input covariance constraint, and a converse for the compound parallel degraded BC under an input covariance constraint; and in [12] to provide a converse for communication under an MMSE disturbance constraint.

In [13] we demonstrated a bound that complements the SCPP, that bounds the MMPE for all SNR values *below* a certain value at which the MMSE is known, and allows for a finer characterization of the *phase transition* phenomenon that manifests as a discontinuity of the MMSE as a function of SNR, as the length of the codeword goes to infinity. This plays an important role in characterizing achievable rates of the capacity achieving codes [14], [15]. One of the applications of the tools presented in this work is an improvement on the bound in [13, Th. 1].

Many other properties of the MMSE in relation to the I-MMSE have been studied in [9] and [16]–[18]. For a comprehensive survey on results, applications and extensions of the I-MMSE relationship we refer the reader to [11] and [19].

While the MMSE has received considerable attention and is well understood, non-quadratic cost functions are only understood in special cases, such as under the assumption of Gaussian statistics. For example, in [20] it was shown that under scalar Gaussian statistics, for a large class of symmetric loss functions the optimal linear MMSE (LMMSE) estimator is also optimal. The result of [20] was extended in [21] to a large class of cost functions that also include asymmetric loss functions. Other early work in this direction includes also [22].

Tan *et al.* [23] studied the expected L_∞ norm of the error, when the input is assumed to be a Gaussian mixture. The authors showed that, as the dimension of the signal goes to infinity, the optimal LMMSE estimator minimizes the expected maximum error.

Hall and Wise [24], [25] studied a class of *even and non-decreasing* and *even and convex*, respectively, loss functions and gave a sufficient condition on the conditional distribution of the input X given the output Y, so that the conditional expectation $\mathbb{E}[X|Y]$ is the optimal estimator.

Akyol *et al.* [26] studied a scalar additive noise channel and an L_p cost function and showed a necessary and sufficient condition on the noise and the input distributions to guarantee that the optimal estimator is linear. Moreover, under the derived sufficient and necessary conditions, if the source and noise variances are the same, then the optimal estimator is linear if and only if the input and the noise distributions are identical.

Weinberger and Merhav [27] and Merhav [28] considered the problem of transmitting a modulated signal over a discrete memoryless channel where the performance criterion was taken to be the L_p cost function. To that end, the authors showed tight exponential bounds for very small and very large values of p.

Saerens [29] focused on designing an appropriate cost function such that the output of the trained model approximates the desired summary statistics, such as the conditional expectation, the geometric mean or the variance.

Livadiotis [30] focus on characterizing expectation and variance based on L_p norms and emphasized that a parameter p provides a new degree of freedom in analyzing of new phenomena in statistical physics. The interested reader is also referred to [31] where the interplay between the L_p means and means generalized in terms convex functions is considered.

In non-Bayesian estimation L_p cost functions have been considered in [32] and [33], in a context of *minimax* estimation, and the authors gave lower and upper bounds on the exponential behavior of the cost function. For a non-Bayesian treatment of non-quadratic cost functions we refer the reader to [34] and [35].

Looking into non-quadratic cost functions is further motivated by the fact that often the quadratic cost function may not be the correct measure of signal fidelity for certain applications. This is especially true in image processing where error metrics, more sensitive to structural changes of the input signal, better capture human perceptions of quality. We refer the reader to [36] for a survey of recent results in this direction.

B. Paper Outline and Main Contributions

In this work we are interested in studying a cost function, termed the Minimum Mean p-th Error (MMPE),[2] the scalar version of which is given by

$$\mathrm{mmpe}(X \mid Y; p) = \inf_f \mathbb{E}\left[|X - f(Y)|^p\right], \tag{2}$$

where the infimum is over all measurable estimators $f(Y)$.

Our contributions are as follows:

1) In Section II we formally define the vector version of the MMPE in (2) and introduce related definitions.
2) In Section III we study properties of the optimal MMPE estimator and show:
   - In Section III-A, Proposition 1 shows that the MPPE optimal estimator indeed exists;
   - In Section III-B, Proposition 2 derives an *orthogonality-like principle* that serves as a necessary and sufficient condition for an estimator to be MMPE optimal;
   - Section III-C gives examples of optimal MMPE estimators. In particular, in Proposition 3 we find the MMPE for Gaussian random vectors, and in Proposition 4 for discrete binary random variables; and
   - In Section III-D, Proposition 5 shows some basic properties of the optimal MMPE estimator in terms

[2] The abbreviation MMPE has been used before in [11, Ch. 8] for the Minimum Mean Poisson Error.

of input distribution, such as, *linearity, stability, degradedness*, etc. Moreover, via an example it is shown that in general the MMPE optimal estimator is biased on average (i.e., the first moment of the error (bias) is not zero). However, it is shown that the p-th order estimator is unbiased on average in the sense that the $(p-1)$-th moment of the error is zero.

3) In Section IV we study properties of the MMPE as a function of order p, SNR and the input distribution that will be useful in a number of applications:
   - In Section IV-A, Proposition 6 shows that the MMPE is invariant under translations of the input random vector and derives basic scaling properties;
   - In Section IV-B, Proposition 7 shows that, as far as estimation error over the channel $\mathbf{Y} = \sqrt{\mathsf{snr}}\mathbf{X}+\mathbf{Z}$ is concerned the estimation of the input $\mathbf{X}$ is equivalent to the estimation of the Gaussian noise $\mathbf{Z}$; and
   - In Section IV-C, Proposition 8 gives a 'change of measure' result that allows one to take the expectation in the definition of the MMPE with respect to an output at a different SNR.
4) In Section V we discuss basic bounds on the MMPE and show:
   - In Section V-A, Proposition 10 develops basic ordering bounds between MMPE's of different orders and bounds equivalent to that of the LMMSE bound;
   - In Section V-B, Proposition 11 shows that, under an appropriate moment constraint on the input distribution, the Gaussian input is asymptotically the 'hardest' to estimate;
   - In Section V-C, Proposition 12 derives interpolation bounds for the MMPE. One of the consequences of such bounds is Proposition 13, which shows that the MMPE is a continuous function of order p; and
   - In Section V-D, Proposition 15 derives bounds on the MMPE with discrete vector inputs.
5) In Section VI we define the conditional MMPE and show:
   - Proposition 16 shows that conditioning reduces the MMPE; and
   - Proposition 17 shows that the MMPE estimation of $\mathbf{X}$ from two AWGN observations is equivalent to estimating $\mathbf{X}$ from a single observation with a higher SNR. This implies that the MMPE is a decreasing function of SNR.
6) In Section VII we show applications of the developed tools:
   - In Proposition 18, by using the tools developed for the conditional MMPE, a simple proof of the SCPP for the MMSE is given, and extended to the MMPE;
   - In Proposition 19 we use the change of measure result in Proposition 8 to show a bound that complements the SCPP bound, that is, it bounds the MMPE for all SNR values *below* a certain SNR value at which the MMPE is known; and
   - In Proposition 20, by using change of measure result in Proposition 19 and continuity of the MMPE in p from Proposition 13, we show that for any finite dimensional input the MMPE is a continuos function of SNR.
7) In Section VIII we apply the developed bounds and generalize or improve some well known information theoretic MMSE bounds:
   - In Section VIII-A, Theorem 1 gives a general inequality that bounds the conditional differential entropy via the MMPE of which the continuous analog of Fano's from [6, Th. 8.6.6] is a special case;
   - In Section VIII-B, Theorem 2 generalizes the Ozarow-Wyner bound [37] on the mutual information achieved by a discrete input on an AWGN channel, to vector discrete inputs and yields the sharpest known version of this bound. Moreover, in Theorem 3 we show how the bound behaves as the dimension of the input goes to infinity;
   - In Section VIII-C, Theorem 4 improves on the previous characterizations of the width of the phase transition region of finite-length code of length n given by $O(\frac{1}{n})$ in [13] to $O(\frac{1}{\sqrt{n}})$. This in turn also improves the converse result on the communications under disturbance constrained problem studied in [13]; and
   - In Section VIII-D, Proposition 21 we show how the MMPE can be used to provide new lower and upper bounds on the derivative of the MMSE.

C. Notation

Throughout the paper we adopt the following notational conventions:

- Deterministic scalar and vector quantities are denoted by lower case and bold lower case letters, respectively. Matrices are denoted by bold upper case letters;
- Random variables and vectors are denoted by upper case and bold upper case letters, respectively, where r.v. is short for either random variable or random vector, which should be clear from the context;
- If A is a r.v. we denote the support of its distribution by $\mathsf{supp}(A)$;
- The symbol $|\cdot|$ may denote different things: $|\mathbf{A}|$ is the determinant of the matrix $\mathbf{A}$, $|\mathcal{A}|$ is the cardinality of the set $\mathcal{A}$, $|X|$ is the cardinality of $\mathsf{supp}(X)$, or $|x|$ is the absolute value of the real-valued x;
- The symbol $\|\cdot\|$ denotes the Euclidian norm;
- $\mathbb{E}[\cdot]$ denotes the expectation operator;
- We denote the covariance of r.v. $\mathbf{X}$ by $\mathbf{K}_{\mathbf{X}}$;
- $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_{\mathbf{X}}, \mathbf{K}_{\mathbf{X}})$ denotes the density of a real-valued Gaussian r.v. $\mathbf{X}$ with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{K}_{\mathbf{X}}$;
- The identity matrix is denoted by $\mathbf{I}$;
- Reflection of the matrix $\mathbf{A}$ along its main diagonal, or the transpose operation, is denoted by $\mathbf{A}^T$;
- The trace operation on the matrix $\mathbf{A}$ is denoted by $\mathrm{Tr}(\mathbf{A})$;
- The Order notation $\mathbf{A} \succeq \mathbf{B}$ implies that $\mathbf{A}-\mathbf{B}$ is a positive semidefinite matrix;
- $\log(\cdot)$ denotes logarithms in base 2;

- $[n_1 : n_2]$ is the set of integers from n_1 to $n_2 \ge n_1$;
- For $x \in \mathbb{R}$ we let $\lfloor x \rfloor$ denote the largest integer not greater than x;
- For $x \in \mathbb{R}$ we let $[x]^+ := \max(x, 0)$ and $\log^+(x) := [\log(x)]^+$;
- Let $f(x), g(x)$ be two real-valued functions. We use the Landau notation $f(x) = O(g(x))$ to mean that for *some* $c > 0$ there exists an x_0 such that $f(x) \le c\, g(x)$ for all $x \ge x_0$, and $f(x) = o(g(x))$ to mean that for *every* $c > 0$ there exists an x_0 such that $f(x) < cg(x)$ for all $x \ge x_0$;
- We denote the conditional r.v. $\mathbf{X}|\mathbf{Y} = \mathbf{y} \sim p_{\mathbf{X}|\mathbf{Y}}(\cdot|\mathbf{y})$ as $\mathbf{X}_{\mathbf{y}}$;
- We denote the upper incomplete gamma function and the gamma function by

$$\Gamma(x; a) := \int_a^\infty t^{x-1} e^{-t} dt, \ x \in \mathbb{R}, a \in \mathbb{R}^+, \quad \text{(3a)}$$

$$\Gamma(x) := \Gamma(x; 0). \quad \text{(3b)}$$

The generalized Q-function is denoted by

$$\bar{Q}(x; a) := \frac{\Gamma(x; a)}{\Gamma(x)}. \quad \text{(3c)}$$

In particular, the generalized Q-function can be related to the standard Q-function, by using the relationship $Q(\sqrt{2}x) = \frac{1}{2\sqrt{\pi}}\Gamma\left(\frac{1}{2}; x^2\right)$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, as $\bar{Q}\left(\frac{1}{2}; a^2\right) = 2Q(\sqrt{2}a)$; and
- We define the volume of the region S embedded in $\mathbb{R}^n$ as

$$\text{Vol}(S) := \int_S 1 \ dx_1 dx_2 \cdots dx_n. \quad (4)$$

In particular, the volume of the n-dimensional ball $B(r)$ of radius r centered at the origin is given by

$$\text{Vol}(B(r)) = \frac{\pi^{\frac{n}{2}} r^n}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

II. Cost Function Definition

Motivated by the study of cost functions with non-quadratic error we define the following norm.

Definition 1: For the r.v. $\mathbf{U} \in \mathbb{R}^n$ *and* $p > 0$

$$\|\mathbf{U}\|_p := \left(\frac{1}{n}\mathbb{E}\left[\|\mathbf{U}\|^p\right]\right)^{\frac{1}{p}} = \left(\frac{1}{n}\mathbb{E}\left[\left(\text{Tr}(\mathbf{U}\mathbf{U}^T)\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}. \quad (5)$$

For $p \ge 1$ the function in (5) defines a norm and obeys the triangle inequality

$$\|\mathbf{U} + \mathbf{V}\|_p \le \|\mathbf{U}\|_p + \|\mathbf{V}\|_p. \quad (6)$$

Therefore, throughout the paper we define the L_p space, for $p \ge 1$, as the space of r.v. on a fixed probability space $(\Omega, \sigma(\Omega), \mathbb{P})$ such that the norm in (5) is finite. However, many of our results will hold for $0 < p < 1$, for which (5) is not a norm.

In particular, for $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ the norm in (5) is given by

$$n\|\mathbf{Z}\|_p^p = \mathbb{E}\left[\left(\sum_{i=1}^n Z_i^2\right)^{\frac{p}{2}}\right] = 2^{\frac{p}{2}} \frac{\Gamma\left(\frac{n+p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \quad \text{for } n \in \mathbb{N}, p \ge 0, \quad (7)$$

and for $\mathbf{V}$ uniform over the n dimensional ball of radius r the norm in (5) is given by

$$\begin{aligned} n\|\mathbf{V}\|_p^p &= \frac{1}{\text{Vol}(B(r))} \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_0^r \rho^p \rho^{n-1} d\rho \\ &= \frac{n}{2p + 2n} r^p, \quad \text{for } n \in \mathbb{N}, p \ge 0. \end{aligned} \quad (8)$$

Note that for $n = 1$ we have that $\|U\|_p^p = \mathbb{E}\left[|U|^p\right]$ and therefore from now on we will refer to $\|\mathbf{U}\|_p^p$ as p-th moment of $\mathbf{U}$. Naturally, for $n > 1$, there are many other ways for defining the moments, see for example [38]. However, in view of the information theoretic problems we are interested in, such for example from previous work [13], the definition in (5) arises naturally.

Definition 2: For any $p > 0$*, we define the minimum mean* p*-th error (MMPE) of estimating* $\mathbf{X}$ *from* $\mathbf{Y}$ *as*

$$\text{mmpe}(\mathbf{X}|\mathbf{Y}; p) := \inf_f \|\mathbf{X} - f(\mathbf{Y})\|_p^p, \quad (9)$$

and where the minimization is over all possible Borel measurable functions $f(\mathbf{Y})$*. Whenever the optimal MMPE estimator exists and is unique (up to a set of measure zero) we shall denote it by* $f_p(\mathbf{X}|\mathbf{Y})$*.*[3] *The optimal estimator in* (9) *might not be unique (i.e., there could be two or more estimators that do not agree on a set of a positive measure) in which case we define*

$$f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y}) = \sup\{\|f(\mathbf{y})\| : f(\cdot) \text{ is a minimizer in (9)}\}. \quad (10)$$

Remark 1: The notation $f_p(\mathbf{X}|\mathbf{Y})$*, for the optimal estimator in* (9) *is inspired by the conditional expectation* $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$*, and* $f_p(\mathbf{X}|\mathbf{Y})$ *should be thought of as an operator on* $\mathbf{X}$ *and a function of* $\mathbf{Y}$*. Indeed, for* $p = 2$*, the MMPE reduces to the MMSE; that is,* $\text{mmpe}(\mathbf{X}|\mathbf{Y}; 2) = \text{mmse}(\mathbf{X}|\mathbf{Y})$ *and* $f_2(\mathbf{X}|\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$*. The properties of* $f_p(\mathbf{X}|\mathbf{Y})$ *as an operator on* $\mathbf{X}$ *will be investigated in Proposition 5.*

Finally, similarly to the conditional expectation, the notation $f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$ *should be understood as an evaluation for a realization of a random variable* $\mathbf{Y}$*, while* $f_p(\mathbf{X}|\mathbf{Y})$ *should be understood as a function of a random variable* $\mathbf{Y}$ *which itself is a random variable.*

We shall denote

$$\text{mmpe}(\mathbf{X}|\mathbf{Y}; p) = \text{mmpe}(\mathbf{X}, \mathsf{snr}, p), \quad (11)$$

if $\mathbf{Y}$ and $\mathbf{X}$ are related as

$$\mathbf{Y} = \sqrt{\mathsf{snr}}\, \mathbf{X} + \mathbf{Z}, \quad (12)$$

where $\mathbf{Z}, \mathbf{X}, \mathbf{Y} \in \mathbb{R}^n$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is independent of $\mathbf{X}$, and $\mathsf{snr} \ge 0$ is the SNR. When it will be necessary to emphasize the SNR at the output $\mathbf{Y}$, we will denote it with $\mathbf{Y}_{\mathsf{snr}}$. Since the distribution of the noise is fixed $\text{mmpe}(\mathbf{X}|\mathbf{Y}; p)$ is completely determined by the distribution of $\mathbf{X}$ and snr and there is no ambiguity in using the notation $\text{mmpe}(\mathbf{X}, \mathsf{snr}, p)$. Applications to the Gaussian noise channel will be the main focus of this paper.

[3] The restriction to measurable functions, in Definition 2, is necessary. See [39] for surprising complications that can arise without this assumption.

Note that there are other ways of defining the loss function in (9); our definition in (9) is motivated by:

- For $X \in \mathbb{R}^1$ the error in (9) reduces to a natural expression with loss function given by $\|X - f(Y)\|^p = |X - f(Y)|^p$;
- The definition in (9) naturally appears in applications of Hölder's or Jensen's inequalities to $\mathrm{mmse}(X|Y)$; and
- The norm in (5) used in the definition of (9) can be related to information theoretic quantities, such as differential entropy and Rényi entropy, via the vector moment entropy inequality from [40].

We shall also look at the p-th error achieved by the suboptimal (unless $p = 2$) estimator $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$, that is,

$$\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^p, \tag{13}$$

which represents higher order moments of the MMSE loss function and serves (see below) as an upper bound on (9).

III. Properties of the Optimal MMPE Estimator

A. Existence of Optimal Estimator

It is important to point out that $\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p$, in general is not equal to the MMPE, as $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ might not be the optimal estimator under the p-th norm. The first result of this section shows that for the AWGN channel the optimal estimator $f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$ indeed exists.

Proposition 1: For $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)$, $p > 0$, $\mathsf{snr} > 0$ *the optimal estimator is given by the following point-wise relationship:*

$$f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y}) = \max \Big\{\mathbf{v} : \mathbb{E}\left[\|\mathbf{X} - \mathbf{v}\|^p | \mathbf{Y} = \mathbf{y}\right] = \min_{\mathbf{a}\in\mathbb{R}^n} \mathbb{E}\left[\|\mathbf{X} - \mathbf{a}\|^p | \mathbf{Y} = \mathbf{y}\right]\Big\}. \tag{14}$$

Moreover, for $p > 1$ *the optimal estimator is unique and is given by*

$$f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y}) = \arg\min_{\mathbf{v}\in\mathbb{R}^n} \mathbb{E}\left[\|\mathbf{X} - \mathbf{v}\|^p | \mathbf{Y} = \mathbf{y}\right]. \tag{15}$$

Finally, if $\|\mathbf{X}\|_p < \infty$ *then* (14) *and* (15) *are also valid for* $\mathsf{snr} = 0^+$.

Proof: See Appendix A. □

A result similar to that in Proposition 1 can be found in [35, Th. 4.1.1] where it has been shown that for a given $\mathbf{X}$ an estimator $f_p(\mathbf{X}|\mathbf{Y})$ is optimal provided that the minimum on the right hand side of (14) exists. In contrast to [35, Th. 4.1.1], Proposition 1 shows that the minimum in (14) exists for any $\mathbf{X}$, and $f_p(\mathbf{X}|\mathbf{Y})$ is the MMPE optimal estimator for any $\mathbf{X}$.

Proposition 1 immediately implies the following corollary on the interchange of the expectation and infimum which will be used in many of the following proofs.

Corollary 1: For $p > 0$ *and* $\mathsf{snr} > 0$

$$\begin{aligned} \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) &= \inf_f \frac{1}{n}\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^p\right] \\ &= \frac{1}{n}\mathbb{E}\left[\inf_f \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^p | \mathbf{Y}\right]\right]. \end{aligned} \tag{16}$$

Proof: In the proof of Proposition 1 it is shown that

$$\mathbb{E}\left[\inf_f \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^p | \mathbf{Y}\right]\right] = \mathbb{E}\left[\|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})\|^p\right],$$

for $f_p(\mathbf{X}|\mathbf{Y})$ in (14). Therefore, we have the following chain of inequalities

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})\|^p\right] &= \mathbb{E}\left[\inf_f \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^p | \mathbf{Y}\right]\right] \\ &\le \inf_f \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^p\right] \\ &\le \mathbb{E}\left[\|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})\|^p\right]. \end{aligned} \tag{17}$$

This concludes the proof. □

B. Orthogonality-Like Property

The MMPE for $p \neq 2$ differs from MMSE in a number of aspects. The main difference is that the norm defined in (5) is not a Hilbert space norm in general (unless $p = 2$); as a result, there is no notion of inner product or orthogonality, and $f_p(\mathbf{X}|\mathbf{Y})$, unlike $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$, can no longer be thought of as an orthogonal projection. Therefore, the orthogonality principle—an important tool in the analysis of the MMSE—is no longer available when studying the MMPE for $p \neq 2$. However, an orthogonality-like property can indeed be shown for the MMPE.

Proposition 2 (Necessary and Sufficient Condition for the Optimality of $f_p(\mathbf{X}|\mathbf{Y})$*): For any* $\mathbf{X}$, *any* $\mathsf{snr} > 0$, $p \ge 1$, $f_p(\mathbf{X}|\mathbf{Y})$ *is an optimal estimator if and only if*

$$\mathbb{E}\left[\|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})\|^{p-2} \cdot \left(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})\right)^T \cdot g(\mathbf{Y})\right] = 0, \tag{18a}$$

for any deterministic function $g : \mathbb{R}^n \to \mathbb{R}^n$, *that is,*

$$\mathbb{E}\left[\left(\mathbf{W}^T\mathbf{W}\right)^{\frac{p-2}{2}} \cdot \mathbf{W}^T \cdot g(\mathbf{Y})\right] = 0, \tag{18b}$$

where $\mathbf{W} = \mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})$. *Moreover, for* $0 < p < 1$ *the condition in* (18a) *is necessary for optimality.*

Proof: See Appendix B. □

Note that Proposition 2 for $n = 1$ and $p \in \mathbb{R}^+$ reduces to

$$\mathbb{E}[|X - f_p(X|Y)|^{p-2}(X - f_p(X|Y))g(Y)] = 0, \tag{19}$$

which for $\frac{p}{2} \in \mathbb{N}$ further reduces to

$$\mathbb{E}[(X - f_p(X|Y))^{p-1}g(Y)] = 0. \tag{20}$$

Moreover, for $p = 2$ Proposition 2 reduces to the familiar orthogonality principle

$$\begin{aligned} \mathbb{E}\left[(\mathbf{X} - f_2(\mathbf{X}|\mathbf{Y}))^T \cdot g(\mathbf{Y})\right] &= \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^T \cdot g(\mathbf{Y})\right] \\ &= 0. \end{aligned} \tag{21}$$

Remark 2: In the analysis of the MMSE the orthogonality property is an important tool, used for example to show that $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ *is the unique minimizer. The argument goes as follows: assume that we have another optimal estimator* $f(\mathbf{Y}) \neq \mathbb{E}[\mathbf{X}|\mathbf{Y}]$, *then by orthogonality principle,*

$$\begin{aligned} 0 &= \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^T g(\mathbf{Y})] - \mathbb{E}[(\mathbf{X} - f(\mathbf{Y}))^T g(\mathbf{Y})] \\ &= \mathbb{E}[(f(\mathbf{Y}) - E[\mathbf{X}|\mathbf{Y}])^T g(\mathbf{Y})]. \end{aligned} \tag{22}$$

By choosing $g(\mathbf{Y}) = (f(\mathbf{Y}) - E[\mathbf{X}|\mathbf{Y}])$ *we see that* $\mathbb{E}[(f(\mathbf{Y}) - E[\mathbf{X}|\mathbf{Y}])^T(f(\mathbf{Y}) - E[\mathbf{X}|\mathbf{Y}])] > 0$*, arriving at a contradiction. This implies that* $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ *is the unique estimator up to a set of measure zero.*

In [26, Lemma 1], by replicating the above argument and by assuming that $\frac{p}{2} \in \mathbb{N}$ *and* $n = 1$*, it was shown that the optimal MMPE estimator is* unique. *However, since the proof relies heavily on the assumption that* $\frac{p}{2} \in \mathbb{N}$ *and* $n = 1$*, this argument cannot be extended in a straightforward way to* $p \in \mathbb{R}^+$ *or* $n > 1$.

C. Examples of Optimal MMPE Estimators

In general we do not have a closed form solution for the MMPE optimal estimator in (14). Interestingly, the optimal estimator for Gaussian inputs can be found and is linear for all $p \geq 1$. Note that similar results have been demonstrated in [20] and [26] for scalar Gaussian inputs. Next we extend this result to vector inputs and give two alternative proofs of the linearity of the optimal MMPE estimator for Gaussian inputs, via Proposition 1 and via Proposition 2.

Proposition 3: For input $\mathbf{X}_G \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ *and* $p \geq 1$

$$\mathrm{mmpe}(\mathbf{X}_G, \mathsf{snr}, p) = \frac{\|\mathbf{Z}\|_p^p}{(1+\mathsf{snr})^{\frac{p}{2}}}, \tag{23a}$$

with optimal estimator given by

$$f_p(\mathbf{X}_G|\mathbf{Y} = \mathbf{y}) = \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\,\mathbf{y}. \tag{23b}$$

Proof: The proof follows by observing that $\mathbf{W} = \mathbf{X}_G - \frac{\sqrt{\mathsf{snr}}\mathbf{Y}}{1+\mathsf{snr}}$ has a Gaussian distribution and is independent of $\mathbf{Y}$. So, for any two functions $f(\cdot)$ and $g(\cdot)$ we have

$$\mathbb{E}[f(\mathbf{W})g(\mathbf{Y})] = 0. \tag{24}$$

Therefore, by using (24) for estimator $f_p(\mathbf{X}_G|\mathbf{Y} = \mathbf{y}) = \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\,\mathbf{y}$ the necessary and sufficient conditions in Proposition 2 hold and thus the linear estimator must be an optimal one. Finally observe that

$$\left\|\mathbf{X}_G - \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\mathbf{Y}\right\|_p^p = \|\hat{\mathbf{Z}}\|_p^p = \frac{\|\mathbf{Z}\|_p^p}{(1+\mathsf{snr})^{\frac{p}{2}}}, \tag{25}$$

where we have used $\hat{\mathbf{Z}} = \mathbf{X}_G - \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\mathbf{Y} \sim \mathcal{N}\left(0, \frac{1}{1+\mathsf{snr}}\mathbf{I}\right)$.

For a proof that uses only Proposition 1 see Appendix C. $\square$

The optimal MMPE estimator is in general a function of p as shown next.

Proposition 4: For $X \in \{x_1, x_2\}$ *with* $\mathbb{P}[X = x_1] = 1 - \mathbb{P}[X = x_2] = q \in (0, 1)$ *and for* $p \geq 1$ *we have that*

$$f_p(X|Y=y) = \frac{x_1 \cdot q^{\frac{1}{p-1}} \cdot \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2(p-1)}} + x_2 \cdot (1-q)^{\frac{1}{p-1}} \cdot \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2(p-1)}}}{q^{\frac{1}{p-1}} \cdot \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2(p-1)}} + (1-q)^{\frac{1}{p-1}} \cdot \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2(p-1)}}}. \tag{26a}$$

In particular, for $p = 1$*, we have that*

$$f_{p=1}(X|Y=y) = \begin{cases} x_1, & a \geq 1 \\ x_2, & \text{otherwise}, \end{cases} \tag{26b}$$

where

$$a = \frac{q}{q-1}\mathrm{e}^{-\mathsf{snr}\frac{(x_1-x_2)(x_1+x_2)}{2} + \sqrt{\mathsf{snr}}y(x_1-x_2)}. \tag{26c}$$

Proof: See Appendix D. $\square$

Proposition 4 will be useful in demonstrating several examples and counter examples in the following sections. Note that for the practically relevant case of BPSK modulation, or $x_1 = -x_2 = 1$ and $q = \frac{1}{2}$, the optimal estimator in (26a) reduces to

$$f_p(X|Y=y) = \tanh\left(\frac{y\sqrt{\mathsf{snr}}}{p-1}\right), \tag{27a}$$

which for $p = 1$ is the hard decision decoder

$$f_{p=1}(X|Y=y) = \begin{cases} -1, & y \leq 0 \\ +1, & y > 0. \end{cases} \tag{27b}$$

By Proposition 4 we can show that the orthogonality principle only holds for $p = 2$ (when MMPE corresponds to MMSE) as shown in Fig. 1a, where we plot $h(p) := \mathbb{E}[(X - f_p(X|Y))Y]$ vs. p for BPSK input and observe it is zero only for $p = 2$.

D. Basic Properties of the Optimal MMPE Estimator

Interestingly many of the known properties of $f_2(\mathbf{X}|\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$ for MMSE are still exhibited by $f_p(\mathbf{X}|\mathbf{Y})$ for any $p > 0$.

Proposition 5: For any $p > 0$ *the optimal MMPE estimator has the following properties:*

1) *if* $0 \leq X \in \mathbb{R}^1$ *then* $0 \leq f_p(X|Y)$*,*
2) *(Linearity)* $f_p(a\mathbf{X} + b|\mathbf{Y}) = af_p(\mathbf{X}|\mathbf{Y}) + b$ *for* $a, b \in \mathbb{R}$*,*
3) *(Stability)* $f_p(g(\mathbf{Y})|\mathbf{Y}) = g(\mathbf{Y})$ *for any deterministic function* $g(\cdot)$*,*
4) *(Idempotent)* $f_p(f_p(\mathbf{X}|\mathbf{Y})|\mathbf{Y}) = f_p(\mathbf{X}|\mathbf{Y})$*,*
5) *(Degradedness)* $f_p\left(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}, \mathbf{Y}_{\mathsf{snr}}\right) = f_p\left(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}\right)$ *for a Markov chain* $\mathbf{X} \to \mathbf{Y}_{\mathsf{snr}_0} \to \mathbf{Y}_{\mathsf{snr}}$*,*
6) *(Orthogonality-like Principle) See Proposition 2.*

Proof: See Appendix E. $\square$

It is important to point out that in general, the linearity property does not hold for the sum of random variables. That is, the following property:

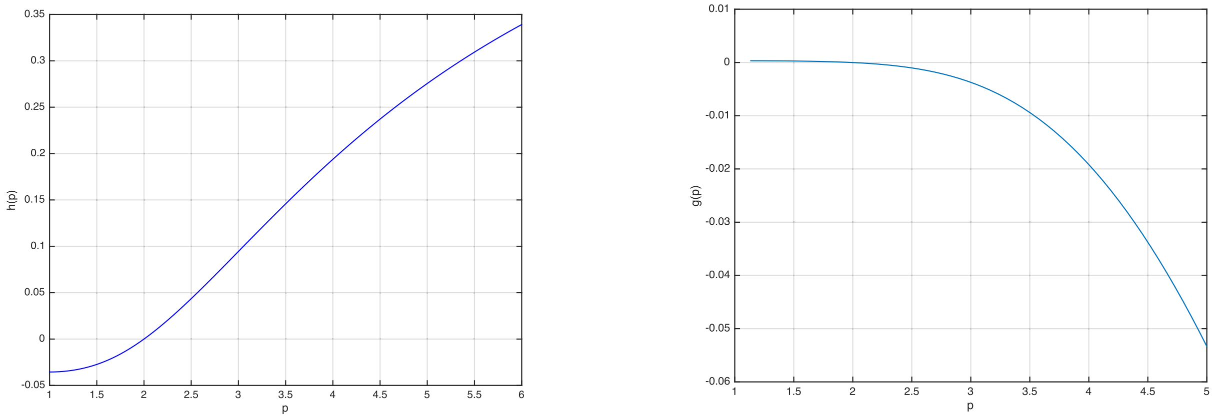
$$f_p(a\mathbf{X}_1 + b\mathbf{X}_2|\mathbf{Y}) = af_p(\mathbf{X}_1|\mathbf{Y}) + bf_p(\mathbf{X}_2|\mathbf{Y}), \tag{28}$$

in general is not true.

Remark 3 (Average Bias of the MMPE Optimal Estimator): An estimator $f_p(\mathbf{X}|\mathbf{Y})$ *is said to be unbiased on average if* $\mathbb{E}[\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})] = \mathbf{0}$*. In general* $f_p(\mathbf{X}|\mathbf{Y})$ *is unbiased on average only for* $p = 2$*, since*

$$\mathbb{E}[f_{p=2}(\mathbf{X}|\mathbf{Y})] = \mathbb{E}[\mathbb{E}[\mathbf{X}|\mathbf{Y}]] = \mathbb{E}[\mathbf{X}]. \tag{29}$$

Fig. 1b shows that in general the optimal MMPE estimator is biased on average; it plots $\mathbb{E}[X - f_p(X|Y)]$ *vs.* p *for* $X \in \{-3, 1\} : \mathbb{P}[X = -3] = 0.01$ *and* $\mathsf{snr} = 1$*, with* $f_p(X|Y)$ *as in Proposition 4. This comes as no surprise as it is very common in Bayesian estimation that the optimal estimator is biased [41].*



(a) (b)

Fig. 1. Counter examples for the orthogonality principle and the bias of the MMPE optimal estimator. (a) Plot of $h(p) := \mathbb{E}[(X - f_p(X|Y))Y]$ vs. p, for $X \in \{\pm 1\}, \mathbb{P}[X = 1] = \frac{1}{2}$ and $\mathsf{snr} = 1$ and $f_p(X|Y)$ given in (27a). (b) Plot of $g(p) := \mathbb{E}[(X - f_p(X|Y)]$ vs. p, for $X \in \{-3, 1\}$, $\mathbb{P}[X = -3] = 0.01$ and $\mathsf{snr} = 1$ with $f_p(X|Y)$ given in (26a).

IV. Properties of the MMPE

In this section we explore properties of the MMPE as a function of SNR and of the input distribution.

A. Basic Properties

The next two properties of the MMPE directly follow from the properties of $f_p(\mathbf{X}|\mathbf{Y})$ in Proposition 5.

Proposition 6: For any $p > 0$

$$\mathrm{mmpe}(\mathbf{X} + a, \mathsf{snr}, p) = \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p), \tag{30a}$$

$$\mathrm{mmpe}(a\mathbf{X}, \mathsf{snr}, p) = a^p \mathrm{mmpe}(\mathbf{X}, a^2\mathsf{snr}, p). \tag{30b}$$

Proposition 6 implies that the MMPE, like the MMSE, is invariant under translations, and that scaling the input results in scaling the SNR and the error.

B. Estimation of the Input is Equivalent to Estimation of the Noise

The following lemma is commonly applied in the analysis of the MMSE.

Lemma 1: For $\mathbf{X}, \mathbf{Z}, \mathbf{Y}$ *given in* (12)

$$\sqrt{\mathsf{snr}}(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]) = -(\mathbf{Z} - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]). \tag{31a}$$

Moreover,

$$\sqrt{\mathsf{snr}}\ \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p = \|\mathbf{Z} - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]\|_p. \tag{31b}$$

Lemma 1 states that estimating the noise is equivalent to estimating the input signal if one uses the conditional expectation as an estimator.

Next we show that an equivalent statement holds for the MMPE.

Proposition 7: For $\mathbf{X}, \mathbf{Z}, \mathbf{Y}$ *given in* (12)

$$\sqrt{\mathsf{snr}}\ \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}|\mathbf{Y}; p) = \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{Z}|\mathbf{Y}; p). \tag{32a}$$

Moreover,

$$\sqrt{\mathsf{snr}}\ (\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})) = -(\mathbf{Z} - f_p(\mathbf{Z}|\mathbf{Y})). \tag{32b}$$

Proof: From the definition of the MMPE in (9)

$$\begin{aligned}
&\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p)\\
&= \inf_{f(\mathbf{y})} \|\mathbf{X} - f(\mathbf{Y})\|_p \qquad (33)\\
&= \inf_{f(\mathbf{y})} \left\| \frac{1}{\sqrt{\mathsf{snr}}}(\mathbf{Y} - \mathbf{Z}) - f(\mathbf{Y}) \right\|_p\\
&= \frac{1}{\sqrt{\mathsf{snr}}} \inf_{f(\mathbf{y})} \left\| \mathbf{Z} - \left(\sqrt{\mathsf{snr}} f(\mathbf{Y}) - \mathbf{Y}\right) \right\|_p\\
&= \frac{1}{\sqrt{\mathsf{snr}}} \inf_{g(\mathbf{y}):\ g(\mathbf{y}) = \mathbf{y} - \sqrt{\mathsf{snr}} f(\mathbf{y})} \|\mathbf{Z} - g(\mathbf{Y})\|_p \qquad (34)\\
&= \frac{1}{\sqrt{\mathsf{snr}}} \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{Z}|\mathbf{Y}; p).
\end{aligned}$$

This shows the equality in (32a). Moreover, since $f_p(\mathbf{X}|\mathbf{Y})$ exists and the infimum in (33) is attainable by Proposition 1, so is the infimum in (34). Therefore, from (34) we have that $f_p(\mathbf{Z}|\mathbf{Y})$ exists and is given by

$$\begin{aligned}
f_p(\mathbf{Z}|\mathbf{Y}) &= \mathbf{Y} - \sqrt{\mathsf{snr}} f_p(\mathbf{X}|\mathbf{Y})\\
&= \sqrt{\mathsf{snr}}\mathbf{X} + \mathbf{Z} - \sqrt{\mathsf{snr}} f_p(\mathbf{X}|\mathbf{Y}), \qquad (35)
\end{aligned}$$

which leads to (32b). This concludes the proof. □

C. Change of Measure

The next result enables us to change the expectation from $\mathbf{Y}_{\mathsf{snr}}$ to $\mathbf{Y}_{\mathsf{snr}_0}$ in the definition of the MMPE in (9) whenever $\mathsf{snr} \le \mathsf{snr}_0$. This is particularly useful when we know the MMPE, or the structure of the optimal MMPE estimator, at one SNR value but not at another smaller SNR value.

Proposition 8: For any $\mathbf{X}$, $\mathsf{snr} \in (0, \mathsf{snr}_0]$ *and* $p > 0$, *we have*

$$\begin{aligned}
&\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)\\
&= \inf_f \frac{1}{n} \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^p \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathrm{e}^{\frac{\mathsf{snr}_0 - \mathsf{snr}}{2\mathsf{snr}_0} \sum_{i=1}^n Z_i^2} \right]. \qquad (36)
\end{aligned}$$

Proof: See Appendix F. □

One must be careful when evaluating Proposition 8. For example, since we have that

$$\lim_{\mathsf{snr}\to 0^+} \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2} = 0,$$

at first glance it appears that the expectation on the right of (36) is zero while $\mathrm{mmpe}(X, 0, p)$ is not, thus violating the equality. However, a more careful examination shows that when $\mathsf{snr} \to 0$ the limit and expectation in (36) cannot be exchanged; indeed we have that

$$\begin{aligned}
&\lim_{\mathsf{snr}\to 0^+} \mathbb{E}\left[\sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2}\right] \\
&= \lim_{\mathsf{snr}\to 0^+} \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathbb{E}\left[\mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2}\right] \\
&= \lim_{\mathsf{snr}\to 0^+} \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \frac{1}{\sqrt{1-\frac{\mathsf{snr}_0-\mathsf{snr}}{\mathsf{snr}_0}}} = 1,
\end{aligned}$$

where in the last equality we used the moment generating function of the Cauchy r.v. Z^2. As an example, Proposition 8 for $X \sim \mathcal{N}(0, 1)$ with the optimal linear estimator from Proposition 3, i.e., $f(y) = ay$ for some a, evaluates to

$$\begin{aligned}
&\mathbb{E}\left[\|X - f(Y_{\mathsf{snr}_0})\|^2 \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2}\right] \\
&\overset{a)}{=} (1-\sqrt{\mathsf{snr}_0}a)^2 \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathbb{E}[X^2] \mathbb{E}\left[\mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2}\right] \\
&\quad + a^2 \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathbb{E}\left[Z^2 \mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2}\right] \\
&\overset{b)}{=} \frac{1}{1+\mathsf{snr}},
\end{aligned}$$

where the equalities follow from: a) linearity of expectation and the fact that Z and X are independent; and b) since $\mathbb{E}\left[\mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2}\right] = \sqrt{\frac{\mathsf{snr}_0}{\mathsf{snr}}}$ and $\mathbb{E}\left[Z^2\mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}Z^2}\right] = \left(\frac{\mathsf{snr}_0}{\mathsf{snr}}\right)^{3/2}$ and by choosing $a = \frac{\mathsf{snr}}{\sqrt{\mathsf{snr}_0}(1+\mathsf{snr})}$ in order to minimize the expression in a).[4]

V. Bounds on the MMPE

In this section we develop bounds on the MMPE, many of which generalize well known MMSE bounds. However, we also show bounds that are unique to the MMPE and emphasize the usefulness of the MMPE.

A. Extension of Basic MMSE Bounds

An important upper bound on the MMSE often used in practice is the LMMSE.

Proposition 9 (LMMSE [9]): For any input $\mathbf{X}$ *and* $\mathsf{snr} > 0$

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \frac{1}{\mathsf{snr}}. \tag{37a}$$

If $\|\mathbf{X}\|_2^2 = \sigma^2 < \infty$*, then for any* $\mathsf{snr} \ge 0$

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \frac{\sigma^2}{1+\sigma^2\mathsf{snr}}, \tag{37b}$$

where equality in (37b) *is achieved iff* $\mathbf{X} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.

The next bound generalizes Proposition 9 to higher order errors.

Proposition 10: For $\mathsf{snr} \ge 0$, $0 < q \le p$*, and input* $\mathbf{X}$

$$\begin{aligned}
n^{\frac{p}{q}-1}\mathrm{mmpe}^{\frac{p}{q}}(\mathbf{X}, \mathsf{snr}, q) &\le \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \\
&\le \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^p, \quad (38a)
\end{aligned}$$

and where

$$\text{for } p \ge 2: \ \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^p \le 2^p \min\left(\frac{\|\mathbf{Z}\|_p^p}{\mathsf{snr}^{\frac{p}{2}}}, \|\mathbf{X}\|_p^p\right) \tag{38b}$$

$$\begin{aligned}
&\text{for } 1 \le p \le 2: \ \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^p \\
&\le \min\left(\frac{\left(\|\mathbf{Z}\|_p + n^{\frac{1}{2}-\frac{1}{p}}\|\mathbf{Z}\|_2\right)^p}{\mathsf{snr}^{\frac{p}{2}}}, \left(\|\mathbf{X}\|_p + n^{\frac{1}{2}-\frac{1}{p}}\|\mathbf{X}\|_2\right)^p\right) \quad (38c)
\end{aligned}$$

$$\text{for } p \ge 0: \ \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \le \min\left(\frac{\|\mathbf{Z}\|_p^p}{\mathsf{snr}^{\frac{p}{2}}}, \|\mathbf{X}\|_p^p\right), \tag{38d}$$

where $\|\mathbf{Z}\|_p^p$ *is given in* (7).

Proof: See Appendix G. □

It is interesting to point out that in the derivation of the bounds in Proposition 10 no assumption is put on the distribution of $\mathbf{Z}$, and thus the bounds hold in great generality. If $\mathbf{Z}$ is composed of independent identically distributed (i.i.d.) Gaussian elements, then the moment $\|\mathbf{Z}\|_p^p$ in Proposition 10 can be tightly approximated in terms of factorials as

$$\begin{aligned}
\frac{n}{2^{\frac{p}{2}}}\|\mathbf{Z}\|_p^p &= \frac{\Gamma\left(\frac{n}{2}+\frac{p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \\
&\le \frac{\Gamma\left(\left\lceil\frac{n}{2}+\frac{p}{2}\right\rceil\right)}{\Gamma\left(\left\lfloor\frac{n}{2}\right\rfloor\right)} = \frac{\left(\left\lceil\frac{n}{2}+\frac{p}{2}\right\rceil-1\right)!}{\left(\left\lfloor\frac{n}{2}\right\rfloor-1\right)!} = O(n^{\frac{p}{2}}),
\end{aligned} \tag{39}$$

which is tight for even n and integer $\frac{p}{2}$.

It is not difficult to check that for $p = 2$ Proposition 10 reduces to Proposition 9. The reason that the bounds on $\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p$ are only available for $p \ge 2$, while the bounds on $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)$ are available for $p \ge 0$, is because the proof of the bound in (38b) uses Jensen's inequality, which requires $p \ge 2$, while the proof of the bound in (38d) does not.

B. Gaussian Inputs Are the Hardest to Estimate

Note that the bounds in Proposition 10 are similar to the bound in (37a) and blow up at $\mathsf{snr} = 0^+$. Therefore, it is desirable to have bounds as in (37b). The next result

[4]Note that this optimal a is evident from the specific change of measure that we have used. Instead of having the estimator according to Proposition 3 as $\frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}$ we get it with the normalization by $\sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}}$.

demonstrates such a bound and shows that Gaussian inputs are asymptotically the hardest to estimate.

Proposition 11: For $\mathsf{snr} \geq 0$, $p \geq 1$, *and a random variable* $\mathbf{X}$ *such that* $\|\mathbf{X}\|_p^p \leq \sigma^p \|\mathbf{Z}\|_p^p$, *we have*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \leq \kappa_{p,\sigma^2\mathsf{snr}} \cdot \frac{\sigma^p \|\mathbf{Z}\|_p^p}{(1+\mathsf{snr}\sigma^2)^{\frac{p}{2}}}, \tag{40a}$$

where

$$for\ \ p=2:\ \kappa_{p,\sigma^2\mathsf{snr}}^{\frac{1}{p}} = 1, \tag{40b}$$

$$for\ \ p \neq 2:\ 1 \leq \kappa_{p,\sigma^2\mathsf{snr}}^{\frac{1}{p}} = \frac{1+\sqrt{\sigma^2\mathsf{snr}}}{\sqrt{1+\sigma^2\mathsf{snr}}} \leq 1 + \frac{1}{\sqrt{1+\sigma^2\mathsf{snr}}}. \tag{40c}$$

Moreover, a Gaussian $\mathbf{X}$ *with per-dimension variance* σ^2 *(i.e.,* $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$*) asymptotically achieves the bound in* (40a)*, since* $\lim_{\mathsf{snr}\to\infty} \kappa_{p,\sigma^2\mathsf{snr}} = 1$.

Proof: See Appendix H. □

C. Interpolation Bounds and Continuity

One of the key advantages of using the MMPE is that the MMPE of order q can be tightly predicted based on the knowledge of the MMPE at lower orders p and higher orders r. At the heart of this analysis is the interpolation result of L_p spaces [42]: given $0 < p \leq q \leq r$ and $\alpha \in (0,1)$ such that $\frac{1}{q} = \frac{\alpha}{p} + \frac{1-\alpha}{r}$, the q-th norm can be bounded as

$$\|V\|_q \leq \|V\|_p^{\alpha} \|V\|_r^{(1-\alpha)}, \tag{41}$$

which implies that the norm is log-convex and thus a continuous function of p [43, Th. 5.1.1]. Next, we present several interpolation results for the MMPE.

Proposition 12 (Log-Convexity and Interpolation): For any $0 < p \leq q \leq r \leq \infty$ *and* $\alpha \in (0,1)$ *such that*

$$\frac{1}{q} = \frac{\alpha}{p} + \frac{\bar{\alpha}}{r} \iff \alpha = \frac{q^{-1}-r^{-1}}{p^{-1}-r^{-1}}, \tag{42a}$$

where $\bar{\alpha} = 1-\alpha$*, we have for any measurable* $f(\mathbf{Y})$

$$\|\mathbf{X}-f(\mathbf{Y})\|_q \leq \|\mathbf{X}-f(\mathbf{Y})\|_p^{\alpha}\ \|\mathbf{X}-f(\mathbf{Y})\|_r^{\bar{\alpha}}. \tag{42b}$$

In particular,

$$\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_q \leq \|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^{\alpha}\ \|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_r^{\bar{\alpha}}. \tag{42c}$$

Moreover,

$$\mathrm{mmpe}^{\frac{1}{q}}(\mathbf{X}, \mathsf{snr}, q) \leq \inf_f \|\mathbf{X}-f(\mathbf{Y})\|_p^{\alpha}\|\mathbf{X}-f(\mathbf{Y})\|_r^{\bar{\alpha}}. \tag{42d}$$

In particular,

$$\mathrm{mmpe}^{\frac{1}{q}}(\mathbf{X}, \mathsf{snr}, q) \leq \|\mathbf{X}-f_r(\mathbf{X}|\mathbf{Y})\|_p^{\alpha}\ \mathrm{mmpe}^{\frac{\bar{\alpha}}{r}}(\mathbf{X}, \mathsf{snr}, r), \tag{42e}$$

$$\mathrm{mmpe}^{\frac{1}{q}}(\mathbf{X}, \mathsf{snr}, q) \leq \mathrm{mmpe}^{\frac{\alpha}{p}}(\mathbf{X}, \mathsf{snr}, p)\|\mathbf{X}-f_p(\mathbf{X}|\mathbf{Y})\|_r^{\bar{\alpha}}. \tag{42f}$$

Proof: The bound in (42b) follows by applying (41) with $V = \|\mathbf{X}-f(\mathbf{Y})\| \in \mathbb{R}$. The bound in (42c) follow by choosing $f(\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$.

The bound in (42d) follows by

$$\begin{aligned}\mathrm{mmpe}^{\frac{1}{q}}(\mathbf{X}, \mathsf{snr}, q) &= \inf_f \|\mathbf{X}-f(\mathbf{Y})\|_q \\ &\leq \inf_f \|\mathbf{X}-f(\mathbf{Y})\|_p^{\alpha}\|\mathbf{X}-f(\mathbf{Y})\|_r^{\bar{\alpha}},\end{aligned} \tag{43}$$

where the last inequality follows from (41) by choosing $V = \|\mathbf{X}-f(\mathbf{Y})\| \in \mathbb{R}$.

Finally, the bounds in (42e) and (42f) follow by choosing $f(\mathbf{Y})$ in (42d) equal to $f_r(\mathbf{X}|\mathbf{Y})$ and $f_p(\mathbf{X}|\mathbf{Y})$ respectively. This concludes the proof. □

From log-convexity we can deduce continuity.

Proposition 13 (Continuity): For any $\mathbf{X}$ *and* $\mathsf{snr} > 0$, $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)$ *and* $\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p$ *are continuous functions of* $p > 0$.

Proof: Continuity of $\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p$ follows from log-convexity in (42c) while the continuity of MMPE follows from

$$\begin{aligned}&\lim_{q\to p} |\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) - \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, q)| \\ &\quad\leq \lim_{q\to p} \max\Big(\|\mathbf{X}-f_q(\mathbf{X}|\mathbf{Y})\|_p^p - \|\mathbf{X}-f_q(\mathbf{X}|\mathbf{Y})\|_q^q, \\ &\qquad \|\mathbf{X}-f_p(\mathbf{X}|\mathbf{Y})\|_q^q - \|\mathbf{X}-f_p(\mathbf{X}|\mathbf{Y})\|_p^p\Big) = 0,\end{aligned}$$

where the last inequality is due to the continuity of the norm. □

An interesting question is whether the following interpolation inequality holds:

$$\mathrm{mmpe}^{\frac{1}{q}}(\mathbf{X}, \mathsf{snr}, q) \leq \mathrm{mmpe}^{\frac{\alpha}{p}}(\mathbf{X}, \mathsf{snr}, p)\ \mathrm{mmpe}^{\frac{\bar{\alpha}}{r}}(\mathbf{X}, \mathsf{snr}, r) \tag{44}$$

instead of (42e) and (42f). A counter example to the interpolation inequality in (44) is shown in Fig. 2 where we take a binary input $X \in \{\pm 1\}$ equality likely, $p = 2, r = 8$, and $\mathsf{snr} = 1$ and show:

- The MMPE of X of order q versus $\alpha \in [0,1]$ where q is computed according to (42a) (blue-solid line);
- The interpolation bound in (42e) (purple dashed-dotted line);
- The interpolation bound in (42f) (yellow solid-dotted line);
- The interpolation bound in (42d) with $f(Y) = f_{\frac{p+r}{2}}(X|Y)$ (green dashed line); and
- The right-hand side of the conjectured inequality in (44) (red-dotted line).

This shows that (44) is not true in general.

D. Bounds on Discrete Inputs

Next, we investigate properties of the MMPE under the assumption that the input is a discrete r.v. Discrete inputs are commonly encountered in practice and, therefore, it is worthwhile to investigate their performance.

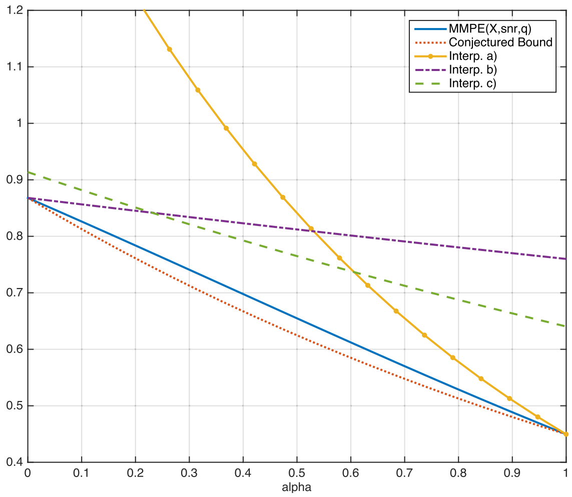


Fig. 2. Interpolation bounds from Proposition 12 and the conjectured bound in (44) versus α. Clearly the conjectured bound is below the true MMPE thus (44) cannot be true.

Proposition 14: Let $\mathbf{X}_D$ *be a discrete r.v. with* $|\mathsf{supp}(\mathbf{X}_D)| = N$ *and* $\mathbb{P}[\mathbf{X}_D = \mathbf{x}_i] = p_i$ *for* $\mathbf{x}_i \in \mathsf{supp}(\mathbf{X}_D)$*. Then for any* $\hat{\mathbf{X}}_D : \mathbb{R}^n \to \mathsf{supp}(\mathbf{X}_D)$

$$\mathrm{mmpe}(\mathbf{X}_D, \mathsf{snr}, p) \le \frac{d_{\max}^p(\mathbf{X}_D)}{n} P_e^{(n)}(\mathsf{snr}), \tag{45a}$$

where

$$P_e^{(n)}(\mathsf{snr}) = \mathbb{P}[\mathbf{X}_D \neq \hat{\mathbf{X}}_D(\mathbf{Y})], \tag{45b}$$

$$d_{\max}(\mathbf{X}_D) = \max_{\mathbf{x}_j, \mathbf{x}_i \in \mathsf{supp}(\mathbf{X}_D)} \|\mathbf{x}_j - \mathbf{x}_i\|. \tag{45c}$$

Proof: The proof follows by upper bounding the MMPE with the probability of detection error. Consider

$$\begin{aligned} &n\, \mathrm{mmpe}(\mathbf{X}_D, \mathsf{snr}, p) \\ &\overset{a)}{\le} \mathbb{E}\left[\|\mathbf{X}_D - \hat{\mathbf{X}}_D(\mathbf{Y})\|^p\right] \\ &= \mathbb{E}\left[\|\mathbf{X}_D - \hat{\mathbf{X}}_D(\mathbf{Y})\|^p \mid \mathbf{X}_D = \hat{\mathbf{X}}_D(\mathbf{Y})\right] \mathbb{P}[\mathbf{X}_D = \hat{\mathbf{X}}_D(\mathbf{Y})] \\ &\quad + \mathbb{E}\left[\|\mathbf{X}_D - \hat{\mathbf{X}}_D(\mathbf{Y})\|^p \mid \mathbf{X}_D \neq \hat{\mathbf{X}}_D(\mathbf{Y})\right] \mathbb{P}[\mathbf{X}_D \neq \hat{\mathbf{X}}_D(\mathbf{Y})] \\ &\overset{b)}{=} \mathbb{E}\left[\|\mathbf{X}_D - \hat{\mathbf{X}}_D(\mathbf{Y})\|^p \mid \mathbf{X}_D \neq \hat{\mathbf{X}}_D(\mathbf{Y})\right] \mathbb{P}[\mathbf{X}_D \neq \hat{\mathbf{X}}_D(\mathbf{Y})] \\ &\overset{c)}{\le} d_{\max}^p(\mathbf{X}_D) \mathbb{P}[\mathbf{X}_D \neq \hat{\mathbf{X}}_D(\mathbf{Y})], \end{aligned}$$

where the (in)-equalities follow from: a) choosing a suboptimal estimator; b) using the fact that $\mathbb{E}\left[\|\mathbf{X}_D - \hat{\mathbf{X}}_D(\mathbf{Y})\|^p \mid \mathbf{X}_D = \hat{\mathbf{X}}_D(\mathbf{Y})\right] = 0$; and c) using a bound $\|\mathbf{X}_D - \hat{\mathbf{X}}_D(\mathbf{Y})\| \le \max_{\mathbf{x}_j, \mathbf{x}_i \in \mathsf{supp}(\mathbf{X}_D)} \|\mathbf{x}_j - \mathbf{x}_i\| = d_{\max}(\mathbf{X}_D)$. This concludes the proof. ☐

So far, by using Proposition 10, we have shown that the MMPE as a function of snr decreases as $O\left(\frac{2^{\frac{p}{2}} n^{\frac{p}{2}-1}}{\mathsf{snr}^{\frac{p}{2}}}\right)$. Next we show that the MMPE can decrease exponentially in snr by choosing $\hat{X}_D(\mathbf{Y})$ to be the maximum a posteriori (MAP) decoder. Such behavior has been already observed for the MMSE in [17] and [44].

Proposition 15: Let $\mathbf{X}_D$ *be a discrete r.v. with* $|\mathsf{supp}(\mathbf{X}_D)| = N$ *and* $\mathbb{P}[\mathbf{X}_D = \mathbf{x}_i] = p_i$ *for* $\mathbf{x}_i \in \mathsf{supp}(\mathbf{X}_D)$ *then:*

$$\begin{aligned} &\mathrm{mmpe}(\mathbf{X}_D, \mathsf{snr}, p) \\ &\le \frac{d_{\max}^p(\mathbf{X}_D)}{n} \sum_{i=1}^N p_i \sum_{j=1:j\neq i}^N Q\left(\frac{\sqrt{\mathsf{snr}}\, d_{ij}}{2} - \frac{\log\left(\frac{p_j}{p_i}\right)}{\sqrt{\mathsf{snr}}\, d_{ij}}\right), \end{aligned} \tag{46a}$$

where

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|. \tag{46b}$$

Proof: See Appendix I. ☐

A slightly weaker bound than that in Proposition 15, yet computationally simpler, can be derived by choosing $\hat{X}_D(\mathbf{Y})$ to be a threshold (or sphere) decoder. This weaker bound would be used later on the mutual information in Section VIII-B.

Corollary 2: For any discrete r.v. $\mathbf{X}_D$

$$\mathrm{mmpe}(\mathbf{X}_D, \mathsf{snr}, p) \le d_{\max}^p(\mathbf{X}_D) \frac{\bar{Q}\left(\frac{n}{2}; \frac{\mathsf{snr}\ d_{\min}^2(\mathbf{X}_D)}{8}\right)}{n}, \tag{47a}$$

where

$$d_{\min}(\mathbf{X}_D) = \min_{i,j:i\neq j} d_{ij}. \tag{47b}$$

VI. Conditional MMPE

We define the conditional MMPE as follows.

Definition 3: For any $\mathbf{X}$ *and* $\mathbf{U}$*, the conditional MMPE of* $\mathbf{X}$ *given* $\mathbf{U}$ *is defined as*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{U}) := \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}}, \mathbf{U})\|_p^p. \tag{48}$$

The conditional MMPE in (48) reflects the fact that the optimal estimator has been given additional information in the form of $\mathbf{U}$. Note that when $\mathbf{Z}$ is independent of $(\mathbf{X}, \mathbf{U})$ we can write the conditional MMPE for $\mathbf{X}_{\mathbf{u}} \sim P_{\mathbf{X}|\mathbf{U}}(\cdot|\mathbf{u})$ as

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{U}) = \int \mathrm{mmpe}(\mathbf{X}_{\mathbf{u}}, \mathsf{snr}, p)\, dP_{\mathbf{U}}(\mathbf{u}). \tag{49}$$

Since giving extra information does not increase the estimation error, we have the following result.

Proposition 16 (Conditioning Reduces the MMPE): For every $\mathsf{snr} \ge 0$*, and random variable* $\mathbf{X}$*, we have*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \ge \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{U}). \tag{50}$$

Finally, the following Proposition generalizes [11, Proposition 3.4] and states that the MMPE estimation of $\mathbf{X}$ from two observations is equivalent to estimating $\mathbf{X}$ from a single observation with a higher SNR.

Proposition 17: For every $\mathbf{X}$ *and* $p \ge 0$*, let* $\mathbf{U} = \sqrt{\Delta} \cdot \mathbf{X} + \mathbf{Z}_\Delta$ *where* $\mathbf{Z}_\Delta \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ *and where* $(\mathbf{X}, \mathbf{Z}, \mathbf{Z}_\Delta)$ *are mutually independent. Then*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}_0, p|\mathbf{U}) = \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}_0 + \Delta, p). \tag{51}$$

Proof: For two independent observations $\mathbf{Y}_{\mathsf{snr}_0} = \sqrt{\mathsf{snr}_0}\mathbf{X} + \mathbf{Z}$ and $\mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta$ where $\mathbf{Z}_\Delta$ and $\mathbf{Z}$ are independent, by using maximal ratio combining, we have that

$$\begin{aligned}\mathbf{Y}_{\mathsf{snr}} &= \frac{\sqrt{\Delta}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{Y}_\Delta + \frac{\sqrt{\mathsf{snr}_0}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{Y}_{\mathsf{snr}_0}\\ &= \sqrt{\mathsf{snr}_0+\Delta}\mathbf{X} + \mathbf{W},\end{aligned}$$

where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Next by using the same argument as in [11, Proposition 3.4], we have that the conditional probabilities are

$$p_{\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0},\mathbf{Y}_\Delta}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}_0}, \mathbf{y}_\Delta) = p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}}) \tag{52}$$

for $\mathbf{y}_{\mathsf{snr}} = \frac{\sqrt{\Delta}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{y}_\Delta + \frac{\sqrt{\mathsf{snr}_0}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{y}_{\mathsf{snr}_0}$. The equivalence of the posterior probabilities implies that the estimation of $\mathbf{X}$ from $\mathbf{Y}_{\mathsf{snr}}$ is as good as the estimation of $\mathbf{X}$ from $(\mathbf{Y}_{\mathsf{snr}_0}, \mathbf{Y}_\Delta)$. This concludes the proof. □

Propositions 17 and Proposition 16 imply that, for fixed $\mathbf{X}$ and p

$$\begin{aligned}\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) &\geq \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\sqrt{\Delta}\mathbf{X} + \mathbf{Z}')\\ &= \mathrm{mmpe}(\mathbf{X}, \mathsf{snr} + \Delta, p),\end{aligned} \tag{53}$$

and we have the following:

Corollary 3: $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)$ *is a* non-increasing *function of* snr.

VII. SCPP Bound and Its Complement

The SCPP is a powerful tool that can be used to show the advantage of Gaussian inputs over arbitrary inputs in certain channels with Gaussian noise. In conjunction with the I-MMSE relationship, the SCPP provides simple and insightful converse proofs to the capacity of multi-user AWGN channels. The original proof of the SCPP in [9] and [10] relied on bounding the MMSE. Next we give a simpler proof of the SCPP that does not require knowledge of the derivative of the MMSE and can easily be extended to the MMPE of any order p.

First observe that, in light of the bound in (38d), for any $\mathsf{snr} > 0$ we can always find a $\beta \geq 0$ such that

$$\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}, p) = \frac{\beta\|\mathbf{Z}\|_p^2}{1+\beta\mathsf{snr}},$$

since

$$\lim_{\beta\to\infty} \frac{\beta\|\mathbf{Z}\|_p^2}{1+\beta\mathsf{snr}} = \frac{\|\mathbf{Z}\|_p^2}{\mathsf{snr}}.$$

Next we generalize the SCPP bound to the MMPE.

Proposition 18: Let $\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}_0, p) = \frac{\beta\|\mathbf{Z}\|_p^2}{1+\beta\mathsf{snr}_0}$ *for some* $\beta \geq 0$. *Then*

$$\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}, p) \leq c_p \cdot \frac{\beta\|\mathbf{Z}\|_p^2}{1+\beta\ \mathsf{snr}},\ \textit{for}\ \mathsf{snr} \geq \mathsf{snr}_0, \tag{54a}$$

where

$$c_p = \begin{cases} 2 & p \geq 2\\ 1 & p = 2.\end{cases} \tag{54b}$$

Proof: Let $\mathsf{snr} = \mathsf{snr}_0 + \Delta$ for $\Delta \geq 0$, and let $\mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta$. Then

$$\begin{aligned}\mathbf{Y}_{\mathsf{snr}} &= \frac{\sqrt{\Delta}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{Y}_\Delta + \frac{\sqrt{\mathsf{snr}_0}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{Y}_{\mathsf{snr}_0}\\ &= \sqrt{\mathsf{snr}_0+\Delta}\mathbf{X} + \mathbf{W},\end{aligned}$$

where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Next, let

$$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}_0, p) = \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})\|_p^2, \tag{55}$$

and define a suboptimal estimator given $(\mathbf{Y}_\Delta, \mathbf{Y}_{\mathsf{snr}_0})$ as

$$\hat{\mathbf{X}} = \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Y}_\Delta + \gamma f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}), \tag{56}$$

for some $\gamma \in \mathbb{R}$ to be determined later. Then

$$\mathbf{X} - \hat{\mathbf{X}} = \gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Z}_\Delta,$$

and

$$\begin{aligned}&\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p)\\ &\quad= \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}})\|_p\\ &\quad\overset{a)}{=} \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_\Delta, \mathbf{Y}_{\mathsf{snr}_0})\|_p\\ &\quad\overset{b)}{\leq} \|\mathbf{X} - \hat{\mathbf{X}}\|_p = \left\|\gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Z}_\Delta\right\|_p\\ &\quad\overset{c)}{=} \frac{\left\|\|\mathbf{Z}\|_p^2(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \sqrt{\Delta}\cdot m\cdot\mathbf{Z}_\Delta\right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m},\end{aligned} \tag{57}$$

where the (in)-equalities follow from: a) Proposition 17; b) by using the sub-optimal estimator in (56); and c) by choosing $\gamma = \frac{\|\mathbf{Z}\|_p^2}{\|\mathbf{Z}\|_p^2+\Delta\cdot m}$ for m defined in (55).

Next, by applying the triangle inequality to (57) we get

$$\begin{aligned}&\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p)\\ &\quad\leq \frac{\left\|\|\mathbf{Z}\|_p^2(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}))\right\|_p + \left\|\sqrt{\Delta}\cdot m\cdot\mathbf{Z}_\Delta\right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}\\ &\quad= \frac{\sqrt{m}\|\mathbf{Z}\|_p\cdot(\|\mathbf{Z}\|_p + \sqrt{\Delta}\cdot\sqrt{m})}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}\\ &\quad\leq \sqrt{2}\frac{\sqrt{m}\|\mathbf{Z}\|_p}{\sqrt{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}},\end{aligned} \tag{58}$$

where in the last step we used $(a+b) \leq \sqrt{2}\sqrt{a^2+b^2}$.

Note that for the case $p = 2$, instead of using the triangular inequality in (58), the term in (57) can be expanded into a quadratic equation for which it is not hard to see that the choice of $\gamma = \frac{\|\mathbf{Z}\|_p^2}{\|\mathbf{Z}\|_p^2+\Delta\cdot m}$ is optimal and leads to the bound

$$\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p) \leq \frac{\sqrt{m}\|\mathbf{Z}\|_p}{\sqrt{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}}.$$

The proof is concluded by noting that $\beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \mathsf{snr}_0 m}$. □

Remark 4: We conjecture that the multiplicative constant c_p *can be sharpened to* 1 *for all* $p \geq 1$. *However, in order to* *make such a claim one must solve the following optimization problem*

$$\min_{\gamma \in [0,1]} \|(1-\gamma)\mathbf{W} + \gamma \mathbf{Z}\|_p, \tag{59}$$

where $\mathbf{W}$ *and* $\mathbf{Z}$ *are independent and* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$*. It is not clear how to solve* (59) *for* $p \neq 2$ *and thus we leave it for the future work.*

Remark 5: Note that the proof of Proposition 18 does not require the assumption that $\mathbf{Z}$ *is Gaussian and only requires the assumptions of Proposition 17. That is, we only require that a channel is such that the estimation of* $\mathbf{X}$ *from two observations is equivalent to estimating* $\mathbf{X}$ *from a single observation with a higher SNR.*

A. Complementary SCPP Bound

In this section we give a bound that complements the SCPP bound, that is, while the SCPP bounds the MMPE for all $\mathsf{snr} \geq \mathsf{snr}_0$, we give a bound that bounds the MMPE for all $\mathsf{snr} \leq \mathsf{snr}_0$ where it is assumed that the MMPE is known at snr_0.

The next result enables us to bound the MMPE at snr with values of the MMPE at snr_0 while varying the order.

Proposition 19: For $0 < \mathsf{snr} \leq \mathsf{snr}_0$*,* $\mathbf{X}$ *and* $p \geq 0$*, we have*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \leq \kappa_{n,t}\ \mathrm{mmpe}^{\frac{1-t}{1+t}}\left(\mathbf{X}, \mathsf{snr}_0, \frac{1+t}{1-t}\cdot p\right),$$

where

$$\kappa_{n,t} := \left(\frac{2^n}{n^2}\right)^{\frac{t}{t+1}} \left(\frac{1}{1-t}\right)^{\frac{nt}{t+1}-\frac{1}{2}}, \quad t = \frac{\mathsf{snr}_0 - \mathsf{snr}}{\mathsf{snr}_0}.$$

Proof: From Proposition 8 we have that

$$\begin{aligned}
&\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)\\
&= \inf_f \frac{1}{n}\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^p \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}\sum_{i=1}^n Z_i^2}\right]\\
&\overset{a)}{\leq} \inf_f \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}}\frac{1}{n}\left(\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^{m\cdot p}\right]\right)^{\frac{1}{m}}\\
&\quad\cdot\left(\mathbb{E}\left[\mathrm{e}^{\frac{r(\mathsf{snr}_0-\mathsf{snr})}{2\mathsf{snr}_0}\sum_{i=1}^n Z_i^2}\right]\right)^{\frac{1}{r}}\\
&\overset{b)}{=} \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} n^{\frac{1}{m}-1}\mathrm{mmpe}^{\frac{1}{m}}(\mathbf{X}, \mathsf{snr}_0, m\cdot p)\\
&\quad\cdot\left(1 - \frac{r(\mathsf{snr}_0 - \mathsf{snr})}{\mathsf{snr}_0}\right)^{-\frac{n}{2r}},
\end{aligned} \tag{60}$$

where the (in)-equalities follow from: a) Hölder's inequality with conjugate exponents $1 \leq m, r$ such that $\frac{1}{m} + \frac{1}{r} = 1$; and b) by recognizing that the expectation of the exponential is the moment generating function of a Chi-square distribution of degree n, which exists only if $\frac{r(\mathsf{snr}_0-\mathsf{snr})}{2\ \mathsf{snr}_0} < \frac{1}{2}$.

Next, we let $t = \frac{\mathsf{snr}_0-\mathsf{snr}}{\mathsf{snr}_0}$ and let $r = \frac{t+1}{2t}$ in (60), so that $m = \frac{1+t}{1-t}$. Observe that now the bound in (60) holds for all values of $\mathsf{snr} \leq \mathsf{snr}_0$ since

$$\frac{r(\mathsf{snr}_0 - \mathsf{snr})}{\mathsf{snr}_0} = rt = \frac{(t+1)t}{2t} = \frac{t+1}{2} < 1. \tag{61}$$

With the choice of $m = \frac{1+t}{1-t}$ the bound in (60) becomes

$$\begin{aligned}
&\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)\\
&\leq \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}}\left(\frac{(1-r)\mathsf{snr}_0 + r\mathsf{snr}}{\mathsf{snr}_0}\right)^{-\frac{n}{2r}} n^{\frac{1}{m}-1}\\
&\quad\cdot(\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}_0, m\cdot p))^{\frac{1}{m}}\\
&= \kappa_{n,t}\ \mathrm{mmpe}^{\frac{1-t}{1+t}}\left(\mathbf{X}, \mathsf{snr}_0, \frac{1+t}{1-t}\cdot p\right).
\end{aligned}$$

This concludes the proof. □

The bound in Proposition 19 is the key in showing new bounds on the phase transitions region for the MMSE, presented in the next section.

As an application of Proposition 19 we show that the MMPE is a continuous function of SNR.

Proposition 20: For fixed $\mathbf{X}$ *and* p*,* $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)$ *is a continuous function of* $\mathsf{snr} > 0$*.*

Proof: Assume without loss of generality that $\mathsf{snr}_0 \geq \mathsf{snr}$

$$\begin{aligned}
&\lim_{\mathsf{snr}\to\mathsf{snr}_0} |\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) - \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}_0, p)|\\
&\overset{a)}{=} \lim_{\mathsf{snr}\to\mathsf{snr}_0} \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) - \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}_0, p)\\
&\overset{b)}{\leq} \lim_{\mathsf{snr}\to\mathsf{snr}_0} \kappa_{n,t}\ \mathrm{mmpe}^{\frac{1-t}{1+t}}\left(\mathbf{X}, \mathsf{snr}_0, \frac{1+t}{1-t}\cdot p\right)\\
&\quad - \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0, p)\\
&\overset{c)}{=} \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0, p) - \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0, p) = 0,
\end{aligned}$$

where the (in)-equalities follow from: a) since the MMPE is a decreasing function of SNR and since $\mathsf{snr}_0 \geq \mathsf{snr}$; b) by using Proposition 19; and c) by definition of t in Proposition 19 we have that $\lim_{\mathsf{snr}\to\mathsf{snr}_0} t = 0$ and $\lim_{\mathsf{snr}\to\mathsf{snr}_0} k_{n,t} = 1$, and by continuity of the MMPE in p from Proposition 13. This concludes the proof. □

VIII. Applications

We next show how the MMPE can be used to derive tighter versions of some well known bounds. It is important to point out that even though the focus of this paper is on the AWGN setting, the results that follow (Theorem 1, Theorem 2 and Theorem 3) apply to any additive channel model in which the noise is an absolutely continuous random variable, without the need for the i.i.d. assumption.

A. Bounds on the Differential Entropy

For any random vector $\mathbf{U}$ such that $|\mathbf{K}_\mathbf{U}| < \infty$, $h(\mathbf{U}) < \infty$, and any random vector $\mathbf{V}$, the following inequality is considered to be a continuous analog of Fano's inequality [6]:

$$\begin{aligned}
h(\mathbf{U}|\mathbf{V}) &\leq \frac{n}{2}\log(2\pi\mathrm{e}\ |\mathbf{K}_{\mathbf{U}|\mathbf{V}}|^{\frac{1}{n}})\\
&\leq \frac{n}{2}\log(2\pi\mathrm{e}\ \mathrm{mmse}(\mathbf{U}|\mathbf{V})),
\end{aligned} \tag{62}$$

where the inequality in (62) is a consequence of the arithmetic-mean geometric-mean inequality, that is, for any $0 \preceq \mathbf{A}$ we have used $|\mathbf{A}|^{\frac{1}{n}} = \left(\prod_{i=1}^n \lambda_i\right)^{\frac{1}{n}} \leq \frac{\sum_{i=1}^n \lambda_i}{n} = \frac{\mathrm{Tr}(\mathbf{A})}{n}$ where λ_i's are the eigenvalues of $\mathbf{A}$.

By applying (62) to the AWGN setting, for any $\mathbf{X}$ such that $|\mathbf{K_X}| < \infty$, $h(\mathbf{X}) < \infty$, by using Proposition 10 with $q = 1$, we can arrive at the trivial bound: for any $p \geq 2$

$$h(\mathbf{X}|\mathbf{Y}) \leq \frac{n}{2} \log \left(2\pi \mathrm{e} \cdot n^{\frac{2-p}{p}} \cdot \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p) \right). \quad (63)$$

Next, we show that the inequality in (62) can be generalized in terms of the norm in (5), and the trivial bound in (63) can be improved.

Theorem 1: For any $\mathbf{U} \in \mathbb{R}^n$ *such that* $h(\mathbf{U}) < \infty$ *and* $\|\mathbf{U}\|_p < \infty$ *for some* $p \in (0, \infty)$*, and for any* $\mathbf{V} \in \mathbb{R}^n$*, we have*

$$h(\mathbf{U}|\mathbf{V}) \leq \frac{n}{2} \log \left(k_{n,p}^2 \cdot n^{\frac{2}{p}} \cdot \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{U}|\mathbf{V}; p) \right), \quad (64)$$

where

$$k_{n,p} := \frac{\sqrt{\pi} \left(\frac{p}{n}\right)^{\frac{1}{p}} \mathrm{e}^{\frac{1}{p}} \Gamma^{\frac{1}{n}} \left(\frac{n}{p} + 1\right)}{\Gamma^{\frac{1}{n}} \left(\frac{n}{2} + 1\right)}. \quad (65)$$

Proof: See Appendix J. □

Note that the result in Theorem 1 holds in great generality, i.e., the AWGN assumption is not necessary. As an application of Theorem 1 to the AWGN setting we have the following stronger version of the inequality in (63)

$$h(\mathbf{X}|\mathbf{Y}) \leq \frac{n}{2} \log \left(k_{n,p}^2 \cdot n^{\frac{2}{p}} \cdot \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}, p) \right).$$

B. Generalized Ozarow-Wyner Bound

In [37] the following "Ozarow-Wyner lower bound" on the mutual information achieved by a discrete input X_D transmitted over an AWGN channel was shown:

$$[H(X_D) - \mathsf{gap}]^+ \leq I(X_D; Y) \leq H(X_D), \quad (66a)$$

$$\mathsf{gap} \leq \frac{1}{2} \log \left(\frac{\pi \mathrm{e}}{6} \right) + \frac{1}{2} \log \left(1 + \frac{\mathrm{lmmse}(X, \mathsf{snr})}{d_{\min}(X_D)^2} \right), \quad (66b)$$

where $\mathrm{lmmse}(X, \mathsf{snr})$ is the LMMSE.

The advantage of the bound in (66) compared to existing bounds is its computational simplicity, and the bound has been shown to be useful for problems such as two-user Gaussian interference channels [45], [46], communication with a disturbance constraint [13], energy harvesting problems [47], [48], and information-theoretic security [49].

The bound on the gap in (66) has been sharpened in [45, Remark 2] to

$$\mathsf{gap} \leq \frac{1}{2} \log \left(\frac{\pi \mathrm{e}}{6} \right) + \frac{1}{2} \log \left(1 + \frac{\mathrm{mmse}(X, \mathsf{snr})}{d_{\min}(X_D)^2} \right),$$

since $\mathrm{lmmse}(X, \mathsf{snr}) \geq \mathrm{mmse}(X, \mathsf{snr})$.

Next, we generalize the bound in (66) to discrete vector inputs and give the sharpest known bound on the gap term.

Theorem 2 (Generalized Ozarow-Wyner Bound): Let $\mathbf{X}_D$ *be a discrete random vector with finite entropy, such that* $p_i = \mathbb{P}[\mathbf{X}_D = \mathbf{x}_i]$*, and* $\mathbf{x}_i \in \mathsf{supp}(\mathbf{X}_D)$*. Moreover, for any* $p > 0$ *let* $\mathcal{K}_p$ *be a set of continuous random vectors, independent of* $\mathbf{X}_D$*, such that for every* $\mathbf{U} \in \mathcal{K}_p$*,* $h(\mathbf{U}), \|\mathbf{U}\|_p < \infty$*, and*

$$\mathsf{supp}(\mathbf{U} + \mathbf{x}_i) \cap \mathsf{supp}(\mathbf{U} + \mathbf{x}_j) = \emptyset, \quad \forall\, \mathbf{x}_i, \ \mathbf{x}_j \in \mathsf{supp}(\mathbf{X}_D),\ i \neq j. \quad (67a)$$

Then for any $p > 0$

$$[H(\mathbf{X}_D) - \mathsf{gap}_p]^+ \leq I(\mathbf{X}_D; \mathbf{Y}) \leq H(\mathbf{X}_D), \quad (67b)$$

where

$$n^{-1}\mathsf{gap}_p \leq \inf_{\mathbf{U} \in \mathcal{K}_p} \left(G_{1,p}(\mathbf{U}, \mathbf{X}_D) + G_{2,p}(\mathbf{U}) \right),$$

$$G_{1,p}(\mathbf{U}, \mathbf{X}_D) = \log \left(\frac{\|\mathbf{U} + \mathbf{X}_D - f_p(\mathbf{X}_D|\mathbf{Y})\|_p}{\|\mathbf{U}\|_p} \right) \quad (67c)$$

$$\overset{\text{for } p \geq 1}{\leq} \log \left(1 + \frac{\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}_D, \mathsf{snr}, p)}{\|\mathbf{U}\|_p} \right), \quad (67d)$$

$$G_{2,p}(\mathbf{U}) = \log \left(\frac{k_{n,p} \cdot n^{\frac{1}{p}} \cdot \|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n} h_{\mathrm{e}}(\mathbf{U})}} \right). \quad (67e)$$

Proof: See Appendix K. □

It is interesting to note that the lower bound in (67b) resembles the bound for lattice codes in [50, Th. 1], where $\mathbf{U}$ can be thought of as dither, $G_{2,p}$ corresponds to the log of the normalized p-moment of a compact region in $\mathbb{R}^n$, $G_{1,p}$ corresponds to the log of the normalized MMSE term, and $H(\mathbf{X}_D)$ corresponds with the capacity C.

In order to show the advantage of Theorem 2 over the original Ozarow-Wyner bound (case of $n = 1$ and with LMMSE instead of MMPE), we consider X_D uniformly distributed with the number of points equal to $N = \lfloor \sqrt{1 + \mathsf{snr}} \rfloor$, that is, we choose the number of points such that $H(X_D) \approx \frac{1}{2} \log(1 + \mathsf{snr})$. Fig. 3 shows:

- The solid cyan line is the "shaping loss" $\frac{1}{2} \log \left(\frac{\pi \mathrm{e}}{6} \right)$ for a one-dimensional infinite lattice and is the limiting gap if the number of points N grows faster than $\sqrt{\mathsf{snr}}$;
- The solid magenta line is the gap in the original Ozarow-Wyner bound in (66); and
- The dashed purple, dashed-dotted blue and dotted green lines are the new gap due to Theorem 2 for value of $p = 2, 4, 6$, respectively, and where we chose $U \sim \mathcal{U}\left[-\frac{d_{\min}(X_D)}{2}, \frac{d_{\min}(X_D)}{2} \right]$.

We note that the version of the Ozarow-Wyner bound in Theorem 2 provides the sharpest bound for the gap term. An open question, for $n = 1$, is what value of p provides the smallest gap and if that coincide with the ultimate "shaping loss".

Next we turn our attention to the case of $n > 1$. Another interesting question is how the gap behaves as $n \to \infty$.

Theorem 3: Let $\mathbf{U}$ *be uniform over the ball of radius* $r = \frac{d_{\min}(\mathbf{X}_D)}{2}$*. Then for any* $p > 0$

$$G_{2,p}(\mathbf{U}) = O\left(\frac{1}{n} \log(n) \right), \quad (68a)$$

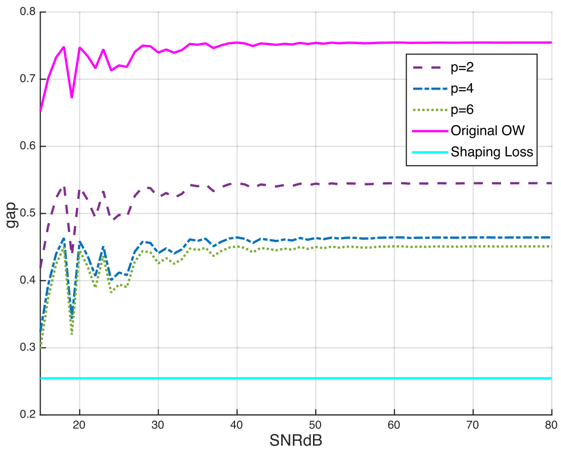


Fig. 3. Gap in equation (66a) and (67) vs. snr.

and therefore $\lim_{n\to\infty} G_{2,p}(\mathbf{U}) = 0$. *Therefore,*

$$\frac{1}{n}H(\mathbf{X}_D) \geq \frac{1}{n}I(\mathbf{X}_D;\mathbf{Y}) \geq \frac{1}{n}H(\mathbf{X}_D) - G_{1,p}(\mathbf{U},\mathbf{X}_D) - O\left(\frac{1}{n}\log(n)\right), \tag{68b}$$

where

$$e^{G_{1,p}(\mathbf{U},\mathbf{X}_D)} \overset{\text{for } p\geq 1}{\leq} 1 + 2\frac{d_{\max}(\mathbf{X}_D)}{d_{\min}(\mathbf{X}_D)}\sqrt[p]{\frac{(p+n)}{n}\bar{Q}\left(\frac{n}{2};\frac{\mathsf{snr}d_{\min}^2(\mathbf{X}_D)}{8}\right)}. \tag{68c}$$

Proof: See Appendix L. □

For recent applications of the bound in Theorem 2 to non-Gaussian and MIMO channels the reader is referred to [51]–[53], respectively.

C. New Bounds on the MMSE and Phase Transitions

The SCPP is instrumental in showing the behavior of the MMSE of capacity achieving codes. For example, as the length of any capacity achieving code goes to infinity, the MMSE behaves as follows:

$$\limsup_{n\to\infty} \mathrm{mmse}(\mathbf{X},\mathsf{snr}) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & 0 \leq \mathsf{snr} \leq \mathsf{snr}_0 \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr}_0 \leq \mathsf{snr} \leq \mathsf{snr}_1 \\ \frac{\gamma}{1+\gamma\,\mathsf{snr}}, & \mathsf{snr} \geq \mathsf{snr}_1, \end{cases} \tag{69}$$

as shown: in [14], for the Gaussian point-to-point channel with the output Y_{snr_0} with $\beta = \gamma = 0$; in [15], for the Gaussian BC with outputs Y_{snr_1} and Y_{snr_0}, where $\mathsf{snr}_0 \leq \mathsf{snr}_1$ and rate pair $(R_1, R_2) = \left(\frac{1}{2}\log(1+\beta\mathsf{snr}_1), \frac{1}{2}\log\left(\frac{1+\mathsf{snr}_0}{1+\beta\mathsf{snr}_0}\right)\right)$ for some $\beta \in [0,1]$, with $\gamma = 0$; in [15], for the Gaussian wiretap channel with outputs Y_{snr_0} (primary) and Y_{snr_1} (eavesdropper) with maximum equivocation $D_{\max}$ and rate $R \geq D_{\max}$, for $\beta = \gamma = 0$; and in [12], for the Gaussian point-to-point channel with output Y_{snr_1} and an MMSE disturbance constraint at Y_{snr_0} measured by $\mathrm{mmse}(\mathbf{X},\mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}$ for some $\beta \in [0,1]$ with $\gamma = \beta$. The jump discontinuities in (69) at $\mathsf{snr} = \mathsf{snr}_0$ and $\mathsf{snr} = \mathsf{snr}_1$ are referred to as the *phase transitions*.

Based on the above, an interesting question is how the MMSE in (69) behaves for codes of finite length. In [13], in order to study the phase transition phenomenon for inputs of finite length, the following optimization problem was proposed:

Definition 4:

$$\mathrm{M}_n(\mathsf{snr},\mathsf{snr}_0,\beta) := \sup_{\mathbf{X}} \mathrm{mmse}(\mathbf{X},\mathsf{snr}), \tag{70a}$$

$$\textit{s.t. } \|\mathbf{X}\|_2^2 \leq 1, \textit{ and } \mathrm{mmse}(\mathbf{X},\mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}, \tag{70b}$$

for some $\beta \in [0,1]$.

Investigation in [13] revealed that $\mathrm{M}_n(\mathsf{snr},\mathsf{snr}_0,\beta)$ in (70a) must be of the following form:

$$\mathrm{M}_n(\mathsf{snr},\mathsf{snr}_0,\beta) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & \mathsf{snr} \leq \mathsf{snr}_L \\ T_n(\mathsf{snr},\mathsf{snr}_0,\beta), & \mathsf{snr}_L \leq \mathsf{snr} \leq \mathsf{snr}_0 \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr}_0 \leq \mathsf{snr}, \end{cases}$$

for some snr_L and some function $T_n(\mathsf{snr},\mathsf{snr}_0,\beta)$, where the region $\mathsf{snr}_L \leq \mathsf{snr} \leq \mathsf{snr}_0$ is referred to as the *phase transition region* and its width is defined as $W(n) := \mathsf{snr}_0 - \mathsf{snr}_L$. In [13] the following was established for $T_n(\mathsf{snr},\mathsf{snr}_0,\beta)$ and $W(n)$:

$$\mathrm{mmse}(\mathbf{X},\mathsf{snr}) \leq \mathrm{mmse}(\mathbf{X},\mathsf{snr}_0) + \kappa_n\left(\frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0}\right), \quad \text{where } \kappa_n \leq n+2, \tag{71}$$

and the width of phase transition region scales as $W(n) = O\left(n^{-1}\right)$.

The main result of this subsection is shown next. It uses Propositions 19 and Proposition 12.

Theorem 4: For $0 < \mathsf{snr} \leq \mathsf{snr}_0$,

$$\mathrm{mmse}(\mathbf{X},\mathsf{snr}) \leq \min_{r>\frac{2}{\gamma}} \kappa(r,\gamma,n)\left(\frac{\beta}{1+\beta\mathsf{snr}_0}\right)^{\frac{\gamma r-2}{r-2}}, \tag{72a}$$

where $\gamma := \frac{\mathsf{snr}}{2\mathsf{snr}_0 - \mathsf{snr}} \in (0,1]$, *and*

$$\kappa(r,\gamma,n) := \frac{\sqrt{2}}{n^{1-\gamma}}\left(\frac{1+\gamma}{\gamma}\right)^{\frac{n(1-\gamma)-1}{2}} M_r^{\frac{2(1-\gamma)}{r-2}}, \tag{72b}$$

$$M_r := \left\|\mathbf{X} - \mathbb{E}\left[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}\right]\right\|_r^r \leq 2^r \min\left(\frac{\|\mathbf{Z}\|_r^r}{\mathsf{snr}_0^{\frac{r}{2}}}, \|\mathbf{X}\|_r^r\right), \tag{72c}$$

and where the minimizing r *in* (72a) *can be approximated by* r_{opt}

$$\approx \begin{cases} 2\ln\left(\frac{4e}{\mathsf{snr}_0\mathrm{mmse}(\mathbf{X},\mathsf{snr}_0)}\right), & \frac{2}{\gamma} \leq \ln\left(\frac{4e}{\mathsf{snr}_0\mathrm{mmse}(\mathbf{X},\mathsf{snr}_0)}\right) \\ \frac{2}{\gamma}, & \frac{2}{\gamma} > \ln\left(\frac{4}{\mathsf{snr}_0\mathrm{mmse}(\mathbf{X},\mathsf{snr}_0)}\right). \end{cases} \tag{72d}$$

Moreover, the width of the phase transition region scales as

$$W(n) = O\left(n^{-\frac{1}{2}}\right). \tag{72e}$$

Proof: From the SCPP complementary bound in Proposition 19 with $p = 1$ we have that

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \kappa_{n,t}\ \mathrm{mmpe}^{\frac{1-t}{1+t}}\left(\mathbf{X}, \mathsf{snr}_0, \frac{1+t}{1-t}\cdot 2\right). \quad (73)$$

From the interpolation result in Proposition 12 letting $q = \frac{1+t}{1-t}\cdot 2$, $p = 2$ we have that for some r such that $2 \le 2\frac{1+t}{1-t} = q < r$ and

$$\alpha = \frac{\frac{1-t}{2(1+t)} - \frac{1}{r}}{\frac{1}{2} - \frac{1}{r}} \Rightarrow 1 - \alpha = \frac{\frac{2t}{1+t} r}{r-2}, \quad (74)$$

and thus the MMPE term can be bounded as

$$\begin{aligned}
&\mathrm{mmpe}^{\frac{1-t}{1+t}}\left(\mathbf{X}, \mathsf{snr}_0, \frac{1+t}{1-t}\cdot 2\right)\\
&\le \mathrm{mmpe}^{\alpha}(\mathbf{X}, \mathsf{snr}_0, 2)\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}]\|_r^{2(1-\alpha)}\\
&= \mathrm{mmse}^{\alpha}(\mathbf{X}, \mathsf{snr}_0)\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}]\|_r^{2(1-\alpha)}\\
&= \mathrm{mmse}^{\alpha}(\mathbf{X}, \mathsf{snr}_0)\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}]\|_r^{2r\frac{\frac{2t}{1+t}}{r-2}}.
\end{aligned}$$

By Proposition 10 we can bound $\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}]\|_r^{2r\frac{\frac{2t}{1+t}}{r-2}}$ as follows:

$$\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}]\|_r^{r\frac{\frac{4t}{1+t}}{r-2}} \le \left(2^r \min\left(\frac{\|\mathbf{Z}\|_r^r}{\mathsf{snr}^{\frac{r}{2}}}, \|\mathbf{X}\|_r^r\right)\right)^{\frac{4t}{(1+t)(r-2)}}.$$

By putting all of the bounds together, letting $\gamma = \frac{1-t}{1+t}$ and observing that

$$\begin{aligned}
1 - \gamma &= \frac{2t}{1+t},\\
\gamma &= \frac{\mathsf{snr}}{2\mathsf{snr}_0 - \mathsf{snr}},\\
\frac{\mathsf{snr}_0}{\mathsf{snr}} &= \frac{1+\gamma}{2\gamma},
\end{aligned}$$

we get the bound in (72a). Finally, the proof of approximately optimal r in (72d) is given in Appendix M. $\square$

The bounds in Theorem 4 and in (71) are shown in Fig. 4. The bound in Theorem 4 is asymptotically tighter than the one in (71). This follows since the phase transition region shrinks as $O\left(\frac{1}{\sqrt{n}}\right)$ for Theorem 4, and as $O\left(\frac{1}{n}\right)$ for the bound in (71). It is not possible in general to assert that Theorem 4 is tighter than (71). In fact, for small values of n, the bound in (71) can offer advantages, as seen for the case $n = 1$ shown in Fig. 4b. Another advantage of the bound in (71) is its analytical simplicity.

D. Bounds on the Derivative of the MMSE

The MMPE can be used to study the second derivative of mutual information (or first derivative of MMSE), as initiated for $n = 1$ in [9] and for $n \ge 1$ in [10], namely,

$$\begin{aligned}
\frac{d^2 I(\mathbf{X}, \mathbf{Y})}{d\mathsf{snr}^2} &= n\frac{d\ \mathrm{mmse}(\mathbf{X}, \mathsf{snr})}{d\mathsf{snr}}\\
&= -\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right),\\
\mathbf{Cov}(\mathbf{X}|\mathbf{Y}) &:= \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^T|\mathbf{Y}\right]. \quad (75)
\end{aligned}$$

The second derivative of mutual information is important in characterizing the bandwidth-power trade-off in the wideband regime [54] and [55], and has also been used in the proof of the SCPP in [9] and [10]. Moreover, in [9] it has been shown that the derivative of the MMSE and the quantity in (13) are related by the following bound for $n = 1$:

$$\mathbb{E}\left[\mathrm{Cov}^2(X|Y)\right] \le \|X - \mathbb{E}[X|Y]\|_4^4 \le \frac{3\cdot 2^4}{\mathsf{snr}^2}. \quad (76)$$

The main result of this subsection is the next bound.

Proposition 21: For any input $\mathbf{X}$

$$\begin{aligned}
&\mathrm{mmse}^2(\mathbf{X}, \mathsf{snr})\\
&= \mathrm{mmpe}^2(\mathbf{X}, \mathsf{snr}, 2)\\
&\le \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right) \le n\ \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, 4). \quad (77)
\end{aligned}$$

Proof: See Appendix N. $\square$

It can be observed that, for the case $n = 1$, by using the bound in (38b) from Proposition 10 we have that

$$\mathbb{E}\left[\mathrm{Cov}^2(X|Y)\right] \le \mathrm{mmpe}(X, \mathsf{snr}, 4) \le \frac{3}{\mathsf{snr}^2}, \quad (78)$$

which significantly reduces the constant in (76) from $3\cdot 2^4$ to 3. For a similar but slightly different bound than that in (78) on $\mathbb{E}\left[\mathrm{Cov}^2(X|Y)\right]$ please see [13].

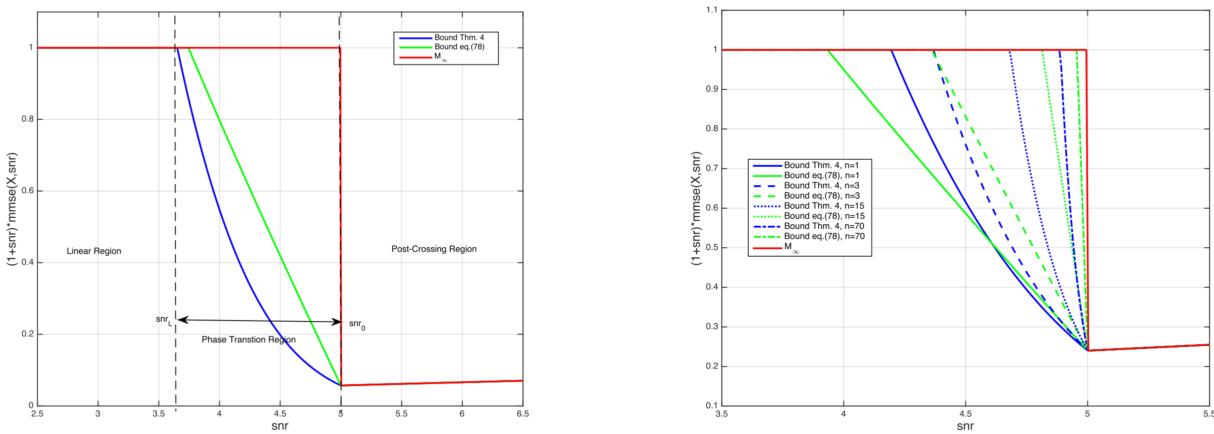
IX. Concluding Remarks

This paper has considered the problem of estimating a random variable from a noisy observation under a general cost function, termed the MMPE. We have shown that many properties of the MMSE and the conditional expectation (i.e., optimal MMSE estimator) are identical or have a natural generalization to the MMPE and the MMPE optimal estimator.

We have also provided a new simpler proof of the SCPP for the MMSE and generalized it to the MMPE. We have shown that the new framework of the MMPE also permits the development of bounds that are complementary to the SCPP which in turn allows for new tighter characterizations of the phase transition phenomena that manifest, in the limit as the length of the capacity achieving code goes to infinity, as a discontinuity of the MMSE as a function of SNR.

We have also shown connections between the MMPE and the conditional differential entropy by generalizing a well know continuous analog of Fano's inequality. The MMPE was further used to refine bounds on the conditional entropy and improve the gap term in the Ozarow-Wyner bound.

Currently, we are investigating the connections between bounds on the MMPE provided in this work and the rate distortion problem with the MMPE distortion measure. Possible future applications of the sharpened version of the Ozarow-Wyner bound include sharpening the bounds on discrete inputs in [56] and [57]. Another interesting future direction is to consider a modified 'information bottleneck problem' [58] where the constraint on the mutual information is replaced by a constraint on the MMPE.



(a) (b)

Fig. 4. Bounds on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs snr. (a) For $\mathsf{snr}_0 = 5$ and $\beta = 0.01$. Here $n = 1$. (b) For $\mathsf{snr}_0 = 5$ and $\beta = 0.05$. Several values of n.

APPENDIX A
PROOF OF PROPOSITION 1

For simplicity, we look at the case $n = 1$. The case for $n > 1$ follows similarly. We first assume that $\mathsf{snr} > 0$. The first direction follows trivially:

$$\inf_f \mathbb{E}\left[|X - f(Y)|^p\right] \le \mathbb{E}\left[|X - f_p(X|Y)|^p\right].$$

The other direction follows by using

$$\inf_f \mathbb{E}\left[|X - f(Y)|^p\right] \ge \mathbb{E}\left[\inf_f \mathbb{E}\left[|X - f(Y)|^p \,\middle|\, Y\right]\right],$$

where we focus on the inner expectation $\inf_f \mathbb{E}\left[|X - f(Y)|^p | Y = y\right]$ and show that the infimum is achieved by $f(y) = f_p(X|Y = y)$ given in (14). Since y is now given, we are simply looking for an optimal solution to the more general problem

$$\inf_{\upsilon \in \mathbb{R}} \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right], \quad (79)$$

where $X_y \sim p_{X|Y}(\cdot|y)$. The goal is to show that the infimum in (79) is achievable. Clearly, the infimum exists since

$$\begin{aligned} 0 \le \inf_{\upsilon \in \mathbb{R}} \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right] &\le \mathbb{E}\left[\left|X_y - 0\right|^p\right] \\ &= \mathbb{E}[|X_y|^p] < \infty, \end{aligned} \quad (80)$$

where the last inequality follows from [9, Proposition 6] which asserts that for any $p < \infty$, X_y is a sub-Gaussian random variable and hence all conditional moments are finite.

Next, we show that $g(\upsilon) = \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right]$ is a continuous function of υ. Recall, that any given function $h(x)$ is continuous if $x_n \to x$ implies $h(x_n) \to h(x)$ as $n \to \infty$.

For arbitrary $|\upsilon| < \infty$ take a sequence υ_n such that $\upsilon_n \to \upsilon$, we want to show that

$$\begin{aligned} \lim_{n\to\infty} g(\upsilon_n) &= \lim_{n\to\infty} \mathbb{E}\left[\left|X_y - \upsilon_n\right|^p\right] \\ &= \mathbb{E}\left[\lim_{n\to\infty} \left|X_y - \upsilon_n\right|^p\right] = g(\upsilon). \end{aligned}$$

This can be done with the help of the dominated convergence theorem. We must find an integrable random variable θ such that $\left|X_y - \upsilon_n\right|^p \le \theta$ for all n; this is found as

$$\begin{aligned} \left|X_y - \upsilon_n\right|^p &\overset{a)}{\le} 2^p \left(|X_y|^p + |\upsilon_n|^p\right) \\ &\overset{b)}{\le} 2^p \left(|X_y|^p + K\right) = \theta, \end{aligned}$$

where the inequalities follow from: a) $\left|X_y - \upsilon_n\right|^p \le (2\max(|X_y|, |\upsilon_n|))^p \le 2^p \left(|X_y|^p + |\upsilon_n|^p\right)$ which holds for any $p \ge 0$; and b) recall that every convergent sequence is bounded and since the sequence υ_n converges to υ it is also bounded by some finite K for every n. The integrability of $\theta = 2^p \left(|X_y|^p + K\right)$ follows again by the sub-Gaussian argument from [9, Proposition 6]. Therefore, we conclude that the function $g(\upsilon)$ is continuous.

Next, we show that the infimum is attainable by some $|\upsilon_0| < \infty$. By definition of the infimum there exists some υ_n (not necessarily convergent) such that

$$\liminf_{n\to\infty} \mathbb{E}[|X_y - \upsilon_n|^p] = \inf_{\upsilon \in \mathbb{R}} \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right].$$

Towards a contradiction, assume that $\upsilon_n \to \infty$. Then by Fatou's lemma

$$\liminf_{n\to\infty} \mathbb{E}[|X_y - \upsilon_n|^p] \ge \mathbb{E}[\liminf_{n\to\infty} |X_y - \upsilon_n|^p] = \infty.$$

However, this contradicts the result in (80) and therefore sequence υ_n must be bounded. This, together with the fact that $g(\upsilon)$ is continuos, implies that the infimum is attainable and thus

$$\inf_{\upsilon \in \mathbb{R}} \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right] = \min_{\upsilon \in \mathbb{R}} \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right]. \quad (81)$$

Therefore, for each $y \in \mathbb{R}$ there exists $|\upsilon| < K$ that minimize the expression $\min_{\upsilon \in \mathbb{R}} \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right]$. Note, that the optimizing υ might not be unique and the set of optimal values is given by

$$S_y = \{\upsilon : \min_{\upsilon \in \mathbb{R}} \mathbb{E}\left[\left|X_y - \upsilon\right|^p\right]\}.$$

According to the Definition 2 we have that

$$\begin{aligned} f_p(X|Y = y) &= \max\{\upsilon : S_y\} \\ &= \max\{\upsilon : \min_{\upsilon \in \mathbb{R}} \mathbb{E}\left[|X - \upsilon|^p \,|Y = y\right]\}. \end{aligned} \quad (82)$$

Note, that we have show that for every y the optimal value v is bounded and therefore in (82) we can take the max instead of the sup.

Moreover, as will be shown in Proposition 2, due to the strict convexity of $|\cdot|^p$ for $p > 1$ the optimizer is indeed unique and can be given by

$$f_p(X|Y = y) = \arg\min_{v\in\mathbb{R}} \mathbb{E}\left[|X - v|^p \,|Y = y\right].$$

For the case of $\mathsf{snr} = 0^+$ the problem reduces to

$$\inf_{\mathbf{v}\in\mathbb{R}^n} \|\mathbf{X} - \mathbf{v}\|_p,$$

which is bounded if and only if $\|\mathbf{X}\|_p < \infty$. This concludes the proof.

Appendix B
Proof of Proposition 2

We take a classical approach used in estimation theory to find an optimal estimator by using tools from calculus of variations [59, Ch.7, Th.1]. A necessary condition for f to be a minimizer in (14) is expressed through a functional derivative as

$$\begin{aligned}&\nabla_g \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^p\right]\\ &= \lim_{\epsilon\to 0} \mathbb{E}\left[\frac{\|\mathbf{X} - (f(\mathbf{Y}) + \epsilon g(\mathbf{Y}))\,\|^p - \|\mathbf{X} - f(\mathbf{Y})\|^p}{\epsilon}\right] = 0,\end{aligned} \tag{83}$$

for all admissible $g(\mathbf{Y})$.

Therefore, we focus on the following limit:

$$\lim_{\epsilon\to 0} \mathbb{E}\left[\frac{\|\mathbf{X} - (f(\mathbf{Y}) + \epsilon g(\mathbf{Y}))\,\|^p - \|\mathbf{X} - f(\mathbf{Y})\|^p}{\epsilon}\right]. \tag{84}$$

We seek to apply the dominated convergence theorem to (84) in order to interchange the order of the limit and the expectation. To that end we let $\mathbf{v} = \mathbf{x} - f(\mathbf{y})$ and

$$\|\mathbf{x} - f(\mathbf{y})\|^p = \left(\mathbf{v}^T\mathbf{v}\right)^{\frac{p}{2}},$$

and

$$\begin{aligned}&\|\mathbf{x} - (f(\mathbf{y}) + \epsilon g(\mathbf{y}))\,\|^p\\ &\quad= \left((\mathbf{v} - \epsilon g(\mathbf{y}))^T(\mathbf{v} - \epsilon g(\mathbf{y}))\right)^{\frac{p}{2}}\\ &\quad= \left(\mathbf{v}^T\mathbf{v} - \epsilon g(\mathbf{y})^T\mathbf{v} - \epsilon\mathbf{v}^T g(\mathbf{y}) + \epsilon^2 g(\mathbf{y})^T g(\mathbf{y})\right)^{\frac{p}{2}}.\end{aligned}$$

Next for the integrand

$$\frac{\|\mathbf{x} - (f(\mathbf{y}) + \epsilon g(\mathbf{y}))\,\|^p - \|\mathbf{x} - f(\mathbf{y})\|^p}{\epsilon} \tag{85}$$

we observe that all the terms in (85) are of order no more than p, and since all of the terms are in L_p (or p integrable) the quantity in (85) is integrable for any ϵ. Therefore, the dominated convergence theorem applies and we can interchange the order of limit and expectation in (84).

Next, observe that we can re-write the limit as a derivative, that is,

$$\begin{aligned}&\lim_{\epsilon\to 0} \frac{\|\mathbf{x} - (f(\mathbf{y}) + \epsilon g(\mathbf{y}))\,\|^p - \|\mathbf{x} - f(\mathbf{y})\|^p}{\epsilon}\\ &\qquad= \frac{d}{d\epsilon}\|\mathbf{x} - (f(\mathbf{y}) + \epsilon g(\mathbf{y}))\,\|^p\Big|_{\epsilon=0}.\end{aligned} \tag{86}$$

By using chain rules of differentiation of matrix calculus we arrive at

$$\begin{aligned}&\frac{d}{d\epsilon}\|\mathbf{x} - (f(\mathbf{y}) + \epsilon g(\mathbf{y}))\,\|^p\Big|_{\epsilon=0}\\ &= \frac{d}{d\epsilon}\left((\mathbf{v} - \epsilon g(\mathbf{y}))^T(\mathbf{v} - \epsilon g(\mathbf{y}))\right)^{\frac{p}{2}}\Big|_{\epsilon=0}\\ &= -p\left((\mathbf{v} - \epsilon g(\mathbf{y}))^T(\mathbf{v} - \epsilon g(\mathbf{y}))\right)^{\frac{p}{2}-1}(\mathbf{v} - \epsilon g(\mathbf{y}))^T g(\mathbf{y})\Big|_{\epsilon=0}\\ &= -p\,\mathrm{Tr}^{\frac{p}{2}-1}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{v}^T g(\mathbf{y}).\end{aligned} \tag{87}$$

Therefore, the function derivative is given by

$$\begin{aligned}&\lim_{\epsilon\to 0} \mathbb{E}\left[\frac{\|\mathbf{X} - (f(\mathbf{Y}) + \epsilon g(\mathbf{Y}))\,\|^p - \|\mathbf{X} - f(\mathbf{Y})\|^p}{\epsilon}\right]\\ &\qquad= \mathbb{E}\left[-p\cdot\|\mathbf{X} - f(\mathbf{Y})\|^{p-2}(\mathbf{X} - f(\mathbf{Y}))^T g(\mathbf{Y})\right].\end{aligned}$$

Finally, for $f_p(\mathbf{X}|\mathbf{Y})$ to be optimal it must satisfy

$$\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^{p-2}(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}))^T g(\mathbf{Y})\right] = 0,$$

for any admissible $g(\mathbf{Y})$. This verifies the necessary condition for optimality for $p > 0$.

To verify that this is a sufficient condition for optimality we take the second variational derivative of $\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y})\|^p\right]$ and demonstrated that it is always positive for $p \geq 1$. The fact that

$$\frac{d^2}{d\epsilon^2}\|\mathbf{x} - (f(\mathbf{y}) + \epsilon g(\mathbf{y}))\,\|^p\Big|_{\epsilon=0} \geq 0 \text{ for } p \geq 1,$$

follows since $\|\mathbf{x} - (f(\mathbf{y}) + \epsilon g(\mathbf{y}))\,\|^p$ is a convex function of ϵ for $p \geq 1$.

This verifies the sufficient condition for $p \geq 1$ and concludes the proof.

Appendix C
Proof of Proposition 3

In Proposition 1 we let $X_y \sim p_{X|Y}(\cdot|y)$ and therefore have to solve for all y

$$\min_{v\in\mathbb{R}} \mathbb{E}\left[|X_y - v|^p\right]. \tag{88}$$

We know that X_y is Gaussian with $X_y \sim \mathcal{N}\left(\frac{\sqrt{\mathsf{snr}}y}{1+\mathsf{snr}}, \frac{1}{1+\mathsf{snr}}\right)$. The optimization problem in (88) can be transformed into

$$\min_{v\in\mathbb{R}} \mathbb{E}\left[\left|\frac{Z}{\sqrt{1+\mathsf{snr}}} + \frac{\sqrt{\mathsf{snr}}y}{1+\mathsf{snr}} - v\right|^p\right] = \frac{\min_{a\in\mathbb{R}} \mathbb{E}\left[|Z - a|^p\right]}{(1+\mathsf{snr})^p} \tag{89}$$

where

$$a = \sqrt{1+\mathsf{snr}}\; v - \frac{\sqrt{\mathsf{snr}}y}{\sqrt{1+\mathsf{snr}}}, \tag{90}$$

and where $Z \sim \mathcal{N}(0, 1)$. Next, by taking the derivative with respect to a in (89)

$$\begin{aligned}f'(a) &= \frac{d}{da}\mathbb{E}\left[|Z - a|^p\right] = \mathbb{E}\left[\frac{d}{da}|Z - a|^p\right]\\ &= \mathbb{E}\left[-p\,\mathsf{sign}(Z - a)\,|Z - a|^{p-1}\right],\end{aligned} \tag{91}$$

where the interchange of the order of differentiation and expectation in (91) is possible by Leibniz integral rule [60] which requires verifying that for

$$g(a,z) = \frac{d}{da}|z-a|^p = -p\ \mathsf{sign}(z-a)\,|z-a|^{p-1}\,, \quad (92)$$

we have that $|g(a,z)| \le \theta(z)$ where $\theta(z)$ is integrable. This is indeed the case since

$$|p\ \mathsf{sign}(z-a)\,|z-a|^{p-1}\,| \le p2^p\left(|z|^{p-1}+|a|^{p-1}\right) = \theta(z).$$

Clearly, $\theta(z)$ is integrable, so the change of the order of differentiation and expectation in (89) is justified.

Next, observe that for a fixed a the function $g(z,a)$ in (92) is a decreasing function of z for any $p \ge 1$ and in addition $g(z,a)$ is an odd function around $z = a$. Since $f'(a)$ is an average value of $g(a,z)$ this means that the sign of $f'(a)$ is the same as the sign of a, that is, $f'(a) > 0$ if $a > 0$ and $f'(a) < 0$ if $a < 0$. Moreover, if $a = 0$

$$f'(a=0) = \mathbb{E}\left[-p\ \mathsf{sign}(Z)\,|Z|^{p-1}\right] = 0.$$

All this implies that $a = 0$ is a critical and a minimum point. Therefore, the optimal $\hat{a} = 0$ for the optimization problem in (89) and the optimal $\hat{v}$ for the original optimization problem is found through (90) to be

$$\hat{v} = \frac{\sqrt{\mathsf{snr}}\ y}{1+\mathsf{snr}}.$$

Finally, we compute the mmpe(X,snr,p) for $X \sim \mathcal{N}(0,1)$

$$\begin{aligned}
\text{mmpe}(X,\mathsf{snr},p) &= \mathbb{E}\left[\left|X - \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}Y\right|^p\right] \\
&= \mathbb{E}\left[\left|\frac{X}{1+\mathsf{snr}} - \frac{\sqrt{\mathsf{snr}}Z}{1+\mathsf{snr}}\right|^p\right] \\
&\overset{a)}{=} \mathbb{E}\left[\left|\frac{\hat{Z}}{\sqrt{1+\mathsf{snr}}}\right|^p\right] \\
&\overset{b)}{=} \frac{2^{\frac{p}{2}}\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}(1+\mathsf{snr})^{\frac{p}{2}}},
\end{aligned}$$

where the equalities follow from: a) follows since X and Z are independent Gaussian r.v.'s and have an equivalent distribution given by $\frac{\hat{Z}}{\sqrt{1+\mathsf{snr}}}$ where $\hat{Z} \sim \mathcal{N}(0,1)$; and b) follows from (7) by setting $n = 1$. This concludes the proof.

APPENDIX D
PROOF OF PROPOSITION 4

From Proposition 1 we have to minimize $\mathbb{E}\left[|X_y - v|^p\right]$ where $X_y \sim p_{X|Y}(\cdot|y)$. We have that the joint probability density function of (X,Y) is given by

$$p_{X,Y}(x,y) = \frac{q}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2}}\delta(x-x_1) + \frac{1-q}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2}}\delta(x-x_2).$$

Without loss of generality we assume that $x_1 \le x_2$. By using Bayes' formula we have that

$$\mathbb{E}\left[|X_y - v|^p\right] = \frac{|x_1-v|^p\frac{q}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2}} + |x_2-v|^p\frac{1-q}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2}}}{p_Y(y)}. \quad (93)$$

The minimization of (93) with respect to v is equivalent to minimizing

$$g(v) = a|x_1 - v|^p + |x_2 - v|^p,$$

where

$$a = \frac{q\ \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2}}}{(1-q)\ \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2}}}.$$

In piecewise form we can write $g(v)$ as

$$g(v) = \begin{cases} a(v-x_1)^p + (v-x_2)^p, & x_2 \le v \\ a(v-x_1)^p + (x_2-v)^p, & x_1 < v < x_2 \\ a(x_1-v)^p + (x_2-v)^p, & v \le x_1, \end{cases}$$

with the derivative of $g(v)$ given by

$$g'(v) = \begin{cases} ar(v-x_1)^{p-1} + r(v-x_2)^{p-1}, & x_2 \le v \\ ar(v-x_1)^{p-1} - r(x_2-v)^{p-1}, & x_1 < v < x_2 \\ -ar(x_1-v)^{p-1} - r(x_2-v)^{p-1}, & v \le x_1, \end{cases} \quad (94)$$

From (94) we see that for the regime $x_2 \le v$ the derivative is positive and therefore the minimum occurs at $v = x_2$. For the regime $v \le x_1$ we have that the derivative is always negative so the minimum occurs at $v = x_1$. For the regime $x_1 < v < x_2$ the optimal v soves

$$g'(v) = ap(v-x_1)^{p-1} - p(x_2-v)^{p-1} = 0,$$

that is,

$$v = \frac{a^{\frac{1}{p-1}}x_1 + x_2}{a^{\frac{1}{p-1}} + 1}.$$

Next, by comparing the three candidates for the minimizing v, we have that

$$\begin{aligned} g(v = x_2) &= a|x_2 - x_1|^p, \\ g(v = x_1) &= |x_2 - x_1|^p, \end{aligned}$$

and

$$\begin{aligned}
&g\left(v = \frac{a^{\frac{1}{p-1}}x_1 + x_2}{a^{\frac{1}{p-1}} + 1}\right) \\
&= a\left|x_1 - \frac{a^{\frac{1}{p-1}}x_1 + x_2}{a^{\frac{1}{p-1}} + 1}\right|^p + \left|x_2 - \frac{a^{\frac{1}{p-1}}x_1 + x_2}{a^{\frac{1}{p-1}} + 1}\right|^p \\
&= \frac{a}{(a^{\frac{1}{p-1}}+1)^p}|x_1 - x_2|^p + \frac{a^{\frac{p}{p-1}}}{(a^{\frac{1}{p-1}}+1)^p}|x_2 - x_1|^p \\
&= |x_2 - x_1|^p\,\frac{a}{(a^{\frac{1}{p-1}}+1)^{p-1}}.
\end{aligned}$$

Since $\frac{a}{(a^{\frac{1}{p-1}}+1)^{p-1}} \le \min(1,a)$, we have that the minimum of $g(v)$ occurs at

$$v = \frac{a^{\frac{1}{p-1}}x_1 + x_2}{a^{\frac{1}{p-1}}+1} = \frac{q^{\frac{1}{p-1}} \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2(r-1)}} \cdot x_1 + (1-q)^{\frac{1}{p-1}} \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2(p-1)}} \cdot x_2}{q^{\frac{1}{p-1}}\ \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2(p-1)}} + (1-q)^{\frac{1}{p-1}} \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2(p-1)}}}. \tag{95}$$

Therefore, the optimal estimator is given by the RHS of (95).

Note, that for the case of $p=1$ the function $g(v)$ reduces to

$$g(v) = a|x_1 - v| + |x_2 - v|,$$

and the minimum occurs at

$$v = \begin{cases} x_1, & a \ge 1 \\ x_2, & a < 1. \end{cases}$$

This implies that for $p=1$ the optimal estimator is

$$f_p(X|Y=y) = \begin{cases} x_1, & a \ge 1 \\ x_2, & a < 1, \end{cases}$$

where $a = \frac{q\ \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_1)^2}{2}}}{(1-q)\ \mathrm{e}^{-\frac{(y-\sqrt{\mathsf{snr}}x_2)^2}{2}}}$. This concludes the proof.

Appendix E
Proof of Proposition 5

The key to deriving all of the claimed properties is the expression of the optimal estimator in Proposition 1. We prove next all the properties.

1) For $0 \le X \in \mathbb{R}$ suppose that

$$0 > v_y = f_p(X|Y=y) = \arg\min_v \mathbb{E}\left[|X-v|^p|Y=y\right],$$

then

$$\begin{aligned} \min_v \mathbb{E}\left[|X-v|^p|Y=y\right] &= \mathbb{E}\left[|X-v_y|^p|Y=y\right] \\ &\overset{a)}{=} \mathbb{E}\left[(X-v_y)^p|Y=y\right] \\ &\overset{b)}{\ge} \mathbb{E}\left[X^p|Y=y\right], \end{aligned} \tag{96}$$

where the (in)-equalities follow from: a) using the assumption that $X \ge 0$ and $v_y < 0$ so $X - v_y > 0$ and the absolute value is redundant; and b) by using the assumption that $X \ge 0$ and $v_y < 0$ then $X - v_y \ge X$. The expression in (96) leads to a contradiction since it implies that $v_y = 0$ but by assumption $v_y < 0$. Therefore, $v_y = f_p(X|Y=y) \ge 0$. This concludes the proof of property 1).

2) Next we show that $f_p(a\mathbf{X}+b|\mathbf{Y}) = af_p(\mathbf{X}|\mathbf{Y})+b$. Let

$$\mathbf{v_y} = f_p(\mathbf{X}|\mathbf{Y}=\mathbf{y}) = \arg\min_{\mathbf{v}} \mathbb{E}\left[\|\mathbf{X}-\mathbf{v}\|^p|\mathbf{Y}=\mathbf{y}\right],$$

then

$$\begin{aligned} &f_p(a\mathbf{X}+b|\mathbf{Y}=\mathbf{y}) \\ &= \arg\min_{\mathbf{v}} \mathbb{E}\left[\|a\mathbf{X}+b-\mathbf{v}\|^p \middle| \mathbf{Y}=\mathbf{y}\right] \\ &= \arg\min_{\mathbf{v}} a^p\ \mathbb{E}\left[\left\|\mathbf{X}-\frac{\mathbf{v}-b}{a}\right\|^p \middle| \mathbf{Y}=\mathbf{y}\right] \\ &\overset{a)}{=} \arg\min_{\mathbf{v}} \mathbb{E}\left[\left\|\mathbf{X}-\frac{\mathbf{v}-b}{a}\right\|^p \middle| \mathbf{Y}=\mathbf{y}\right] \\ &\overset{b)}{=} a\mathbf{v_y}+b \\ &= af_p(\mathbf{X}|\mathbf{Y}=\mathbf{y})+b, \end{aligned}$$

where the equalities follow from: a) since scaling the objective function does not change the optimizer; and b) since the minimum is attained at $\frac{\mathbf{v}-b}{a} = \mathbf{v_y}$. This concludes the proof of property 2).

3) Next, we show that $f_p(g(\mathbf{Y})|\mathbf{Y}=\mathbf{y}) = g(\mathbf{Y})$. Since,

$$\begin{aligned} &f_p(g(\mathbf{Y})|\mathbf{Y}=\mathbf{y}) \\ &= \arg\min_{\mathbf{v}} \mathbb{E}\left[\|g(\mathbf{Y})-\mathbf{v}\|^p|\mathbf{Y}=\mathbf{y}\right] \\ &= \arg\min_{\mathbf{v}} \int \|g(\mathbf{y})-\mathbf{v}\|^p p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})d\mathbf{x} \\ &= \arg\min_{\mathbf{v}} \|g(\mathbf{y})-\mathbf{v}\|^p \\ &= g(\mathbf{y}). \end{aligned}$$

This concludes the proof of property 3).

4) Follows from property 3) by taking $g(\mathbf{Y}) = f_p(\mathbf{X}|\mathbf{Y})$.

5) Observe that for the Markov chain $\mathbf{X} \to \mathbf{Y}_{\mathsf{snr}_0} \to \mathbf{Y}_{\mathsf{snr}}$ we have

$$p_{\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0},\mathbf{Y}_{\mathsf{snr}}}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}_\mathbf{0}},\mathbf{y}_{\mathsf{snr}}) = p_{\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}_\mathbf{0}}). \tag{97}$$

By using Proposition 1 we have that

$$\begin{aligned} &f_p\left(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}=\mathbf{y}_{\mathsf{snr}_\mathbf{0}},\mathbf{Y}_{\mathsf{snr}}=\mathbf{y}_{\mathsf{snr}}\right) \\ &= \arg\min_{\mathbf{v}\in\mathbb{R}^n} \mathbb{E}\left[\|\mathbf{X}-\mathbf{v}\|^p|\mathbf{Y}_{\mathsf{snr}_0}=\mathbf{y}_{\mathsf{snr}_\mathbf{0}},\mathbf{Y}_{\mathsf{snr}}=\mathbf{y}_{\mathsf{snr}}\right] \\ &= \arg\min_{\mathbf{v}\in\mathbb{R}^n} \int \|\mathbf{x}-\mathbf{v}\|^p p_{\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0},\mathbf{Y}_{\mathsf{snr}}}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}_\mathbf{0}},\mathbf{y}_{\mathsf{snr}})\mathbf{dx} \\ &\overset{a)}{=} \arg\min_{\mathbf{v}\in\mathbb{R}^n} \int \|\mathbf{x}-\mathbf{v}\|^p p_{\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}_\mathbf{0}})\mathbf{dx} \\ &= \arg\min_{\mathbf{v}\in\mathbb{R}^n} \mathbb{E}\left[\|\mathbf{X}-\mathbf{v}\|^p|\mathbf{Y}_{\mathsf{snr}_0}=\mathbf{y}_{\mathsf{snr}_\mathbf{0}}\right] \\ &= f_p\left(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}=\mathbf{y}_{\mathsf{snr}_\mathbf{0}}\right) \end{aligned}$$

where the equality in a) follows from (97).

6) See Fig. 1a for the counter example.

This concludes the proof.

Appendix F
Proof of the Bound in Proposition 8

We define

$$\hat{\mathbf{Y}}_{\mathsf{snr}} = \mathbf{Y}_{\mathsf{snr}_0} + \mathbf{Z}',$$

where $\mathbf{Z}' \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma^2 = \frac{\mathsf{snr}_0 - \mathsf{snr}}{\mathsf{snr}}$ is independent of $\mathbf{Y}_{\mathsf{snr}_0}$, $\mathbf{X}$ and $\mathbf{Z}$. Observe that $\hat{\mathbf{Y}}_{\mathsf{snr}}$ and $\mathbf{Y}_{\mathsf{snr}}$ have the same SNR's and therefore

$$\mathrm{mmpe}(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}};p) = \mathrm{mmpe}(\mathbf{X}|\hat{\mathbf{Y}}_{\mathsf{snr}};p). \tag{98}$$

By performing a change of measure we have

$$n \text{ mmpe}(\mathbf{X}|\hat{\mathbf{Y}}_{\mathsf{snr}}; p) = \inf_f \mathbb{E}\left[\|\mathbf{X} - f(\hat{\mathbf{Y}}_{\mathsf{snr}})\|^p\right] = \inf_f \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^p L(\mathbf{X}, \mathbf{Y}_{\mathsf{snr}_0})\right],$$

where $L(\mathbf{x}, \mathbf{y})$ is given by

$$L(\mathbf{x}, \mathbf{y}) = \frac{p_{\hat{\mathbf{Y}}_{\mathsf{snr}}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{p_{\mathbf{Y}_{\mathsf{snr}_0}|\mathbf{X}}(\mathbf{y}|\mathbf{x})} = \frac{\frac{1}{\sqrt{(2\pi)^n(1+\sigma^2)}} \mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\sqrt{\mathsf{snr}_0}\mathbf{x})^T \frac{1}{1+\sigma^2}\mathbf{I}(\mathbf{y}-\sqrt{\mathsf{snr}_0}\mathbf{x})}}{\frac{1}{\sqrt{(2\pi)^n}} \mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\sqrt{\mathsf{snr}_0}\mathbf{x})^T \mathbf{I}(\mathbf{y}-\sqrt{\mathsf{snr}_0}\mathbf{x})}},$$

and thus

$$\begin{aligned} &n \,\text{mmpe}(\mathbf{X}|\hat{\mathbf{Y}}_{\mathsf{snr}}; p) \\ &= \inf_f \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^p L(\mathbf{X}, \mathbf{Y}_{\mathsf{snr}_0})\right] \\ &= \inf_f \frac{1}{\sqrt{1+\sigma^2}} \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^p \mathrm{e}^{\frac{1}{2}\mathbf{Z}^T\mathbf{I}\mathbf{Z} - \frac{1}{2}\mathbf{Z}^T\frac{1}{1+\sigma^2}\mathbf{I}\mathbf{Z}}\right] \\ &= \inf_f \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^p \mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}\mathbf{Z}^T\mathbf{Z}}\right] \\ &= \inf_f \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{Y}_{\mathsf{snr}_0})\|^p \mathrm{e}^{\frac{\mathsf{snr}_0-\mathsf{snr}}{2\mathsf{snr}_0}\sum_{i=1}^n Z_i^2}\right]. \end{aligned}$$

This concludes the proof.

Appendix G
Proof of Proposition 10

A. Proof of the Bound in (38a)

The upper bound in (38a) follows from the fact that $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ is in general a suboptimal estimator for a given p thus

$$\text{mmpe}(\mathbf{X}, \mathsf{snr}, p) \le \frac{1}{n}\mathbb{E}\left[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|^p\right].$$

The lower bound in (38a) for $p \ge q$ follows by

$$\begin{aligned} \text{mmpe}(\mathbf{X}, \mathsf{snr}, q) &= \inf_f \frac{1}{n}\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{X}|\mathbf{Y})\|^q\right] \\ &= \inf_f \frac{1}{n}\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{X}|\mathbf{Y})\|^{\frac{qp}{p}}\right] \\ &\overset{a)}{\le} \inf_f \frac{1}{n}\left(\mathbb{E}\left[\|\mathbf{X} - f(\mathbf{X}|\mathbf{Y})\|^p\right]\right)^{\frac{q}{p}} \\ &= \left(\frac{1}{n^{\frac{p}{q}}} \inf_f \mathbb{E}\left[\|\mathbf{X} - f(\mathbf{X}|\mathbf{Y})\|^p\right]\right)^{\frac{q}{p}} \\ &= \left(\frac{1}{n^{\frac{p}{q}-1}} \text{mmpe}(\mathbf{X}, \mathsf{snr}, p)\right)^{\frac{q}{p}}, \end{aligned}$$

where the inequality in a) follows from Jensen's inequality and the concavity of $(\cdot)^{\frac{q}{p}}$.

B. Proof of the Bounds in (38b) and (38c)

We now proceed to the proof of the upper bounds in (38b) and (38c). We have

$$\begin{aligned} \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p &\overset{a)}{=} \frac{1}{\sqrt{\mathsf{snr}}}\|\mathbf{Z} - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]\|_p \\ &\overset{b)}{\le} \frac{1}{\sqrt{\mathsf{snr}}}\left(\|\mathbf{Z}\|_p + \|\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\|_p\right), \end{aligned} \tag{99}$$

where the (in)-equalities follow from: a) by using Lemma 1, b) by using the triangle inequality which holds for $p \ge 1$.

Next, the term $\|\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\|_p$ can be further bound as follows:

$$\begin{aligned} n^{\frac{1}{p}}\|\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\|_p &= \mathbb{E}^{\frac{1}{p}}\left[\mathrm{Tr}^{\frac{p}{2}}\left(\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}^T[\mathbf{Z}|\mathbf{Y}]\right)\right] \\ &= \mathbb{E}^{\frac{1}{p}}\left[\left(\sum_{i=1}^n \mathbb{E}^2[Z_i|\mathbf{Y}]\right)^{\frac{p}{2}}\right] \\ &\overset{a)}{\le} \mathbb{E}^{\frac{1}{p}}\left[\left(\sum_{i=1}^n \mathbb{E}[Z_i^2|\mathbf{Y}]\right)^{\frac{p}{2}}\right] \\ &= \mathbb{E}^{\frac{1}{p}}\left[\mathbb{E}^{\frac{p}{2}}\left[\sum_{i=1}^n Z_i^2|\mathbf{Y}\right]\right] \\ &= \mathbb{E}^{\frac{1}{p}}\left[\mathbb{E}^{\frac{p}{2}}\left[\mathrm{Tr}(\mathbf{Z}\mathbf{Z}^T)|\mathbf{Y}\right]\right], \end{aligned} \tag{100}$$

where the inequality in a) follows from using Jensen's inequality. Depending on whether $\frac{p}{2} \le 1$ or $\frac{p}{2} \ge 1$ we bound (100) as follows:

$$\begin{aligned} \text{for } p \ge 2\text{: } &\mathbb{E}^{\frac{1}{p}}\left[\mathbb{E}^{\frac{p}{2}}\left[\mathrm{Tr}(\mathbf{Z}\mathbf{Z}^T)|\mathbf{Y}\right]\right] \\ &\overset{a)}{\le} \mathbb{E}^{\frac{1}{p}}\left[\mathbb{E}\left[\mathrm{Tr}^{\frac{p}{2}}(\mathbf{Z}\mathbf{Z}^T)|\mathbf{Y}\right]\right] \\ &= \mathbb{E}^{\frac{1}{p}}\left[\mathrm{Tr}^{\frac{p}{2}}(\mathbf{Z}\mathbf{Z}^T)\right] \\ &= n^{\frac{1}{p}}\|\mathbf{Z}\|_p, \end{aligned} \tag{101a}$$

$$\begin{aligned} \text{for } 1 \le p < 2\text{: } &\mathbb{E}^{\frac{1}{p}}\left[\mathbb{E}^{\frac{p}{2}}\left[\mathrm{Tr}(\mathbf{Z}\mathbf{Z}^T)|\mathbf{Y}\right]\right] \\ &\overset{b)}{\le} \mathbb{E}^{\frac{1}{2}}\left[\mathbb{E}\left[\mathrm{Tr}(\mathbf{Z}\mathbf{Z}^T)|\mathbf{Y}\right]\right] \\ &= \mathbb{E}^{\frac{1}{2}}\left[\mathrm{Tr}(\mathbf{Z}\mathbf{Z}^T)\right] \\ &= n^{\frac{1}{2}}\|\mathbf{Z}\|_2, \end{aligned} \tag{101b}$$

where the inequalities follow from: a) by using Jensen's inequality on a convex function x^r for $r \ge 1$; and b) by using Jensen's inequality on a concave function x^r for $r \le 1$.

By putting (99), (100) and (101) together we get

$$\text{for } p \ge 2\text{: } \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p \le \frac{2\|\mathbf{Z}\|_p}{\sqrt{\mathsf{snr}}}, \tag{102a}$$

$$\text{for } 1 \le p < 2\text{: } \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p \le \frac{\|\mathbf{Z}\|_p + n^{\frac{1}{2}-\frac{1}{p}}\|\mathbf{Z}\|_2}{\sqrt{\mathsf{snr}}}. \tag{102b}$$

The second term in the minimum of (38b) and (38c) is shown by assuming that $\|\mathbf{X}\|_p$ is finite and by mimicking the steps leading to the bound in (102). We have

$$\text{for } p \ge 1\text{: } \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p \le 2\|\mathbf{X}\|_p, \tag{103a}$$

$$\text{for } 1 \le p < 1\text{: } \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p \le \left(\|\mathbf{X}\|_p + n^{\frac{1}{2}-\frac{1}{p}}\|\mathbf{X}\|_2\right). \tag{103b}$$

Taking the minimum, between (102) and (103) concludes the proof.

C. Proof of the Bound in (38d)

The first part of the bound in (38d) follows by choosing $f(\mathbf{y}) = \frac{\mathbf{y}}{\sqrt{\mathsf{snr}}}$ in the definition of then MMPE, and hence

$$\begin{aligned} \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) &\leq \left\| \mathbf{X} - \frac{\mathbf{Y}}{\sqrt{\mathsf{snr}}} \right\|_p^p \\ &= \frac{1}{\mathsf{snr}^{\frac{p}{2}}} \|\mathbf{Z}\|_p^p. \end{aligned} \tag{104}$$

The bound holds as long as $\|\mathbf{Z}\|_p^p = \mathbb{E}\left[\left(\sum_{i=1}^n Z_i^2\right)^{\frac{p}{2}}\right]$ is finite which is the case for $p \geq 0$.

The second bound follows by choosing $f(\mathbf{y}) = 0$ in the definition of then MMPE, and hence

$$\inf_f \mathbb{E}[\mathrm{Tr}^{\frac{p}{2}}(\mathbf{X} - f(\mathbf{X}|\mathbf{Y}))(\mathbf{X} - f(\mathbf{X}|\mathbf{Y}))^T] \leq \mathbb{E}\left[\mathrm{Tr}^{\frac{p}{2}}\left(\mathbf{X}\mathbf{X}^T\right)\right], \tag{105}$$

which holds for any p as long as $\mathbb{E}\left[\mathrm{Tr}^{\frac{p}{2}}\left(\mathbf{X}\mathbf{X}^T\right)\right]$ exists.

The proof of the upper bound in (38d) is completed by taking the minimum of the bound in (104) and (105). This concludes the proof.

Appendix H
Proof of the Bound in Proposition 11

First we show that if $\|\mathbf{X}\|_p \leq \|\mathbf{Z}\|_p$ then

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \leq \kappa_{p,\mathsf{snr}} \frac{\|\mathbf{Z}\|_p^p}{(1+\mathsf{snr})^{\frac{p}{2}}}. \tag{106}$$

Consider the following sub-optimal estimator $f(\mathbf{Y}) = \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\mathbf{Y}$

$$\begin{aligned} \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) &\leq \left\| \mathbf{X} - \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\mathbf{Y} \right\|_p^p \\ &= \left\| \frac{1}{1+\mathsf{snr}}\mathbf{X} - \frac{\sqrt{\mathsf{snr}}}{1+\mathsf{snr}}\mathbf{Z} \right\|_p^p \\ &= \frac{\left\|\mathbf{X} - \sqrt{\mathsf{snr}}\mathbf{Z}\right\|_p^p}{(1+\mathsf{snr})^p} \\ &\overset{a)}{\leq} \frac{\left(\|\mathbf{X}\|_p + \sqrt{\mathsf{snr}}\|\mathbf{Z}\|_p\right)^p}{(1+\mathsf{snr})^p} \\ &\overset{b)}{\leq} \frac{\left(1+\sqrt{\mathsf{snr}}\right)^p \|\mathbf{Z}\|_p^p}{(1+\mathsf{snr})^p}, \\ &= \kappa_{p,\mathsf{snr}} \frac{\|\mathbf{Z}\|_p^p}{(1+\mathsf{snr})^{\frac{p}{2}}}, \end{aligned} \tag{107}$$

where

$$\kappa_{p,\mathsf{snr}}^{\frac{1}{p}} = \frac{1+\sqrt{\mathsf{snr}}}{\sqrt{1+\mathsf{snr}}}, \tag{108}$$

where the (in)-equalities follow from: a) triangle inequality and scaling property of the norm; and b) by using the assumption that $\|\mathbf{X}\|_p \leq \|\mathbf{Z}\|_p$.

Next, let $\mathbf{X} = \sigma\mathbf{U}$. Then $\|\mathbf{X}\|_p = \|\sigma\mathbf{U}\|_p \leq \sigma\|\mathbf{Z}\|_p$ and therefore $\|\mathbf{U}\|_p \leq \|\mathbf{Z}\|_p$, so by using the bound in (106) we have that

$$\begin{aligned} \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) &= \mathrm{mmpe}(\sigma\mathbf{U}, \mathsf{snr}, p) \\ &\overset{a)}{=} \sigma^p \mathrm{mmpe}(\mathbf{U}, \sigma^2\mathsf{snr}, p) \\ &\overset{b)}{\leq} \kappa_{p,\sigma^2\mathsf{snr}} \sigma^p \frac{\|\mathbf{Z}\|_p^p}{(1+\mathsf{snr}\sigma^2)^{\frac{p}{2}}}, \end{aligned}$$

where the (in)-equalities follow from: a) by using the scaling property of the MMPE in Proposition 6; and b) by using the bound in (106).

Observe that the bound in (106) is achieved asymptotically by using $\mathbf{X}_G \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ since by Proposition 3 and the scaling property in Proposition 6 we have that

$$\mathrm{mmpe}(\mathbf{X}_G, \mathsf{snr}, p) = \frac{\sigma^p \|\mathbf{Z}\|_p^p}{(1+\mathsf{snr}\sigma^2)^{\frac{p}{2}}}.$$

This concludes the proof.

Appendix I
Proof of Proposition 15

We seek to give an upper bound on $P_e^{(n)}(\mathsf{snr})$ in (45) for the MAP decoding rule given by

$$\begin{aligned} \hat{x} &= \arg\max_m p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}_m|\mathbf{y}) \\ &= \arg\max_m p_m \mathrm{e}^{-\frac{\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{x}_m\|^2}{2}} \\ &= \arg\max_m p_m \mathrm{e}^{-\frac{\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{x}_m\|^2}{2}} \\ &= \arg\min_m \left(\frac{\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{x}_m\|^2}{2} + \log\frac{1}{p_m} \right). \end{aligned}$$

To that end, let us denote the following events:

$$\mathcal{E}_i(\mathbf{y}) = \left\{ x_k = \arg\min_m \left(\frac{\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{x}_m\|^2}{2} + \log\frac{1}{p_i} \right), \text{ for some } k \neq i \right\},$$

$$\mathcal{E}_{i,j}(\mathbf{y}) = \left\{ \frac{\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{x}_i\|^2}{2} + \log\frac{p_j}{p_i} \geq \frac{\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{x}_j\|^2}{2} \right\}.$$

We have

$$\begin{aligned} P_e^{(n)}(\mathsf{snr}) &= \mathbb{P}[\hat{\mathbf{X}} \neq \mathbf{X}] \\ &= \sum_{i=1}^N p_i \mathbb{P}[\hat{\mathbf{X}} \neq \mathbf{X}|\mathbf{X} = \mathbf{x}_i] \\ &= \sum_{i=1}^N p_i \mathbb{P}[\mathcal{E}_i(\mathbf{Y})|\mathbf{X} = \mathbf{x}_i] \\ &\leq \sum_{i=1}^N \sum_{j \neq i} p_i \mathbb{P}[\mathcal{E}_{i,j}(\mathbf{Y})|\mathbf{X} = \mathbf{x}_i]. \end{aligned} \tag{109}$$

Next we analyze $\mathbb{P}[\mathcal{E}_{i,j}(\mathbf{Y})|\mathbf{X}=\mathbf{x}_i]$

$$\begin{aligned}
&\mathbb{P}[\mathcal{E}_{i,j}(\mathbf{Y})|\mathbf{X}=\mathbf{x}_i]\\
&= \mathbb{P}\left[\frac{\|\mathbf{Z}\|^2}{2}+\log\frac{p_j}{p_i} \ge \frac{\|\mathbf{Z}+\sqrt{\mathsf{snr}}(\mathbf{x}_i-\mathbf{x}_j)\|^2}{2}\right]\\
&\overset{a)}{=} \mathbb{P}\left[\log\left(\frac{p_j}{p_i}\right)-\frac{\mathsf{snr}\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2} \ge \sqrt{\mathsf{snr}}\mathbf{Z}^T(\mathbf{x}_i-\mathbf{x}_j)\right]\\
&\overset{b)}{=} \mathbb{P}\left[\log\left(\frac{p_j}{p_i}\right)-\frac{\mathsf{snr}\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2} \ge \sqrt{\mathsf{snr}}\|\mathbf{x}_i-\mathbf{x}_j\|Z\right]\\
&\overset{c)}{=} Q\left(\frac{\sqrt{\mathsf{snr}}d_{ij}}{2}-\log\left(\frac{p_j}{p_i}\right)\frac{1}{\sqrt{\mathsf{snr}}d_{ij}}\right), \quad (110)
\end{aligned}$$

where the equalities follow from: a) expanding $\|\mathbf{Z}+2\sqrt{\mathsf{snr}}(\mathbf{x}_i-\mathbf{x}_j)\|^2$; b) using the fact that $\mathbf{Z}^T(\mathbf{x}_i-\mathbf{x}_j)$ has the same distribution as $\|\mathbf{x}_i-\mathbf{x}_j\|Z$; and c) using the definition $d_{ij}=\|\mathbf{x}_i-\mathbf{x}_j\|$.

Combining (109) and (110) concludes the proof.

Appendix J
Proof of Theorem 1

Let $\mathbf{W}_\mathbf{v}=\mathbf{U}_\mathbf{v}-g(\mathbf{v})$ where $g(\cdot)$ is a deterministic function and $\mathbf{U}_\mathbf{v}\sim p_{\mathbf{U}|\mathbf{V}}(\cdot|\mathbf{v})$. By [40, Th. 3] we have

$$\frac{n^{\frac{1}{p}}\|\mathbf{W}_\mathbf{v}\|_p}{e^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{W}_\mathbf{v})}} \ge \frac{1}{k_{n,p}},\ k_{n,p}:=\frac{\sqrt{pi}\left(\frac{p}{n}\right)^{\frac{1}{p}}\mathrm{e}^{\frac{1}{p}}\Gamma^{\frac{1}{n}}\left(\frac{n}{p}+1\right)}{\Gamma^{\frac{1}{n}}\left(\frac{n}{2}+1\right)}, \quad (111)$$

where $h_{\mathrm{e}}(\cdot)$ is the differential entropy measured in nats. Moreover, observe that $h_{\mathrm{e}}(\mathbf{W}_\mathbf{v})=h_{\mathrm{e}}(\mathbf{U}_\mathbf{v}-g(\mathbf{v}))=h_{\mathrm{e}}(\mathbf{U}_\mathbf{v})$ due to the translation invariance of the differential entropy. Therefore, by rearranging (111) and by using the translation invariance of the differential entropy, we get

$$\frac{1}{n}h_{\mathrm{e}}(\mathbf{U}_\mathbf{v})\log(\mathrm{e}) \le \log\left(k_{n,p}\cdot n^{\frac{1}{p}}\|\mathbf{W}_\mathbf{v}\|_p\right), \quad (112)$$

where from (5) we have $n^{\frac{1}{p}}\|\mathbf{W}_\mathbf{v}\|_p = \mathbb{E}^{\frac{1}{p}}[\|\mathbf{U}-g(\mathbf{V})\|^p|\mathbf{V}=\mathbf{v}]$. By taking the expectation on both sides of (112) with respect to $p_\mathbf{V}(\mathbf{v})$ we arrive at

$$\begin{aligned}
&n^{-1}h_{\mathrm{e}}(\mathbf{U}|\mathbf{V})\log(\mathrm{e}) = n^{-1}h(\mathbf{U}|\mathbf{V})\\
&\le \frac{1}{p}\mathbb{E}\left[\log\left(k_{n,p}^p\cdot n\cdot\frac{1}{n}\cdot\mathbb{E}\left[\|\mathbf{U}-g(\mathbf{V})\|^p|\mathbf{V}\right]\right)\right]\\
&\overset{a)}{\le} \frac{1}{p}\log\left(k_{n,p}^p\cdot n\cdot\frac{1}{n}\cdot\mathbb{E}\left[\mathbb{E}\left[\|\mathbf{U}-g(\mathbf{V})\|^p|\mathbf{V}\right]\right]\right)\\
&= \frac{1}{p}\log\left(k_{n,p}^p\cdot n\cdot\frac{1}{n}\cdot\mathbb{E}\left[\|\mathbf{U}-g(\mathbf{V})\|^p\right]\right)\\
&= \log\left(k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{U}-g(\mathbf{V})\|_p\right),
\end{aligned}$$

where the inequality in a) follows from Jensen's inequality. Finally, since this bound holds for any deterministic function $g(\cdot)$, to tighten this bound, and due to the monotonicity of the log function, we may pick $g(\cdot)$ to be the optimal p-th estimator of $\mathbf{U}$. This concludes the proof.

Appendix K
Proof of Theorem 2

Let $(\mathbf{U},\mathbf{X}_D,\mathbf{Z})$ be mutually independent. By the data processing inequality and the assumption in (67a) we have

$$\begin{aligned}
I(\mathbf{X}_D;\mathbf{Y}) &\ge I(\mathbf{X}_D+\mathbf{U};\mathbf{Y})\\
&= h(\mathbf{X}_D+\mathbf{U})-h(\mathbf{X}_D+\mathbf{U}|\mathbf{Y})\\
&= H(\mathbf{X}_D)+h(\mathbf{U})-h(\mathbf{X}_D+\mathbf{U}|\mathbf{Y}). \quad (113)
\end{aligned}$$

Next, by using Theorem 1, we have that the last term of (113) can be bounded as

$$n^{-1}h(\mathbf{X}_D+\mathbf{U}|\mathbf{Y}) \le \log\left(k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{X}_D+\mathbf{U}-g(\mathbf{Y})\|_p\right). \quad (114)$$

Next, by combining (113) and (114) and taking $g(\mathbf{Y})=f_p(\mathbf{X}|\mathbf{Y})$ we have that

$$\begin{aligned}
I(\mathbf{X}_D;\mathbf{Y}) &\ge H(\mathbf{X}_D)-\mathsf{gap}_p, \quad (115)\\
n^{-1}\mathsf{gap}_p &\le \inf_{\mathbf{U}\in\mathcal{K}}\left(L_{1,p}(\mathbf{U},\mathbf{X}_D)+L_{2,p}(\mathbf{U})\right),\\
G_{1,p}(\mathbf{U},\mathbf{X}_D) &= \log\left(\frac{\|\mathbf{U}+\mathbf{X}_D-f_p(\mathbf{X}|\mathbf{Y})\|_p}{\|\mathbf{U}\|_p}\right)\\
&\overset{\text{for } p\ge 1}{\le} \log\left(1+\frac{\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}_D,\mathsf{snr},p)}{\|\mathbf{U}\|_p}\right),\\
G_{2,p}(\mathbf{U}) &= \log\left(\frac{k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{U})}}\right), \quad (116)
\end{aligned}$$

where inequality in (116) follows by the triangle inequality which holds for $p\ge 1$.

Finally, the proof concludes by taking $g(\mathbf{Y})=f_p(\mathbf{X}|\mathbf{Y})$.

Appendix L
Proof of Theorem 3

To show that $\lim_{n\to\infty}G_{2,p}(\mathbf{U})=0$ we show that

$$\lim_{n\to\infty}\frac{k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{U})}}=1.$$

First of all observe that using (111) in Appendix J

$$1\le\frac{k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{U})}}.$$

Next, we show an upper bound. Note that if $\mathbf{U}$ is uniform over a ball $B_0(r)$ of radius $r=d_{\min}(\mathbf{X}_D)/2$ then

$$h(\mathbf{U})=\log\left(\mathrm{Vol}(B_0(r))\right), \quad (117)$$

where

$$\mathrm{Vol}(B_0(r))=\frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}r^n. \quad (118)$$

Moreover, the norm $\mathbf{U}$ can be upper bounded by

$$\begin{aligned}
\|\mathbf{U}\|_p^p &= \frac{1}{n}\frac{1}{\mathrm{Vol}(B_0(r))}\int_{B_0(r)}\left(\sum_{i=1}^n u_i^2\right)^{\frac{p}{2}}du_1du_2\cdots du_n\\
&\le \frac{1}{n}\frac{1}{\mathrm{Vol}(B_0(r))}\int_{B_0(r)}\left(r^2\right)^{\frac{p}{2}}du_1du_2\cdots du_n=\frac{r^p}{n}. \quad (119)
\end{aligned}$$

Therefore, by using (119) and (118)

$$\begin{aligned}\frac{k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{U})}} &\le \frac{k_{n,p}\cdot\Gamma^{\frac{1}{n}}\left(\frac{n}{2}+1\right)}{\sqrt{\pi}}\\ &= \frac{\sqrt{\pi}\left(\frac{p}{n}\right)^{\frac{1}{p}}\mathrm{e}^{\frac{1}{p}}\Gamma^{\frac{1}{n}}\left(\frac{n}{p}+1\right)\Gamma^{\frac{1}{n}}\left(\frac{n}{2}+1\right)}{\Gamma^{\frac{1}{n}}\left(\frac{n}{2}+1\right)\sqrt{\pi}}\\ &= (p\mathrm{e})^{\frac{1}{p}}\left(\frac{1}{n}\right)^{\frac{1}{p}}\Gamma^{\frac{1}{n}}\left(\frac{n}{p}+1\right).\end{aligned}$$

Next by using the Stirling's approximation $\Gamma(x+1) = \sqrt{2\pi x}\left(\frac{x}{\mathrm{e}}\right)^x + o(x)$ we have that

$$\left(\frac{1}{n}\right)^{\frac{1}{p}}\Gamma^{\frac{1}{n}}\left(\frac{n}{p}+1\right) \le \left(\frac{2\pi n}{p}\right)^{\frac{1}{n}}\left(\frac{1}{p\mathrm{e}}\right)^{\frac{1}{p}} + o\left(n\right),$$

and therefore

$$\frac{k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{U})}} \le \left(\frac{2\pi n}{p}\right)^{\frac{1}{n}} + o\left(n\right) \overset{\text{as } n\to\infty}{\to} 1.$$

This shows that $\lim_{n\to\infty} G_{2,p}(\mathbf{U}) = 0$. Next, we show that $\lim_{n\to\infty} G_{1,p}(\mathbf{X}_D,\mathbf{U}) = 0$ by showing that $\lim_{n\to\infty}\frac{\mathrm{mmpe}(\mathbf{X}_D,\mathsf{snr},p)}{\|\mathbf{U}\|_p} = 0$. First, observe that by using the bound in (47a)

$$\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}_D,\mathsf{snr},p) \le \frac{d_{\max} Q^{\frac{1}{p}}\left(\frac{n}{2};\frac{\mathsf{snr}d_{\min}^2}{8}\right)}{n^{\frac{1}{p}}},$$

and by using (8) we have that

$$\begin{aligned}&\frac{\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}_D,\mathsf{snr},p)}{\|\mathbf{U}\|_p}\\ &\le \frac{\frac{d_{\max}(\mathbf{X}_D) Q^{\frac{1}{p}}\left(\frac{n}{2};\frac{\mathsf{snr}d_{\min}^2(\mathbf{X}_D)}{8}\right)}{n^{\frac{1}{p}}}}{\frac{d_{\min}(\mathbf{X}_D)}{2(p+n)^{\frac{1}{p}}}}\\ &= 2\frac{d_{\max}(\mathbf{X}_D)}{d_{\min}(\mathbf{X}_D)}\sqrt[p]{\frac{(p+n)\bar{Q}\left(\frac{n}{2};\frac{\mathsf{snr}d_{\min}^2(\mathbf{X}_D)}{8}\right)}{n}}. \quad (120)\end{aligned}$$

This concludes the proof.

Appendix M
On Finding the Optimal r in the Proof of Theorem 4

We must solve the following optimization problem:

$$\begin{aligned}\min_{r>\frac{2}{\gamma}} g(r) &= M^{\frac{\gamma r-2}{r-2}}\frac{G^{\frac{r(1-\gamma)}{r-2}}}{N^{\frac{2(1-\gamma)}{r-2}}}\Gamma^{\frac{2(1-\gamma)}{r-1}}(n/2+r/2),\\ M &= \mathrm{mmse}(\mathbf{X},\mathsf{snr}_0),\\ G &= \frac{8}{\mathsf{snr}_0},\\ N &= n\Gamma\left(\frac{n}{2}\right) = 2\Gamma\left(\frac{n}{2}+1\right).\end{aligned}$$

Instead of optimizing $g(r)$ we will focus on optimizing $h(r) = \ln(g(r))$ where

$$\begin{aligned}h(r) &= \frac{\gamma r-2}{r-2}\ln(M) + \frac{r(1-\gamma)}{r-2}\ln(G) - \frac{2(1-\gamma)}{(r-2)}\ln(N)\\ &\quad + \frac{2(1-\gamma)}{r-2}\ln\left(\Gamma\left(\frac{n+r}{2}\right)\right). \quad (121)\end{aligned}$$

Unfortunately, a closed form solution for the optimum of (121) is difficult to find and instead we look for an approximate solution. This is done by using Stirling's formula $\Gamma(x+1)\approx\sqrt{2\pi x}\left(\frac{x}{\mathrm{e}}\right)^x$. We have

$$\begin{aligned}\Gamma\left(\frac{n+r}{2}\right) &= \Gamma\left(\frac{n+r}{2}-1+1\right)\\ &\approx \sqrt{2\pi\left(\frac{r+n-2}{2}\right)}\left(\frac{\frac{r+n-2}{2}}{\mathrm{e}}\right)^{\frac{r}{2}+\frac{n-2}{2}}. \quad (122)\end{aligned}$$

Now, we seek to optimize the following expression:

$$g(r) \approx \frac{M^{\frac{\gamma r-2}{r-2}}N^{\frac{2(\gamma-1)}{(r-2)}}\left(\sqrt{2\pi\left(\frac{r+n-2}{2}\right)}\left(\frac{\frac{r+n-2}{2}}{\mathrm{e}}\right)^{\frac{r+n-2}{2}}\right)^{\frac{2(1-\gamma)}{r-2}}}{G^{\frac{r(\gamma-1)}{r-2}}},$$

that is

$$\begin{aligned}h(r) &\approx \frac{\gamma r-2}{r-2}\ln(M) + \frac{r(1-\gamma)}{r-2}\ln(G) - \frac{2(1-\gamma)}{r-2}\ln(N)\\ &\quad + \frac{1-\gamma}{r-2}\ln\left(2\pi\left(\frac{r}{2}+\frac{n-2}{2}\right)\right) - \frac{2(1-\gamma)(\frac{r}{2}+\frac{n-2}{2})}{r-2}\\ &\quad + \frac{2(1-\gamma)(\frac{r}{2}+\frac{n-2}{2})}{r-2}\ln\left(\frac{r}{2}+\frac{n-2}{2}\right). \quad (123)\end{aligned}$$

By taking the derivative of (123) with respect to r we get

$$\begin{aligned}h'(r) &= \frac{1}{2}\frac{1-\gamma}{(\frac{r}{2}-1)^2}f(r)\\ f(r) &= \ln(M) - \ln(\sqrt{2\pi}G) + \log(N) + \frac{r-2}{n-2+r}\\ &\quad - \frac{n+1}{2}\ln\left(\frac{n-2}{2}+\frac{r}{2}\right) + \frac{r}{2} + \frac{n-2}{2}\\ &\approx \ln(M) - \ln(\sqrt{2\pi}G) + \log(N)\\ &\quad - \frac{n+1}{2}\ln\left(\frac{n-2}{2}+1\right) + \frac{r}{2} + \frac{n-2}{2} \quad (124)\end{aligned}$$

where in the last step we used the approximation $\frac{n+1}{2}\ln\left(\frac{n-2}{2}+\frac{r}{2}\right)\approx\frac{n+1}{2}\ln\left(\frac{n-2}{2}+1\right)$ and $\frac{r-2}{n-2+r}\approx 0$ which is reasonable as n becomes large.

Solving $f(r) = 0$ in (124) we get that the approximate solution is

$$\begin{aligned}\frac{r}{2} &= \ln\left(\frac{\sqrt{2\pi}\, G\left(\frac{n}{2}\right)^{\frac{n+1}{2}}}{MN\mathrm{e}^{\frac{n-2}{2}}}\right)\\ &= \ln\left(\frac{8\sqrt{2\pi}\left(\frac{n}{2}\right)^{\frac{n+1}{2}}}{\mathsf{snr}_0\mathrm{mmse}(X,\mathsf{snr}_0)2\Gamma\left(\frac{n}{2}+1\right)\mathrm{e}^{\frac{n-2}{2}}}\right)\\ &\approx \ln\left(\frac{8\sqrt{2\pi}\left(\frac{n}{2}\right)^{\frac{n+1}{2}}}{\mathsf{snr}_0\mathrm{mmse}(X,\mathsf{snr}_0)2\sqrt{2\pi\frac{n}{2}}\left(\frac{n}{2\mathrm{e}}\right)^{\frac{n}{2}}\mathrm{e}^{\frac{n-2}{2}}}\right)\\ &= \ln\left(\frac{4\mathrm{e}}{\mathsf{snr}_0\mathrm{mmse}(X,\mathsf{snr}_0)}\right),\end{aligned}$$

where in the last approximation we have used Stirling's formula.

Since, we have a constraint that $r > \frac{2}{\gamma}$ we set r to be

$$r \approx \begin{cases} 2\ln\left(\frac{4\mathrm{e}}{\mathsf{snr}_0\mathrm{mmse}(X,\mathsf{snr}_0)}\right), & \frac{2}{\gamma} \le \ln\left(\frac{4\mathrm{e}}{\mathsf{snr}_0\mathrm{mmse}(X,\mathsf{snr}_0)}\right)\\ \frac{2}{\gamma}, & \frac{2}{\gamma} > \ln\left(\frac{4\mathrm{e}}{\mathsf{snr}_0\mathrm{mmse}(X,\mathsf{snr}_0)}\right). \end{cases}$$

This concludes the proof.

Appendix N
Proof of Proposition 21

First observe that

$$\inf_f \mathbb{E}[\|\mathbf{X} - f(\mathbf{Y})\|^2|\mathbf{Y}=\mathbf{y}] = \mathbb{E}\left[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|^2|\mathbf{Y}=\mathbf{y}\right]. \tag{125}$$

We will need the following bounds on trace of $\mathbf{A} \succeq 0$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$

$$\frac{1}{n}\mathrm{Tr}(\mathbf{A})^2 \le \mathrm{Tr}(\mathbf{A}^2) \le n\mathrm{Tr}(\mathbf{A})^2. \tag{126}$$

For the upper bound we have that

$$\begin{aligned}&\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right)\\ &= \mathbb{E}\left[\mathrm{Tr}\left(\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right)\right]\\ &\overset{a)}{\le} \mathbb{E}\left[n\mathrm{Tr}^2\left(\mathbf{Cov}(\mathbf{X}|\mathbf{Y})\right)\right]\\ &= \mathbb{E}\left[n\mathrm{Tr}^2\left(\mathbb{E}\left[(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])^T|\mathbf{Y}\right]\right)\right]\\ &= \mathbb{E}\left[n\mathbb{E}^2\left[\mathrm{Tr}(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])^T|\mathbf{Y}\right]\right]\\ &= \mathbb{E}\left[n\mathbb{E}^2\left[\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|^2|\mathbf{Y}\right]\right]\\ &\overset{b)}{=} \mathbb{E}\left[n\left(\inf_f \mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^2|\mathbf{Y}\right]\right)^2\right]\\ &= \mathbb{E}\left[n\inf_f \mathbb{E}^2\left[\|\mathbf{X}-f(\mathbf{Y})\|^2|\mathbf{Y}\right]\right]\\ &\overset{c)}{\le} \mathbb{E}\left[n\inf_f \mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^4|\mathbf{Y}\right]\right]\\ &\overset{d)}{\le} n\inf_f \mathbb{E}\left[\mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^4|\mathbf{Y}\right]\right]\\ &\overset{e)}{=} n\inf_f \mathbb{E}\left[\|\mathbf{X}-f(\mathbf{Y})\|^4\right]\\ &= n^2\mathrm{mmpe}(\mathbf{X},\mathsf{snr},4),\end{aligned}$$

where the (in)-equalities follow from: a) since $\mathbf{Cov}(\mathbf{X}|\mathbf{Y}) \succeq 0$ and using the inequality in (126); and b) by using (125); c) Jensen's inequality; d) by using $\mathbb{E}[X_1] \le \mathbb{E}[X_2]$ if $X_1 \le X_2$; and e) law of total expectation.

For the lower bound

$$\begin{aligned}\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right) &= \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right)\right]\\ &\overset{a)}{\ge} \frac{1}{n}\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}^2\left(\mathbf{Cov}(\mathbf{X}|\mathbf{Y})\right)\right]\\ &\overset{b)}{\ge} \frac{1}{n^2}\mathbb{E}^2\left[\mathrm{Tr}\left(\mathbf{Cov}(\mathbf{X}|\mathbf{Y})\right)\right]\\ &= \mathrm{mmse}^2(\mathbf{X},\mathsf{snr}),\end{aligned}$$

where the inequalities follow from: a) since $\mathbf{Cov}(\mathbf{X}|\mathbf{Y}) \succeq 0$ and by using the inequality in (126); and b) Jensen's inequality.

Acknowledgment

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

References

[1] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, H. V. Poor, and S. Shamai (Shitz), "On the applications of the minimum mean p-th error (MMPE) to information theoretic quantities," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2016, pp. 66–70.

[2] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, H. V. Poor, and S. Shamai (Shitz), "On the minimum mean p-th error in Gaussian noise channels and its applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 1646–1650.

[3] 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.

[4] S. Shamai (Shitz), "From constrained signaling to network interference alignment via an information-estimation perspective," *IEEE Inf. Theory Soc. Newslett.*, vol. 62, no. 7, pp. 6–24, Sep. 2012.

[5] E. Kreyszig, *Introductory Functional Analysis With Applications*, vol. 81. Hoboken, NJ, USA: Wiley, 1989.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[7] S. Verdú and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2165–2166, May 2006.

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

[9] 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.

[10] 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.

[11] D. Guo, S. Shamai (Shitz), and S. Verdú, *The Interplay Between Information and Estimation Measures*. Breda, The Netherlands: Now Publishers, 2013.

[12] R. Bustin and S. Shamai (Shitz), "MMSE of 'bad' codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 733–743, Feb. 2013.

[13] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, H. V. Poor, and S. Shamai (Shitz), "On communication through a Gaussian channel with an MMSE disturbance constraint," *IEEE Trans. Inf. Theory*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/8023840/

[14] N. Merhav, D. Guo, and S. Shamai (Shitz), "Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1400–1416, Mar. 2010.

[15] R. Bustin, R. F. Schaefer, H. V. Poor, and S. Shamai (Shitz), "On the SNR-evolution of the MMSE function of codes for the Gaussian broadcast and wiretap channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2070–2091, Apr. 2016.

[16] 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.

[17] Y. Wu and S. Verdú, "MMSE dimension," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, Aug. 2011.

[18] D. Guo, "Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbation," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2009, pp. 814–818.

[19] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai (Shitz), "A view of information-estimation relations in Gaussian networks," *Entropy*, vol. 19, no. 8, p. 409, 2017.

[20] S. Sherman, "Non-mean-square error criteria," *IRE Trans. Inf. Theory*, vol. 4, no. 3, pp. 125–126, Sep. 1958.

[21] J. Brown, "Asymmetric non-mean-square error criteria," *IRE Trans. Autom. Control*, vol. 7, no. 1, pp. 64–66, Jan. 1962.

[22] V. Pugachev, "A method for determining optimum systems using general Bayes criterion," *IRE Trans. Circuit Theory*, vol. 7, no. 4, pp. 491–505, Dec. 1960.

[23] J. Tan, D. Baron, and L. Dai, "Wiener filters in Gaussian mixture signal estimation with ℓ_∞-norm error," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6626–6635, Oct. 2014.

[24] E. B. Hall and G. L. Wise, "Simultaneous optimal estimation over a family of fidelity criteria," in *Proc. Conf. Inf. Sci. Syst.*, 1987, pp. 25–27.

[25] E. B. Hall and G. L. Wise, "On optimal estimation with respect to a large family of cost functions," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 691–693, May 1991.

[26] E. Akyol, K. B. Viswanatha, and K. Rose, "On conditions for linearity of optimal estimation," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3497–3508, Jun. 2012.

[27] N. Weinberger and N. Merhav, "Lower bounds on parameter modulation–estimation under bandwidth constraints," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3854–3874, Jun. 2017.

[28] N. Merhav, "Exponential error bounds on parameter modulation–estimation for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 832–841, Feb. 2014.

[29] M. Saerens, "Building cost functions minimizing to some summary statistics," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1263–1271, Nov. 2000.

[30] G. Livadiotis, "Expectation values and variance based on ℓ^p-norms," *Entropy*, vol. 14, no. 12, pp. 2375–2396, 2012.

[31] A. Ben-Tal, "On generalized means and generalized convex functions," *J. Optim. Theory Appl.*, vol. 21, no. 1, pp. 1–13, 1977.

[32] M. V. Burnashev, "A new lower bound for the α-mean error of parameter transmission over the white Gaussian channel," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 1, pp. 23–34, Jan. 1984.

[33] M. V. Burnashev, "On minimum attainable mean-square error in transmission of a parameter over a channel with white Gaussian noise," *Problemy Peredachi Inf.*, vol. 21, no. 4, pp. 3–16, 1985.

[34] R. Gonin, *Nonlinear Lp-Norm Estimation*, vol. 100. Boca Raton, FL, USA: CRC Press, 1989,

[35] E. L. Lehmann and G. Casella, *Theory Point Estimation*. New York, NY, USA: Springer, 2006.

[36] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[37] L. H. Ozarow and A. D. Wyner, "On the capacity of the Gaussian channel with a finite number of input levels," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1426–1428, Nov. 1990.

[38] E. Lutwak, S. Lv, D. Yang, and G. Zhang, "Affine moments of a random vector," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5592–5599, Sep. 2013.

[39] G. L. Wise, "A note on a common misconception in estimation," *Syst. Control Lett.*, vol. 5, no. 5, pp. 355–356, 1985.

[40] E. Lutwak, D. Yang, and G. Zhang, "Moment-entropy inequalities for a random vector," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1603–1607, Apr. 2007.

[41] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[42] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*. Hoboken, NJ, USA: Wiley, 2013.

[43] R. Webster, *Convexity*. London, U.K.: Oxford Univ. Press, 1994.

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

[45] A. Dytso, D. Tuninetti, and N. Devroye, "Interference as noise: Friend or foe?" *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3561–3596, Jun. 2016.

[46] A. Dytso, D. Tuninetti, and N. Devroye, "On the two-user interference channel with lack of knowledge of the interference codebook at one receiver," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1257–1276, Mar. 2015.

[47] Y. Dong, F. Farnia, and A. Özgür, "Near optimal energy control and approximate capacity of energy harvesting communication," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 540–557, Mar. 2015.

[48] D. Shaviv, P.-M. Nguyen, and A. Özgür, "Capacity of the energy-harvesting channel with a finite battery," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6436–6458, Nov. 2016.

[49] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2515–2534, Jun. 2008.

[50] G. D. Forney, Jr., "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. 41th Annu. Allerton Conf. Commun., Control Comput.*, 2003, vol. 41. no. 1, pp. 430–439.

[51] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "A generalized Ozarow-Wyner capacity bound with application," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 1058–1062.

[52] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai (Shitz), "On additive channels with generalized Gaussian noise," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 426–430.

[53] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor, "Upper and lower bounds on the capacity of amplitude-constrained MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017.

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

[55] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[56] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Proc. IEEE 48th Annu. Allerton Conf. Commun., Control Comput.*, Sep./Oct. 2010, pp. 620–628.

[57] S. Shamai (Shitz), L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, Nov. 1991.

[58] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, "Information bottleneck for Gaussian variables," *J. Mach. Learn. Res.*, vol. 6, pp. 165–188, Jan. 2005.

[59] D. G. Luenberger, *Optimization by Vector Space Methods*. Hoboken, NJ, USA: Wiley, 1997.

[60] D. V. Widder, *Advanced Calculus*. Chelmsford, MA, USA: Courier Corporation, 1989.

Alex Dytso is currently a Postdoctoral Researcher in the Department of Electrical Engineering at Princeton University. In 2016, he received a Ph.D. degree from the Department of Electrical and Computer Engineering at the University of Illinois, Chicago. He received his B.S. degree in 2011 from the University of Illinois, Chicago, where he also received the International Engineering Consortium's William L. Everitt Student Award of Excellence for outstanding seniors. His current research interest are in the areas of multi-user information theory and estimation theory, and their applications in wireless networks.

Ronit Bustin (S'09–M'14) received a B.Sc. degree in computer science and electrical engineering and a M.Sc. degree in electrical engineering in 2004 and 2006, respectively, from Tel-Aviv university, Israel. She received a Ph.D. in electrical engineering in 2013 from the Technion - Israel Institute of Technology, Haifa. From 2013 to 2015 she was a postdoctoral research associate in the Department of Electrical Engineering at Princeton University. During 2015-2016 she was a post-doctoral research associate in the Department of Electrical Engineering at Tel Aviv University. During 2016-2017 she was a post-doctoral research associate in the Department of Electrical Engineering at the Technion. She currently holds a research position in General Motors - Advanced Technical Center in Israel. Her research interests include multi-user information theory, secrecy constraints, Gaussian MIMO channels, estimation theory, channel coding, interactive communication and communication complexity. Ronit Bustin received the Irwin and Joan Jacobs scholarship for excellence in graduate studies and research, in January 2010. She is a recipient of the Adams fellowship from the Israel Academy of Sciences and Humanities, April 2010, and an Andrew and Erna Finci Viterbi graduate fellow in the faculty of electrical engineering at the Technion for the fall semester 2010-2011. Ronit received the Rothschild fellowship in 2013 and the women postdoctoral scholarship of Israel's Council for Higher Education (VATAT) for her postdoctoral studies at Princeton.

Daniela Tuninetti is currently a Professor within the Department of Electrical and Computer Engineering at the University of Illinois at Chicago (UIC), which she joined in 2005. Dr. Tuninetti got her Ph.D. in Electrical Engineering in 2002 from ENST/Telecom ParisTech (Paris, France, with work done at the Eurecom Institute in Sophia Antipolis, France), and she was a postdoctoral research associate at the School of Communication and Computer Science at the Swiss Federal Institute of Technology in Lausanne (EPFL, Lausanne, Switzerland) from 2002 to 2004.

Dr. Tuninetti is a recipient of a best paper award at the European Wireless Conference in 2002, of an NSF CAREER award in 2007, and named UIC University Scholar in 2015.

Dr. Tuninetti was the editor-in-chief of the IEEE Information Theory Society Newsletter from 2006 to 2008, an editor for IEEE COMMUNICATION LETTERS from 2006 to 2009, and for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2014; she is currently an associate editor for IEEE TRANSACTIONS ON INFORMATION THEORY.

Dr. Tuninetti's research interests are in the ultimate performance limits of wireless interference networks (with special emphasis on cognition and user cooperation), coexistence between radar and communication systems, multi-relay networks, content-type coding, and caching systems.

Natasha Devroye is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Illinois at Chicago (UIC), which she joined in January 2009. From July 2007 until July 2008 she was a Lecturer at Harvard University. Dr. Devroye obtained her Ph.D in Engineering Sciences from the School of Engineering and Applied Sciences at Harvard University in 2007, and a Honors B. Eng in Electrical Engineering from McGill University in 2001. Dr. Devroye was a recipient of an NSF CAREER award in 2011 and was named UIC's Researcher of the Year in the "Rising Star" category in 2012. She has been an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and is currently an Associate Editor for the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING and the IEEE TRANSACTIONS ON INFORMATION THEORY. Her research focuses on multi-user information theory and applications to cognitive and software-defined radio, radar, relay, zero-error and two-way communication networks.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 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 the Michael Henry Strater University Professor of Electrical Engineering. During 2006 to 2016, he served as Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks, energy systems and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Chinese Academy of Sciences and the Royal Society. In 1990, he served as President of the IEEE INFORMATION THEORY SOCIETY, in 2004-07 as the Editor-in-Chief of these TRANSACTIONS, and in 2009 as General Co-chair of the IEEE International Symposium on Information Theory, held in Seoul, South Korea. He received a Guggenheim Fellowship in 2002 and the IEEE Education Medal in 2005. Recent recognition of his work includes the 2016 John Fritz Medal, the 2017 IEEE Alexander Graham Bell Medal, Honorary Professorships from Peking University and Tsinghua University, both conferred in 2016, and a D.Sc. *honoris causa* from Syracuse University awarded in 2017.

Shlomo Shamai (Shitz) (S'80–M'82–SM'88–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975-1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow, an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities and a foreign member of the US National Academy of Engineering. He is the 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.

He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 and 2015 European Commission FP7, Network of Excellence in Wireless COMmunications (NEWCOM++, NEWCOM#) Best Paper Awards, the 2010 Thomson Reuters Award for International Excellence in Scientific Research, the 2014 EURASIP Best Paper Award (for the *EURASIP Journal on Wireless Communications and Networking*), and the 2015 IEEE Communications Society Best Tutorial Paper Award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He has also served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY and on the IEEE Information Theory Society Nominations and Appointments Committee.