On the Applications of the Minimum Mean p-th Error (MMPE) to Information Theoretic Quantities

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*—This paper considers the minimum mean p-th error (MMPE) estimation problem: estimating a random vector in the presence of additive white Gaussian noise (AWGN) in order to minimize an L_p norm of the estimation error.**

The MMPE generalizes the classical minimum mean square error (MMSE) estimation problem. This paper derives basic properties of the optimal MMPE estimator and MMPE functional. Optimal estimators are found for several inputs of interests, such as Gaussian and binary symbols. Under an appropriate p-th moment constraint, the Gaussian input is shown to be asymptotically the hardest to estimate for any $p \geq 1$. By using a conditional version of the MMPE, the famous "MMSE single-crossing point" bound is shown to hold for the MMPE too for all $p \geq 1$, up to a multiplicative constant. Finally, the paper develops connections between the conditional differential entropy and the MMPE, which leads to a tighter version of the Ozarow-Wyner lower bound on the rate achieved by discrete inputs on AWGN channels.

I. Introduction

Notation: Deterministic scalar/vector quantities are denoted by lowercase normal/bold letters, matrices by bold uppercase letters, random variables by uppercase letters, and random vectors by bold uppercase letters. For a random vector $\mathbf{V}$ we denote the support by $\mathsf{supp}(\mathbf{V})$, covariance matrix by $\mathbf{K}_{\mathbf{V}}$, determinant by $|\mathbf{K}_{\mathbf{V}}|$, transpose by $\mathbf{V}^T$, and trace by $\mathrm{Tr}(\mathbf{V})$. $\Gamma(\cdot)$ denotes the gamma function.

Problem Formulation: Consider the classical point-to-point Gaussian channel,

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

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

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

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

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

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 expected absolute difference between the variable of interest and its estimate, which uses the L_1 norm instead of the L_2.

For various theoretical and practical reasons, the following, which generalizes the notion of the p-th moment, has been used [1]:

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

where the quantity in (3) is a norm for $p \geq 1$. In particular, for $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I})$ the norm is given by

$$n\|\mathbf{Z}\|_p^p = 2^{\frac{p}{2}} \frac{\Gamma\left(\frac{n+p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \text{ for } n \in \mathbb{N}, p \geq 0. \tag{4}$$

In [2] the norm in (3) was used to define the following generalization of the MMSE. The minimum mean p-th error (MMPE) in estimating $\mathbf{X}$ from $\mathbf{Y}$ is defined as:

$$\begin{aligned}\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) &:= \inf_f \|\mathbf{X} - f(\mathbf{Y})\|_p^p \\ &= \inf_f n^{-1}\mathbb{E}\left[\mathsf{Err}^{\frac{p}{2}}\left(\mathbf{X}, f(\mathbf{Y})\right)\right],\end{aligned}$$

where the minimization is over all measurable estimators $f(\mathbf{Y})$. The optimal MMPE estimator of $\mathbf{X}$ of order p is denoted by $f_p(\mathbf{X}|\mathbf{Y})$, and $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, 2) = \mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ with $f_2(\mathbf{X}|\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$. The MMPE is a function of the input distribution $\mathbf{X}$, the SNR, and the order p.

Past Work: Properties of the MMSE for the channel in (1) have been throughly explored in [3]. Of particular interest here is the "single-crossing point property" bound developed in [3] for $n = 1$ and in [4] for $n \geq 1$. In [5] the authors considered a scalar additive channel and derived conditions on the input and noise distributions under which the optimal estimator is linear. The current notion of MMPE has been introduced and studied in [6] where amongst other properties it has been shown that:

Proposition 1. *For any $p \geq 0$, the optimal MMPE estimator is given by the following point-wise almost sure relationship:*

$$f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y}) = \arg\min_{\mathbf{v} \in \mathbb{R}^n} \mathbb{E}\left[\mathsf{Err}^{\frac{p}{2}}(\mathbf{X}, \mathbf{v})|\mathbf{Y} = \mathbf{y}\right]. \tag{5}$$

Contributions and Paper Outline: In Section II-A we study properties of the optimal estimator $f_p(\mathbf{X}|\mathbf{Y})$ in (5). In Section II-B we find the optimal estimator for several inputs of interest. In Section II-C we show that in general the optimal estimator is biased. In Section III we develop several bounds on the MMPE. For example, we show that the Gaussian input is the 'hardest' to estimate under an appropriate moment constraint. In Section IV we define and study properties of the conditional MMPE. In Section V we show that the famous 'MMSE single-crossing point' bound [3], [4] holds for the MMPE for all $p \geq 1$, up to a multiplicative constant. Finally, in Section VI we show connections between entropy and the MMPE, which lead to a tighter version of the Ozarow-Wyner bound [7]. Due to space limitations, the proofs are omitted and can be found in the extended version of the paper [2].

II. Properties of Optimal Estimators

A. Basic Properties of the Optimal MMPE Estimator

Interestingly, many of the known properties of $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ – the optimal MMSE / L_2 norm estimator – are still exhibited by $f_p(\mathbf{X}|\mathbf{Y})$ in (5), as shown in [6, Prop. 6 and Prop. 7].

Proposition 2. *([6, Prop. 6]) For any $p > 0$*

$$\mathrm{mmpe}(\mathbf{X}+a, \mathsf{snr}, p) = \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p), \tag{6a}$$

$$\mathrm{mmpe}(a\mathbf{X}, \mathsf{snr}, p) = a^p \mathrm{mmse}(\mathbf{X}, a^2\mathsf{snr}, p). \tag{6b}$$

Proposition 2 asserts that the MMPE is shift invariant and shows how the MMPE behaves under scaling.

Proposition 3. *([2, Prop. 7]) For any $p > 0$ the optimal estimator $f_p(\mathbf{X}|\mathbf{Y})$ has the following properties:*

1) (Non-negativity) if $0 \le X \in \mathbb{R}^1$ then $0 \le f_p(X|Y)$,
2) (Linearity) $f_p(a\mathbf{X}+b|\mathbf{Y}) = a f_p(\mathbf{X}|\mathbf{Y}) + b$,
3) (Stability) $f_p(g(\mathbf{Y})|\mathbf{Y}) = g(\mathbf{Y})$ for any function $g(\cdot)$,
4) (Idempotence) $f_p(f_p(\mathbf{X}|\mathbf{Y})|\mathbf{Y}) = f_p(\mathbf{X}|\mathbf{Y})$,
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-like Property) $\mathbb{E}\left[\mathsf{Err}^{\frac{p-2}{2}}\left(\mathbf{X}, f_p(\mathbf{X}|\mathbf{Y})\right) \cdot \left(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})\right)^T \cdot g(\mathbf{Y})\right] = 0$ for any deterministic function $g(\cdot) \in L^p$.

Proposition 3 shows that the MMPE has many of the well known properties of the MMSE.

B. Examples of Optimal MMPE Estimators

Next, we provide examples of optimal MMPE estimators.

Proposition 4. *For $\mathbf{X}_G \sim \mathcal{N}(0, \mathbf{I})$ and $p \ge 1$*

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

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

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

Proposition 5. *For $X \in \{x_1, x_2\}$ with $P_X[X = x_1] := q \in [0,1]$, and $\bar{q} := 1-q$, and for $p \ge 1$, we have that*

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

For the practically relevant BPSK modulation, or $X_{\mathrm{BPSK}} \in \{\pm 1\}$ with equal probability, from Proposition 5 we have that

$$f_p(X_{\mathrm{BPSK}}|Y=y) = \tanh\left(\frac{y\sqrt{\mathsf{snr}}}{p-1}\right). \tag{7}$$

For $p = 1$ the optimal estimator in (7) becomes a hard decision decoder.

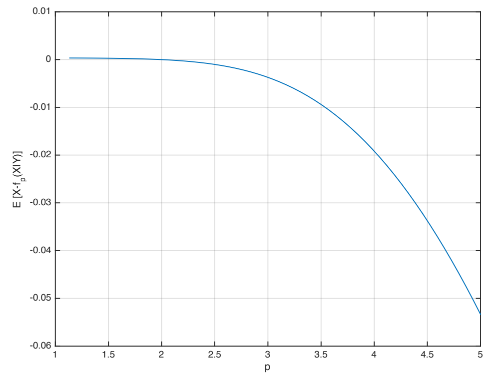

Fig. 1: Plot of $\mathbb{E}[(X - f_p(X|Y)]$ vs. p for $X \in \{-3, 1\}$ and $\Pr[X = -3] = 0.01$, with $f_p(X|Y)$ as in Proposition 5.

C. Average Bias of the Optimal MMPE Estimator

An estimator $f(\mathbf{X}|\mathbf{Y})$ is said to be unbiased on average if $\mathbb{E}[\mathbf{X} - f(\mathbf{X}|\mathbf{Y})] = 0$. In general $f_p(\mathbf{X}|\mathbf{Y})$ is unbiased on average only for $p = 2$. This follows from the law of total expectation

$$\mathbb{E}[f_{p=2}(\mathbf{X}|\mathbf{Y})] = \mathbb{E}[\mathbb{E}[\mathbf{X}|\mathbf{Y}]] = \mathbb{E}[\mathbf{X}]. \tag{8}$$

Fig. 1 shows that in general the optimal MMPE estimator is biased on average, as is common in Bayesian estimation; it plots $\mathbb{E}[(X - f_p(X|Y)]$ vs. p for $X \in \{-3, 1\} : \Pr[X = -3] = 0.01$, with $f_p(X|Y)$ as in Proposition 5. However, the optimal MMPE estimator is unbiased in the sense that the $(p-1)$-th moment of the bias is zero. This can be seen from the orthogonality like property in Proposition 3 by taking $g(\mathbf{Y}) = \mathbf{I}$.

III. Bounds on the MMPE

Our first bound generalizes the linear MMSE (LMMSE) bounds given in [3, Prop. 4] to the MMPE for any inputs $\mathbf{X}$ of any dimension $n \ge 1$.

Proposition 6. *For $\mathsf{snr} \ge 0$, $0 < q \le p$, and any input $\mathbf{X}$,*

$$n^{\frac{p}{q}-1} \mathrm{mmpe}^{\frac{p}{q}}(\mathbf{X}, \mathsf{snr}, q) \le \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p), \tag{9a}$$

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \le \min\left(\frac{\|\mathbf{Z}\|_p^p}{\mathsf{snr}^{\frac{p}{2}}}, \|\mathbf{X}\|_p^p\right). \tag{9b}$$

Moreover, if $\|\mathbf{X}\|_p \le \|\mathbf{Z}\|_p$ for $p \ge 1$, then

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \le k_{p,\mathsf{snr}} \cdot \frac{\|\mathbf{Z}\|_p^p}{(1+\mathsf{snr})^{\frac{p}{2}}}, \tag{9c}$$

where $k_{p=2,\mathsf{snr}} = 1$ and $k_{p\neq 2,\mathsf{snr}}^{\frac{1}{p}} = \frac{1+\sqrt{\mathsf{snr}}}{\sqrt{1+\mathsf{snr}}} \le 1 + \frac{1}{\sqrt{1+\mathsf{snr}}}$.

As an application of Proposition 6 we show that asymptotically a Gaussian input is the hardest to estimate.

Proposition 7. *For every $\mathsf{snr} \ge 0$, $p \ge 1$, and a random variable $\mathbf{X}$ such that $\|\mathbf{X}\|_p^p \le \sigma^p \|\mathbf{Z}\|_p^p$, we have*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \le k_{p,\sigma^2\mathsf{snr}} \frac{\sigma^p \|\mathbf{Z}\|_p^p}{(1+\mathsf{snr}\sigma^2)^{\frac{p}{2}}}. \tag{10}$$

Moreover, $\mathbf{X} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ *asymptotically achieves the bound in* (10) *since* $\lim_{\mathsf{snr}\to\infty} k_{p,\sigma^2\mathsf{snr}} = 1$.

Proof. Let $\mathbf{X} = \sigma\mathbf{U}$. Then $\|\mathbf{X}\|_p = \|\sigma\mathbf{U}\|_p \le \sigma\|\mathbf{Z}\|_p$ and thus $\|\mathbf{U}\|_p \le \|\mathbf{Z}\|_p$. By the bound in (9c), we have that

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) = \mathrm{mmpe}(\sigma\mathbf{U}, \mathsf{snr}, p) \tag{11}$$
$$\overset{a)}{=} \sigma^p \mathrm{mmpe}(\mathbf{U}, \sigma^2\mathsf{snr}, p) \tag{12}$$
$$\overset{b)}{\le} k_{p,\sigma^2\mathsf{snr}} \frac{\sigma^p \|\mathbf{Z}\|_p^p}{(1+\mathsf{snr}\sigma^2)^{\frac{p}{2}}}, \tag{13}$$

where the (in)-equalities follow from: a) the scaling property of the MMPE in (6b), and b) the bound in (9c). By Proposition 4, the bound in (10) is achieved with equality by $\mathbf{X}_G \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. This concludes the proof. □

IV. Conditional MMPE

We define the conditional MMPE as

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{U}) := \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}}, \mathbf{U})\|_p^p. \tag{14}$$

The conditional MMPE in (14) 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

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{U}) = \int \mathrm{mmpe}(\mathbf{X}_{\mathbf{u}}, \mathsf{snr}, p)\, dP_{\mathbf{U}}(\mathbf{u}). \tag{15}$$

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

Proposition 8. *For every* $\mathsf{snr} \ge 0$, $p \ge 0$, *and random variable* $\mathbf{X}$, *we have*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \ge \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{U}). \tag{16}$$

Finally, the following proposition generalizes [8, Proposition 3.4] and states that MMPE estimation of $\mathbf{X}$ from two independent observations is equivalent to estimating $\mathbf{X}$ from a single observation with a higher SNR:

Proposition 9. *For every* $\mathbf{X}$ *and* $p \ge 0$, *let* $\mathbf{U} = \sqrt{\Delta}\cdot\mathbf{X} + \mathbf{Z}_\Delta$ *where* $\mathbf{Z}_\Delta \sim \mathcal{N}(0, \mathbf{I})$ *and where* $(\mathbf{X}, \mathbf{Z}, \mathbf{Z}_\Delta)$ *are mutually independent. Then*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{U}) = \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}+\Delta, p). \tag{17}$$

Together, Propositions 9 and 8 imply that, for fixed $\mathbf{X}$ and p, the MMPE is a non-increasing function of snr.

V. Single Crossing Point Property Bound

The single crossing point property bound is a powerful tool for showing converses for Gaussian noise channels [4]. Next we generalize this important bound to the MMPE.

Proposition 10. *Suppose* $\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}_0, p) = \frac{\beta\|\mathbf{Z}\|_p^2}{1+\beta\mathsf{snr}_0}$ *for some* $\beta \ge 0$. *Then*

$$\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}, p) \le c_p \cdot \frac{\beta\|\mathbf{Z}\|_p^2}{1+\beta\,\mathsf{snr}}, \text{ for } \mathsf{snr} \ge \mathsf{snr}_0, \tag{18}$$

where $c_p = \begin{cases} 2 & p \ge 1 \\ 1 & p = 2 \end{cases}$.

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

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

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

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

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

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

for some $\gamma \in \mathbb{R}$. 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

$$\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p) = \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}})\|_p$$
$$\overset{a)}{=} \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_\Delta, \mathbf{Y}_{\mathsf{snr}_0})\|_p$$
$$\overset{b)}{\le} \|\mathbf{X} - \hat{\mathbf{X}}\|_p = \left\|\gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Z}_\Delta\right\|_p$$
$$\overset{c)}{=} \frac{\left\|\|\mathbf{Z}\|_p^2(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \sqrt{\Delta}\cdot m\cdot\mathbf{Z}_\Delta\right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}, \tag{19}$$

where the (in)-equalities follow from: a) Proposition 9, b) using a sub-optimal estimator, and c) choosing $\gamma = \frac{\|\mathbf{Z}\|_p^2}{\|\mathbf{Z}\|_p^2+\Delta\cdot m}$.

Next, by applying the triangle inequality,

$$\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p)$$
$$\le \frac{\left\|\|\mathbf{Z}\|_p^2(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}))\right\|_p + \left\|\sqrt{\Delta}\cdot m\cdot\mathbf{Z}_\Delta\right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m} \tag{20}$$
$$= \frac{\sqrt{m}\|\mathbf{Z}\|_p\cdot(\|\mathbf{Z}\|_p + \sqrt{\Delta}\cdot\sqrt{m})}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m} \le \sqrt{2}\frac{\sqrt{m}\|\mathbf{Z}\|_p}{\sqrt{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}},$$

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

Note that for the case $p = 2$, instead of using the triangle inequality in (20), the term 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 a bound

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

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

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

$$\min_{\gamma\in[0,1]} \|(1-\gamma)\mathbf{W} + \gamma\mathbf{Z}\|_p, \tag{21}$$

where $\mathbf{W}$ *and* $\mathbf{Z}$ *are independent and* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. *The optimization problem in* (21) *is the subject of current investigation.*

VI. Applications

We next show how the MMPE concept can be used to derive tighter versions of some well known information theoretic bounds. It is important to point out that even though the focus of this paper is the additive white Gaussian noise (AWGN) setting, the results that follow apply to any channel model in which the noise is an absolutely continuous random variable without an assumption of independent and identically distributed noise samples.

A. Bounds on the Differential Entropy

For any random vector $\mathbf{U}$ such that $|\mathbf{K}_\mathbf{U}| < \infty$ and $h(\mathbf{U}) < \infty$, and any random vector $\mathbf{V}$, the following inequality is considered to be a continuous analog of Fano's inequality [9]:

$$h(\mathbf{U}|\mathbf{V}) \le \frac{n}{2} \log(2\pi \mathrm{e}\ \mathrm{mmse}(\mathbf{U}|\mathbf{V})). \tag{22}$$

Therefore, by applying (22) to the AWGN setting, for any $\mathbf{X}$ such that $|\mathbf{K}_\mathbf{U}| < \infty$ and $h(\mathbf{U}) < \infty$, by using Proposition 6 (i.e., $q = 2$) we can arrive at the trivial bound:

$$h(\mathbf{X}|\mathbf{Y}) \le \frac{n}{2} \log\left(2\pi \mathrm{e} \cdot n^{\frac{2-p}{p}} \cdot \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}, p)\right). \tag{23}$$

Next, we show that the trivial bound in (23) can be improved.

Proposition 11. *For any* $\mathbf{U} \in \mathbb{R}^n$ *such that* $\|\mathbf{U}\|_p < \infty$ *and* $h(\mathbf{U}) < \infty$ *for some* $p \in (0, \infty)$ *and any* $\mathbf{V} \in \mathbb{R}^n$*, we have*

$$h(\mathbf{U}|\mathbf{V}) \le \frac{n}{2} \log\left(k_{n,p}^2 \cdot n^{\frac{2}{p}} \cdot \|\mathbf{U} - g(\mathbf{V})\|_p^2\right), \tag{24}$$

for any deterministic function $g(\cdot)$ *and where*

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

Proof. 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 [1, Theorem 3] we have

$$k_{n,p} \cdot n^{\frac{1}{p}} \|\mathbf{W}_\mathbf{v}\|_p \ge e^{\frac{1}{n} h_\mathrm{e}(\mathbf{W}_\mathbf{v})}, \tag{25}$$

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 (25) and by using the translation invariance of the differential entropy, we get

$$n^{-1} h_\mathrm{e}(\mathbf{U}_\mathbf{v}) \log(\mathrm{e}) \le \log\left(k_p \cdot n^{\frac{1}{p}} \|\mathbf{W}_\mathbf{v}\|_p\right), \tag{26}$$

where from (2b) $n^{\frac{1}{p}} \|\mathbf{W}_\mathbf{v}\|_p = \mathbb{E}^{\frac{1}{p}}\left[\mathsf{Err}^{\frac{p}{2}}(\mathbf{U}, g(\mathbf{V}))|\mathbf{V} = \mathbf{v}\right]$. By taking the expectation on both sides of (26) with respect to $p_\mathbf{V}(\mathbf{v})$ we arrive at

$$\begin{aligned} &n^{-1} h_\mathrm{e}(\mathbf{U}|\mathbf{V}) \log(\mathrm{e}) = n^{-1} h(\mathbf{U}|\mathbf{V}) \\ &\le \frac{1}{p} \mathbb{E}\left[\log\left(k_p^p \cdot n \cdot \frac{1}{n} \cdot \mathbb{E}\left[\mathsf{Err}^{\frac{p}{2}}(\mathbf{U}, g(\mathbf{V}))|\mathbf{V}\right]\right)\right] \\ &\overset{a)}{\le} \frac{1}{p} \log\left(k_p^p \cdot n \cdot \frac{1}{n} \cdot \mathbb{E}\left[\mathbb{E}\left[\mathsf{Err}^{\frac{p}{2}}(\mathbf{U}, g(\mathbf{V}))|\mathbf{V}\right]\right]\right) \\ &= \frac{1}{p} \log\left(k_p^p \cdot n \cdot \frac{1}{n} \cdot \mathbb{E}\left[\mathsf{Err}^{\frac{p}{2}}(\mathbf{U}, g(\mathbf{V}))\right]\right) \\ &= \log\left(k_p \cdot n^{\frac{1}{p}} \cdot \|\mathbf{U} - g(\mathbf{V})\|_p\right), \end{aligned}$$

where the inequality in a) follows from Jensen's inequality. This concludes the proof. □

Note that the result in Proposition 11 holds in great generality and the AWGN assumption is not necessary. As an application of Proposition 11 to the AWGN setting we have the following stronger version of the inequality in (23).

Proposition 12. *For any* $\mathbf{X}$ *such that* $\|\mathbf{X}\|_p < \infty$ *and* $h(\mathbf{X}) < \infty$ *for some* $p \in (0, \infty)$*, we have that*

$$n^{-1} h(\mathbf{X}|\mathbf{Y}) \le \frac{1}{2} \log\left(k_{n,p}^2 \cdot n^{\frac{2}{p}} \cdot \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}, p)\right),$$

where $k_{n,p}$ *is defined in Proposition 11.*

Proof. The proof follows by setting $\mathbf{U} = \mathbf{X}$ and $\mathbf{V} = \mathbf{Y}$ and $g(\mathbf{Y}) = f_p(\mathbf{X}|\mathbf{Y})$ in the statement of Proposition 11. □

B. Generalized Ozarow-Wyner Bound

In [7] the following "Ozarow-Wyner lower bound" on the mutual information achieved by a discrete input on an AWGN channel was shown:

$$[H(X_D) - \mathsf{gap}]^+ \le I(X_D; Y) \le H(X_D), \tag{27a}$$

$$\mathsf{gap} = \frac{1}{2} \log\left(\frac{\pi \mathrm{e}}{6}\right) + \frac{1}{2} \log\left(1 + \frac{\mathrm{lmmse}(X, \mathsf{snr})}{d_{\min}(X_D)^2}\right), \tag{27b}$$

where $\mathrm{lmmse}(X|Y)$ is the LMMSE. The advantage of the bound in (27) compared to existing bounds is its computational simplicity. The bound on the gap in (27) has been sharpened in [10, Remark 2] to

$$\mathsf{gap} = \frac{1}{2} \log\left(\frac{\pi \mathrm{e}}{6}\right) + \frac{1}{2} \log\left(1 + \frac{\mathrm{mmse}(X, \mathsf{snr})}{d_{\min}(X_D)^2}\right), \tag{28}$$

since $\mathrm{lmmse}(X, \mathsf{snr}) \ge \mathrm{mmse}(X, \mathsf{snr})$. Next, we generalize the bound in (27) to vector discrete inputs and give the sharpest known bound on the gap term.

Proposition 13. *(*Generalized Ozarow-Wyner Bound*) Let* $\mathbf{X}_D$ *be a discrete random vector such that* $p_i = \mathbb{P}[\mathbf{X}_D = \mathbf{x}_i]$ *for* $\mathbf{x}_i \in \mathsf{supp}(\mathbf{X}_D)$*, and let* $\mathcal{K}$ *be a set of continuous random vectors, independent of* $\mathbf{X}_D$*, such that* $h(\mathbf{U}) < \infty$ *and* $\|\mathbf{U}\|_p < \infty$*, and*

$$\begin{aligned} &\mathsf{supp}(\mathbf{U} + \mathbf{x}_i) \cap \mathsf{supp}(\mathbf{U} + \mathbf{x}_j) = \emptyset, \\ &\forall\ \mathbf{x}_i, \mathbf{x}_j \in \mathsf{supp}(\mathbf{X}_D), i \ne j, \text{ and } \forall\ \mathbf{U} \in \mathcal{K}. \end{aligned} \tag{29}$$

Then for any $p \ge 1$

$$[H(\mathbf{X}_D) - \mathsf{gap}_p]^+ \le I(\mathbf{X}_D; \mathbf{Y}) \le H(\mathbf{X}_D), \tag{30}$$

where

$$\mathsf{gap}_p = \inf_{\mathbf{U} \in \mathcal{K}} \mathsf{gap}(\mathbf{U}), \quad \textit{and} \quad n^{-1} \cdot \mathsf{gap}(\mathbf{U})$$

$$= \log\left(1 + \frac{\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p)}{\|\mathbf{U}\|_p}\right) + \log\left(\frac{k_{n,p} \cdot n^{\frac{1}{p}} \cdot \|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n} h_\mathrm{e}(\mathbf{U})}}\right).$$

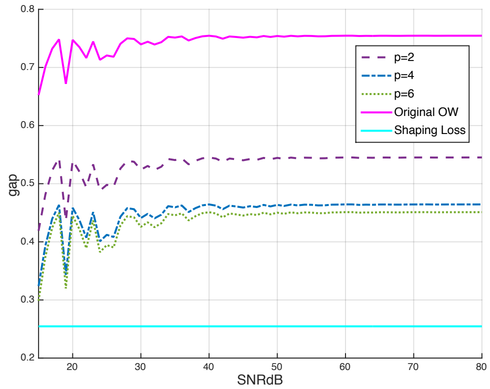

Fig. 2: Gap vs. snr.

Proof. Let $(\mathbf{U}, \mathbf{X}_D, \mathbf{Z})$ be mutually independent. By the data processing inequality and the assumption in (29) we have

$$\begin{aligned} I(\mathbf{X}_D;\mathbf{Y}) &\ge I(\mathbf{X}_D+\mathbf{U};\mathbf{Y}) = h(\mathbf{X}_D+\mathbf{U}) - h(\mathbf{X}_D+\mathbf{U}|\mathbf{Y}) \\ &= H(\mathbf{X}_D) + h(\mathbf{U}) - h(\mathbf{X}_D+\mathbf{U}|\mathbf{Y}). \end{aligned} \tag{31}$$

Next, by using Proposition 11, we have that the last term of (31) can be bounded as:

$$\begin{aligned} n^{-1} h(\mathbf{X}_D+\mathbf{U}|\mathbf{Y}) &\le \log\left(k_{n,p}\cdot n^{\frac{1}{p}} \cdot \|\mathbf{X}_D+\mathbf{U}-g(\mathbf{Y})\|_p\right) \\ &\le \log\left(k_{n,p}\cdot n^{\frac{1}{p}} \cdot (\|\mathbf{U}\|_p + \|\mathbf{X}_D - g(\mathbf{Y})\|_p)\right), \end{aligned} \tag{32}$$

where the last inequality follows by the triangle inequality. Next, by combining (31) and (32) we have that

$$\begin{aligned} I(\mathbf{X}_D;\mathbf{Y}) &\ge H(\mathbf{X}_D) - n\cdot\log\left(1+\frac{\|\mathbf{X}_D - g(\mathbf{Y})\|_p}{\|\mathbf{U}\|_p}\right) \\ &\quad - n\cdot\log\left(\frac{k_{n,p}\cdot n^{\frac{1}{p}}\cdot\|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{U})}}\right). \end{aligned}$$

Finally, the proof concludes by taking $g(\mathbf{Y}) = f_p(\mathbf{X}|\mathbf{Y})$ and taking the supremum over all possible $\mathbf{U}$. □

As a simple example of Proposition 13 consider the case of $n = 1$ and 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 $H(X) \approx \frac{1}{2}\log(1+\mathsf{snr})$. Fig. 2 shows the following: the solid cyan line is the "shaping loss" 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 due to the original Ozarow-Wyner gap in (27); and the dashed purple, dashed-dotted blue and dotted green lines are the new gap due to Proposition 13 for values of $p = 2, 4$ and 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 new version of the Ozarow-Wyner bound provides the sharpest bound for the gap. An open question is what value of p would provide the smallest gap and if that would coincide with the ultimate "shaping loss".

VII. Conclusion

We have considered the problem of estimating a random variable from a noisy observations under a very general cost function, termed the MMPE. We have shown that many of the known properties of the MMSE extend directly to the MMPE. The MMPE has been used to refine bounds on the conditional entropy and improve the gap term in the Ozarow-Wyner bound. Other possible applications of this sharpened version of the Ozarow-Wyner bound include sharpening of the bound on discrete inputs in [11] and [12]. For other applications of the MMPE, such as bounds on the first order phase transition of mutual information, please see [2]. Another recent application of the MMPE can be found in [13] in connections to the transmition of a modulated signal over a continuous-time AWGN channel.

Acknowledgment

The work of Alex Dytso, Daniela Tuninetti and Natasha Devroye was partially funded by the U. S. National Science Foundation (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 the NSF under awards CCF-1420575 and ECCS-1343210. The work of Shlomo Shamai was supported by S. and N. Grand Research Fund, and by the Israel Science Foundation. 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] 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, April 2007.

[2] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, S. Shamai, and H. V. Poor, "On the minimum mean *p*-th error in Gaussian noise channels and its applications." Submitted to *IEEE Trans. Inf. Theory,* https://arxiv.org/abs/1607.01461, 2016.

[3] 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.

[4] 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.

[5] 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, 2012.

[6] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, S. Shamai, and H. V. Poor, "On the minimum mean *p*-th error in Gaussian noise channels and its applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, July 2016.

[7] L. Ozarow and A. 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.

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

[9] T. Cover and J. Thomas, *Elements of Information Theory: Second Edition*. Wiley, 2006.

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

[11] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Proc. 48th Annual Allerton Conf. Commun., Control and Comp.*, Monticello, IL, 2010, pp. 620–628.

[12] S. Shamai, 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, 1991.

[13] N. Weinberger and N. Merhav, "Lower bounds on parameter modulation-estimation under bandwidth constraints." Submitted to *IEEE Trans. Inf. Theory,* http://arxiv.org/abs/1606.06576, Jun. 2016.