On the Minimum Mean p-th Error in Gaussian Noise Channels and Its Applications

Alex Dytso, Ronit Bustin, Daniela Tuninetti, Natasha Devroye, H. Vincent Poor and Shlomo Shamai (Shitz)
Corresponding Author: Alex Dytso, University of Illinois at Chicago, USA, odytso2@uic.edu

***Abstract*—The problem of estimating an arbitrary random variable from its observation corrupted by additive white Gaussian noise, where the cost function is taken to be the minimum mean p-th error (MMPE), is considered. The classical minimum mean square error (MMSE) is a special case of the MMPE. Several bounds and properties of the MMPE are derived and discussed. As applications of the new MMPE bounds, this paper presents: (a) a new upper bound for the MMSE that complements the 'single-crossing point property' for all SNR values *below* a certain value at which the MMSE is known, (b) an improved characterization of the phase-transition phenomenon which manifests, in the limit as the length of the capacity achieving code goes to infinity, as a discontinuity of the MMSE, and (c) new bounds on the second derivative of mutual information, or the first derivative of MMSE, that tighten previously known bounds.**

I. Introduction

Consider the classical point-to-point Gaussian channel:

$$\mathbf{Y} = \sqrt{\mathsf{snr}}\ \mathbf{X} + \mathbf{Z}, \tag{1}$$

where $\mathbf{Z}, \mathbf{X}, \mathbf{Y} \in \mathbb{R}^n$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is independent of $\mathbf{X}$ and where $\mathsf{snr} \geq 0$ is the signal to noise ratio (SNR). When it will be necessary to emphasize the SNR at output $\mathbf{Y}$ we will denote it with $\mathbf{Y}_{\mathsf{snr}}$. The minimum mean squared error (MMSE) of estimating $\mathbf{X}$ from $\mathbf{Y}$ plays a key role in Bayesian statistics and estimation theory and is defined as

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \mathrm{mmse}(\mathbf{X}, \mathsf{snr}) := n^{-1} \inf_f \mathbb{E}\left[\mathsf{Err}\left(\mathbf{X}, f(\mathbf{Y})\right)\right]$$

$$= n^{-1}\mathbb{E}\left[\mathsf{Err}\left(\mathbf{X}, \mathbb{E}[\mathbf{X}|\mathbf{Y}]\right)\right], \text{ where} \tag{2a}$$

$$\mathsf{Err}\left(\mathbf{X}, f(\mathbf{Y})\right) := \mathrm{Tr}\left(\left(\mathbf{X} - f(\mathbf{Y})\right)\left(\mathbf{X} - f(\mathbf{Y})\right)^T\right), \tag{2b}$$

where $\mathrm{Tr}(\cdot)$ is the trace operation. In the Bayesian setting the MMSE in (2a) is understood as a cost function with the quadratic loss function (i.e. L_2 norm) defined in (2b). Another commonly used cost function is the L_1 norm with loss function given by the absolute value of error (i.e., the difference between the variable of interest and its estimate). However, other order errors are far less well understood for the approximation theoretic treatment of L_p spaces see [1]. Motivated by the study of cost functions in which the loss function is of a different error order, for any random variable $\mathbf{U}$ we define a norm for $2p \geq 1$ by

$$\|\mathbf{U}\|_p := n^{-\frac{1}{2p}}\ \mathbb{E}^{\frac{1}{2p}}[\mathrm{Tr}^p\left(\mathbf{U}\mathbf{U}^T\right)], \tag{3}$$

with the triangle inequality shown in [2]. Throughout the paper we define the L_{2p} space, for $2p \geq 1$, as the space of random vectors on a fixed probability space $(\Omega, \sigma(\Omega), \mathbb{P})$ such that the norm defined in (3) is finite. The minimum mean p-th error (MMPE) in estimating $\mathbf{X}$ from $\mathbf{Y}$ is defined as

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) := \inf_f \|\mathbf{X} - f(\mathbf{Y})\|_p^{2p}$$

$$= \inf_f n^{-1}\mathbb{E}\left[\mathsf{Err}^p\left(\mathbf{X}, f(\mathbf{Y})\right)\right], \tag{4}$$

where the minimization is over all possible Borel measurable estimators $f(\mathbf{Y})$. The optimal MMPE estimator of order p of $\mathbf{X}$ is denoted by $f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$. In particular, $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, 1) = \mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ with $f_1(\mathbf{X}|\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$. The definition of loss function used in (4) is motivated by:

- it reduces to a natural expression with loss function given by $\mathsf{Err}^p(X, f(Y)) = |X - f(Y)|^{2p}$ for $X \in \mathbb{R}^1$,
- it naturally appears in applications of Holder's or Jensen's inequality to (2). This is the key motivation for studying MMPE: as a tool to develop new bounds on the MMSE.

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

$$n^{-1}\mathbb{E}\left[\mathsf{Err}^p\left(\mathbf{X}, \mathbb{E}[\mathbf{X}|\mathbf{Y}]\right)\right] = \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^{2p}, \tag{5}$$

which represents higher order moments of the loss function in (2b) and serves (see below) as an upper bound on (4).

MMPE for $2p \neq 2$ differs from MMSE in a number of aspects. The main difference is that the norm defined in (3) is not a Hilbert space norm in general (unless $2p = 2$); 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 $2p \neq 2$.

Notation. We adopt the following notational conventions: deterministic scalar/vector quantities are denoted by lower normal/bold case letters, matrices by bold upper case letters, random variables by upper case letters, and random vectors by bold uppercase letters; $\Gamma(x)$ denotes the gamma function.

Past Work. The key application of MMPE is to derive new bounds on the first (MMSE) and second derivative of mutual information. Next, we review some properties of the MMSE and its derivative that are relevant for this work.

Properties of the MMSE for the channel in (1) have been throughly explored in [3]. Of particular interest to this work is the 'single-crossing point property' bound developed in [4] for $n = 1$ and in [5] for $n \geq 1$, stated next.[1]

[1]The single-crossing point property is not stated in full generality here, see [4, Proposition 16], [5, Theorem 1] for the complete statement of the theorem.

Proposition 1. (Single-crossing point property). *For any fixed* $\mathbf{X}$*, suppose that* $\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\beta}{1+\beta \mathsf{snr}_0}$*, for some* $\beta \geq 0$*. Then for all* $\mathsf{snr} \in [\mathsf{snr}_0, \infty)$ *we have that*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \frac{\beta}{1+\beta \mathsf{snr}}. \quad (6)$$

The single-crossing point property with the I-MMSE relationship $I_n(\mathbf{X}, \mathsf{snr}) := \frac{1}{n} I(\mathbf{X}; \mathbf{Y}_{\mathsf{snr}}) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t) dt$ [6], have been used: in [4] to provide an alternative proof of the converse for the Gaussian broadcast channel (BC) and show a special case of the entropy power inequality (EPI); in [3] to provide a simple proof for the information combining problem and a converse for the BC with confidential messages; in [5], by using various extensions of (6), 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 [7] to provide a converse for communication under an MMSE disturbance constraint.

The single-crossing point property is also 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:

$$\limsup_{n\to\infty} \mathrm{mmse}(\mathbf{X}, \mathsf{snr}) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & 0 \leq \mathsf{snr} \leq \mathsf{snr}_0 \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr}_0 \leq \mathsf{snr} \leq \mathsf{snr}_1 \\ \frac{\gamma}{1+\gamma\mathsf{snr}}, & \mathsf{snr} \geq \mathsf{snr}_1 \end{cases}, \quad (7)$$

as shown: in [8], for the Gaussian point-to-point channel with the output Y_{snr_0} with $\beta = \gamma = 0$; in [9], 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 [9], 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 [7], 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 (7) 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 (7) behaves for codes of finite length. In [10], in order to study the phase transition phenomenon for inputs of finite length, the following optimization problem was proposed:

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) := \sup_{\mathbf{X}} \mathrm{mmse}(\mathbf{X}, \mathsf{snr}), \quad (8a)$$

$$\text{s.t. } \|\mathbf{X}\|_1^2 \leq 1, \text{ and } \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}, \quad (8b)$$

for some $\beta \in [0,1]$. Investigation in [10] revealed that $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (8a) 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 [10] the following was established for $T_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ and $W(n)$.

Theorem 1. *For any fixed* $\mathbf{X}$ *and* $\mathsf{snr} \in [0, \mathsf{snr}_0]$*, let* $\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\beta}{1+\beta\mathsf{snr}_0}$*. Then*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) + \kappa_n \left(\frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0}\right), \quad \text{where } \kappa_n \leq n+2. \quad (9)$$

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

The MMPE will also be used in the study of the second derivative of mutual information (or first derivative of MMSE), as initiated for $n = 1$ in [4] and for $n \geq 1$ in [5], namely,

$$\frac{d^2 I(\mathbf{X}; \mathbf{Y})}{d\mathsf{snr}^2} = n \frac{d\ \mathrm{mmse}(\mathbf{X}, \mathsf{snr})}{d\mathsf{snr}} = -\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right),$$
$$\mathbf{Cov}(\mathbf{X}|\mathbf{Y}) := \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^T | \mathbf{Y}\right]. \quad (10)$$

The second derivative of mutual information is important in characterizing the bandwidth-power trade-off in the wideband regime [11], [12] and it has also been used in the proof of the single-crossing point property in [4] and [5]. Moreover, in [4] it has been shown that the derivative of the MMSE and the quantity in (5) are related by the following bound for $n = 1$:

$$\mathbb{E}\left[\mathrm{Cov}^2(X|Y)\right] \leq \|X - \mathbb{E}[X|Y]\|_2^4 \leq \frac{3 \cdot 2^4}{\mathsf{snr}^2}. \quad (11)$$

Finally, it is worth pointing out that in [13] yet another cost function, referred to as the *score function*, was shown to be useful in establishing information-estimation relationships.

Paper Outline and Main Contributions. In Section II we study properties of the optimal estimator $f_p(\mathbf{X}|\mathbf{Y})$: in Proposition 2 we show that the optimal estimator indeed exists; in Proposition 3 we find the exact value of the MMPE and the optimal estimator for Gaussian inputs, and in Proposition 4 we find the optimal estimator for BPSK inputs; in Proposition 5 we compare well known properties of the MMSE to those for the MMPE. In Section III we develop several bounds on MMPE such as: in Proposition 7 we find bounds, equivalent to that of the linear MMSE (LMMSE) bound, for the MMPE; and in Proposition 8 we derive interpolation bounds for the MMPE; in Proposition 9 we show that MMPE is a continuous function of p; and in Theorem 2 we upper bound the MMPE at a lower SNR with the MMPE of a different order at a higher SNR. In Section IV we show how the tools developed in Section III can be applied to find new bounds on the MMSE and the derivative of MMSE: in Theorem 3 we show a bound that complements the single-crossing point property and improves the characterization of the width phase transition region in Theorem 1, provides a lower bound on how fast a capacity achieving code sequence will converge to (7); and in Theorem 4 we show how the MMPE can be used to provide new lower and upper bounds on the derivative of the MMSE, where for $n = 1$ we show an improvement of the bound in (11) and generalize it to arbitrary n. Due to space limitations, the proofs are omitted and can be found in the extended version of the paper [2].

II. On the Optimal MMPE Estimator

It is important to point out that $\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^{2p}$ in general is not equal to 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 the optimizing $f_p(\mathbf{X}|\mathbf{Y}=\mathbf{y})$ indeed exists.

Proposition 2. *For* $\mathrm{mmpe}(\mathbf{X},\mathsf{snr},p)$ *and* $p \geq 0$ *the optimal estimator is given by the following point-wise relationship:*

$$f_p(\mathbf{X}|\mathbf{Y}=\mathbf{y}) = \arg\min_{\mathbf{v}\in\mathbb{R}^n} \mathbb{E}\left[\mathsf{Err}^p(\mathbf{X},\mathbf{v})|\mathbf{Y}=\mathbf{y}\right]. \quad (12)$$

In general we do not have an analytical solution to (12). Interestingly, the optimal estimator for scalar Gaussian inputs can be found and is the same for all p and is linear.

Proposition 3. *For input* $X \sim \mathcal{N}(0,1)$ *and* $2p \geq 1$

$$\mathrm{mmpe}(X,\mathsf{snr},p) = \frac{2^p\Gamma\left(\frac{2p+1}{2}\right)}{\sqrt{\pi}(1+\mathsf{snr})^p},$$

with optimal estimator given by $f_p(X|Y=y) = \frac{\sqrt{\mathsf{snr}}\,y}{1+\mathsf{snr}}$.

The optimal MMPE estimator is in general a function of p as shown next for the practically relevant case of BPSK modulation, or $X \in \{\pm 1\}$ with equal probability.

Proposition 4. *For BPSK input and* $2p \geq 1$ *we have*

$$f_p(X|Y=y) = \tanh\left(\frac{y\sqrt{\mathsf{snr}}}{2p-1}\right).$$

Not all known properties of $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ and $\mathrm{mmse}(\mathbf{X}|\mathbf{Y})$ are also exhibited by $f_p(\mathbf{X}|\mathbf{Y})$ and $\mathrm{mmpe}(\mathbf{X},\mathsf{snr},p)$.

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

1) *if* $0 \leq X \in \mathbb{R}^1$ *then* $0 \leq f_p(X|Y)$,
2) *(Linearity)* $f_p(a\mathbf{X}+b|\mathbf{Y}) = af_p(\mathbf{X}|\mathbf{Y})+b$,
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) *(Degradeness)* $f_p\left(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0},\mathbf{Y}_{\mathsf{snr}}\right) = f_p\left(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}\right)$, *for a Markov chain* $\mathbf{X}\to\mathbf{Y}_{\mathsf{snr}_0}\to\mathbf{Y}_{\mathsf{snr}}$,
6) *(Orthogonality Principle) It only holds for* $2p=2$ *(when MMPE corresponds to MMSE) as shown in Fig. 1, 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* $2p=2$,
7) *(Shift)* $\mathrm{mmpe}(\mathbf{X}+a,\mathsf{snr},p) = \mathrm{mmpe}(\mathbf{X},\mathsf{snr},p)$, *and*
8) *(Scaling)* $\mathrm{mmpe}(a\mathbf{X},\mathsf{snr},p) = a^{2p}\mathrm{mmse}(\mathbf{X},a^2\mathsf{snr},p)$.

III. Bounds and Properties of the MMPE

a) Bounds: An important upper bound on the MMSE often used in practice is the LMMSE bound.

Proposition 6. (LMMSE *[4].) For any input* $\mathbf{X}$ *and* $\mathsf{snr} > 0$

$$\mathrm{mmse}(\mathbf{X},\mathsf{snr}) \leq \frac{1}{\mathsf{snr}}. \quad \text{(13a)}$$

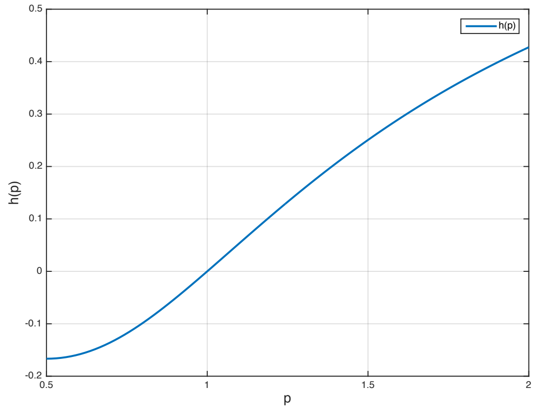


Fig. 1: Plot of the correlation between the error and observation $h(p) := \mathbb{E}[(X-f_p(X|Y))Y]$ vs. p, for $X \sim$ BPSK and $\mathsf{snr}=1$ and where $f_p(X|Y)$ for any $2p \geq 1$ is given in Proposition 4.

If $\|\mathbf{X}\|_1^2 = \sigma^2 < \infty$*, then for any* $\mathsf{snr} \geq 0$

$$\mathrm{mmse}(\mathbf{X},\mathsf{snr}) \leq \frac{\sigma^2}{1+\sigma^2\mathsf{snr}}, \quad \text{(13b)}$$

where equality in (13b) *is achieved iff* $\mathbf{X} \sim \mathcal{N}(0,\sigma^2\mathbf{I})$.

Next, we present bounds on the MMPE and $\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p$ that generalize Proposition 6, as well as, [4, Propositions 4 and 5] to higher order errors and input $\mathbf{X}$ of any dimension $n \geq 1$.

Proposition 7. *For* $\mathsf{snr} \geq 0$, $0 < q \leq p$, *and any input* $\mathbf{X}$

$$n^{\frac{p-q}{q}}\mathrm{mmpe}^{\frac{p}{q}}(\mathbf{X},\mathsf{snr},q) \leq \mathrm{mmpe}(\mathbf{X},\mathsf{snr},p) \leq \|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^{2p}. \quad \text{(14a)}$$

Moreover,

$$\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^{2p} \overset{p\geq 1}{\leq} 2^{2p}\min\left(\frac{\|\mathbf{Z}\|_p^{2p}}{\mathsf{snr}^p},\|\mathbf{X}\|_p^{2p}\right), \quad \text{(14b)}$$

$$\mathrm{mmpe}(\mathbf{X},\mathsf{snr},p) \overset{p\geq 0}{\leq} \min\left(\frac{\|\mathbf{Z}\|_p^{2p}}{\mathsf{snr}^p},\|\mathbf{X}\|_p^{2p}\right). \quad \text{(14c)}$$

If $\|\mathbf{X}\|_p^{2p} < \infty$ *then for* $p \geq 0$

$$\mathrm{mmpe}(\mathbf{X},\mathsf{snr},p) \leq \frac{\left\|\|\mathbf{Z}\|_p^{2p}\mathbf{X}-\sqrt{\mathsf{snr}}\|\mathbf{X}\|_p^{2p}\mathbf{Z}\right\|_p^{2p}}{\left(\|\mathbf{Z}\|_p^{2p}+\mathsf{snr}\|\mathbf{X}\|_p^{2p}\right)^p}, \quad \text{(14d)}$$

where $\|\mathbf{Z}\|_p^{2p} = \frac{2^p\Gamma\left(\frac{n}{2}+p\right)}{n\Gamma\left(\frac{n}{2}\right)}$.

It is not difficult to check that for $2p=2$ Proposition 7 reduces to Proposition 6. The reason that bounds on $\|\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p$ are only available for $p \geq 1$, while the bounds on $\mathrm{mmpe}(\mathbf{X},\mathsf{snr},p)$ are available for $p \geq 0$, is because the proof of the bound in (14b) uses Jensen's inequality, which requires $p \geq 1$, while the proof of the bound in (14c) does not.

b) Interpolation and Continuity: One of the key advantages of using MMPE is that the MMPE of order q can be tightly predicted based on the knowledge of the MMPE at a lower order p and the MMPE at a higher order r. At the heart of this analysis is the interpolation of L^p spaces [14]: given $0 \le p \le q \le r$ 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 $\|X\|_q \le \|X\|_p^{\alpha} \|X\|_r^{(1-\alpha)}$, which implies that the norm is log-convex and thus a continuous function of p [15, Theorem 5.1.1]. Next, we present several interpolation results for the MMPE.

Proposition 8. (Log-Convexity.) *For any* $0 < p < q < r \le \infty$ *and* $\alpha \in (0,1)$ *such that*

$$\frac{1}{q} = \frac{\alpha}{p} + \frac{\bar{\alpha}}{r} \Longleftrightarrow \alpha = \frac{q^{-1} - r^{-1}}{p^{-1} - r^{-1}},$$

where $\bar{\alpha} = 1 - \alpha$, *we have*

$$\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_q \le \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_p^{\alpha} \ \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_r^{\bar{\alpha}},$$

$$\mathrm{mmpe}^{\frac{1}{q}}(\mathbf{X}, \mathsf{snr}, q) \le \|\mathbf{X} - f_r(\mathbf{X}|\mathbf{Y})\|_p^{2\alpha} \ \mathrm{mmpe}^{\frac{\bar{\alpha}}{r}}(\mathbf{X}, \mathsf{snr}, r).$$

From log-convexity we can deduce continuity.

Proposition 9. (Continuity.) *For fixed* $\mathbf{X}$ *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* $p \ge 0$.

c) Change of Measure: The next result enables us to change the expectation from $\mathbf{Y}_{\mathsf{snr}}$ to $\mathbf{Y}_{\mathsf{snr}_0}$; this is particularly useful when we know the MMPE, or the structure of the estimator, at one SNR value but not at another.

Proposition 10. *For any* $\mathbf{X}$ *and any* $\mathsf{snr} \in (0, \mathsf{snr}_0]$, $p \ge 0$

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \tag{16}$$
$$= \inf_f \frac{1}{n} \mathbb{E}\left[\mathsf{Err}^p(\mathbf{X}, f(\mathbf{Y}_{\mathsf{snr}_0})) \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathrm{e}^{\frac{\mathsf{snr}_0 - \mathsf{snr}}{2\mathsf{snr}_0} \sum_{i=1}^n Z_i^2} \right].$$

One must be careful when evaluating Proposition 10. 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 (16) is zero while $\mathrm{mmpe}(X, 0, p)$ is not, by Proposition 7, thus violating the equality. However, closer examination shows that when $\mathsf{snr} \to 0$ the limit and expectation cannot be exchanged

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

where in the last equality we used the moment generating function of (the chi-square random variable) Z^2. As an example, Proposition 10 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

$$\mathbb{E}\left[\mathsf{Err}(X, f(Y_{\mathsf{snr}_0})) \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathrm{e}^{\frac{\mathsf{snr}_0 - \mathsf{snr}}{2\mathsf{snr}_0} Z^2} \right]$$
$$\stackrel{a)}{=} (1 - \sqrt{\mathsf{snr}_0} a)^2 \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathbb{E}[X^2] \mathbb{E}\left[\mathrm{e}^{\frac{\mathsf{snr}_0 - \mathsf{snr}}{2\mathsf{snr}_0} Z^2} \right]$$
$$+ a^2 \sqrt{\frac{\mathsf{snr}}{\mathsf{snr}_0}} \mathbb{E}\left[Z^2 \mathrm{e}^{\frac{\mathsf{snr}_0 - \mathsf{snr}}{2\mathsf{snr}_0} Z^2} \right] \stackrel{b)}{=} \frac{1}{1 + \mathsf{snr}},$$

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 choosing $a = \frac{\mathsf{snr}}{\sqrt{\mathsf{snr}_0}(1 + \mathsf{snr})}$ to minimize the expression.

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

Theorem 2. *For* $0 < \mathsf{snr} \le \mathsf{snr}_0$, *any* $\mathbf{X}$ *and any* $p \ge 0$

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

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

The bound in Theorem 2, whose proof hinges on Hölder's inequality and the change of measure technique in Proposition 10, is the key in showing new bounds on the phase transitions region for MMSE, presented in the next section.

IV. Applications

In this section we show that the MMPE, besides its interest in estimation theory, can be used to study quantities that are important for information theoretic applications such as the first (MMSE) and the second derivative of mutual information.

a) New bounds on the MMSE: The main result of the subsection is shown next, which uses Theorem 2 and Proposition 8.

Theorem 3. *For* $0 < \mathsf{snr} \le \mathsf{snr}_0$,

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \min_{r > \frac{1}{\gamma}} \frac{\sqrt{2}}{n^{1-\gamma}} \left(\frac{1+\gamma}{\gamma} \right)^{\frac{n(1-\gamma)-1}{2}} M_1^{\frac{\gamma r - 1}{r-1}} M_r^{\frac{1-\gamma}{r-1}}, \tag{17a}$$

$$\text{where } \gamma := \frac{\mathsf{snr}}{2\mathsf{snr}_0 - \mathsf{snr}} \in (0, 1], \tag{17b}$$

$$M_1 := \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\beta}{1 + \beta \mathsf{snr}_0}, \tag{17c}$$

$$M_r := \|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}]\|_r^{2r} \le 2^{2r} \min\left(\frac{\|\mathbf{Z}\|_r^{2r}}{\mathsf{snr}_0^r}, \|\mathbf{X}\|_r^{2r} \right), \tag{17d}$$

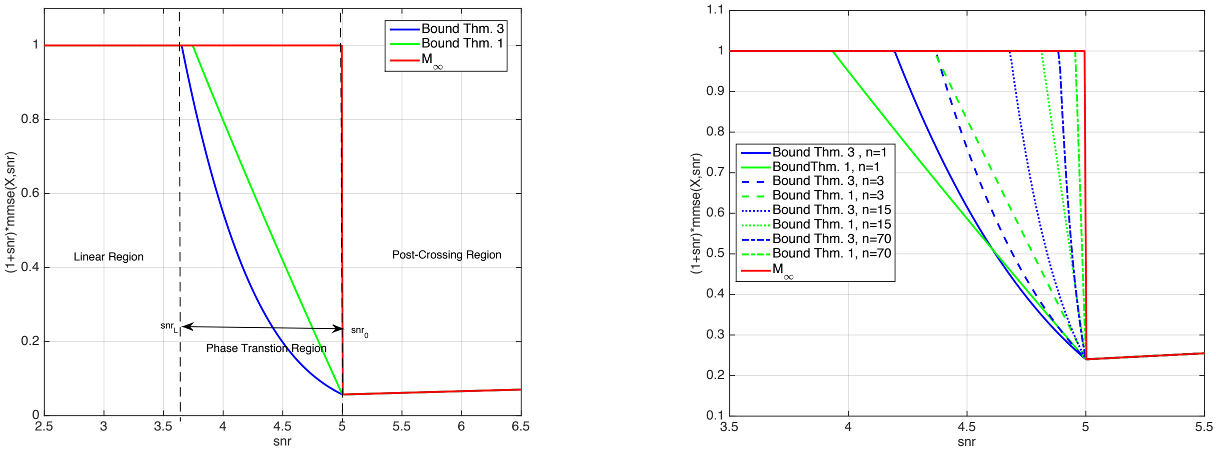
and where the minimizing r *in* (17a) *can be approximated by*

$$r_{opt} \approx \begin{cases} \ln\left(\frac{4\mathrm{e}}{\mathsf{snr}_0 \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0)} \right), & \frac{1}{\gamma} \le \ln\left(\frac{4\mathrm{e}}{\mathsf{snr}_0 \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0)} \right) \\ \frac{1}{\gamma}, & \frac{1}{\gamma} > \ln\left(\frac{4}{\mathsf{snr}_0 \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0)} \right) \end{cases}.$$

Moreover, the width of the phase transition region is given by

$$W(n) = O\left(n^{-\frac{1}{2}} \right). \tag{17e}$$

The bounds in Theorems 3 and 1 are shown in Fig. 2. The bound in Theorem 3 is asymptotically tighter than the one in Theorem 1. This follows since the phase transition region shrinks as $O\left(\frac{1}{\sqrt{n}} \right)$ for Theorem 3, while as $O\left(\frac{1}{n} \right)$ for Theorem 1. It is not possible in general to assert that Theorem 3 is tighter than Theorem 1. In fact, for small values of n, the bound in Theorem 1 can offer advantages, as seen for the case $n = 1$ shown in Fig. 2b. Another advantage of the bound in Theorem 1 is its analytical simplicity.



(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.

Fig. 2: Bounds on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs snr.

b) Bounds on the derivative of the MMSE in (10)*:* The main result of this subsection is the next bound.

Theorem 4. *For any input* $\mathbf{X}$

$$\mathrm{mmse}^2(\mathbf{X}, \mathsf{snr}) = \mathrm{mmpe}^2(\mathbf{X}, \mathsf{snr}, 1)$$
$$\leq \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right) \leq n \ \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, 2). \quad (18)$$

It can be observed that, for the case $n = 1$, by using the bound in (14b) from Proposition 7 we have that

$$\mathbb{E}\left[\mathrm{Cov}^2(X|Y)\right] \leq \mathrm{mmpe}(X, \mathsf{snr}, 2) \leq \frac{3}{\mathsf{snr}^2}, \quad (19)$$

which significantly reduces the constant in (11) from $3 \cdot 2^4$ to 3. A bound similar to that in (19) has been shown in [10, Proposition 9 and 10] via a different method.

V. Concluding Remark

This paper has considered the problem of estimating a random variable from a noisy observation under a very general cost function, termed the MMPE. As a tool the MMPE has been applied to show a new bound on the MMSE that complements the single-crossing point property. The MMPE has also been used to refine bounds on the derivative of the MMSE. Even though not reported here, due to space limitations, the MMPE can also be used to improve, asymptotically, the converse bound of the disturbance constrained problem studied in [10].

An interesting future direction is to study interactions between the signal dimension n and p as was done in [16].

Acknowledgment. The work of Alex Dytso, Daniela Tuninetti and Natasha Devroye was partially funded by NSF under award 1422511. The work of Ronit Bustin was supported in part by the women postdoctoral scholarship of Israel's Council for Higher Education (VATAT) 2014-2015. The work of H. Vincent Poor and Ronit Bustin was partially supported by NSF under awards CCF-1420575 and ECCS-1343210. The work of Shlomo Shamai was supported by the Israel Science Foundation (ISF). 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] M. J. D. Powell, *Approximation Theory and Methods*. Cambridge University Press, 1981.

[2] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, H. V. Poor, and S. Shamai, "On the minimum mean p-th error in Gaussian noise channels and its applications," To be submitted to *IEEE Trans. Inf. Theory,* http://odytso2.people.uic.edu/papers/MMPEjournal.pdf, 2016.

[3] D. Guo, S. Shamai, and S. Verdú, *The Interplay Between Information and Estimation Measures*. now Publishers Incorporated, 2013.

[4] D. Guo, Y. Wu, S. Shamai, 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, April 2011.

[5] R. Bustin, M. Payaró, D. P. Palomar, and S. Shamai, "On MMSE crossing properties and implications in parallel vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 818–844, Feb 2013.

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

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

[8] N. Merhav, D. Guo, and S. Shamai, "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, March 2010.

[9] 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, April 2016.

[10] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, S. Shamai, and H. V. Poor, "On communications through a Gaussian channel with a MMSE disturbance constraint," in *Proc. IEEE Int. Symp. Inform. Theory and its Applications*, Feb 2016.

[11] 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.

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

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

[14] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2013.

[15] R. Webster, *Convexity*. Oxford University Press, 1994.

[16] 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.