On the Applications of the Minimum Mean p^{th} Error to Information Theoretic Quantities

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

ELECTRICAL AND COMPUTER ENGINEERING COLLEGE OF ENGINEERING UIC

PRINCETON UNIVERSITY

TECHNION Israel Institute of Technology

Outline

- Notation and Past Work
- Minimum Mean p^{th} Error (MMPE)
- Properties of the MMPE
- Conditional MMPE
- Applications to Information Theory Problems

PRINCETON UNIVERSITY

Notation

$$\mathbf{X} \in \mathbb{R}^n, \ \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \leq 1,$$

$$\mathbf{Y}_{\mathrm{snr}} = \sqrt{\mathrm{snr}}\mathbf{X} + \mathbf{Z},$$

$$I_n(\mathbf{X}, \mathrm{snr}) = \frac{1}{n} I(\mathbf{X}; \mathbf{Y}_{\mathrm{snr}}),$$

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}}) = \mathrm{mmse}(\mathbf{X}, \mathrm{snr}) := \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathrm{cov}(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}})\right]\right)$$

I-MMSE relationship

$$\frac{d}{d\mathrm{snr}} I_n(\mathbf{X}, \mathrm{snr}) = \frac{1}{2}\mathrm{mmse}(\mathbf{X}, \mathrm{snr})$$

$$I_n(\mathbf{X}, \mathrm{snr}) = \frac{1}{2}\int_0^{\mathrm{snr}} \mathrm{mmse}(\mathbf{X}, t) dt$$

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

PRINCETON UNIVERSITY

Ozarow-Wyner Bound

(Ozarow-Wyner Bound.) Let X_D be a discrete random variable then

$$
\begin{aligned}
H(X_D) - \mathrm{gap} &\leq I(X_D; Y) \leq H(X_D), \\
\mathrm{gap} &= \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \mathrm{lmmse}(X_D|Y)}{d_{\min}^2(X_D)}\right), \\
\mathrm{lmmse}(X_D|Y) &= \frac{\mathbb{E}[X_D^2]}{1 + \mathrm{snr} \cdot \mathbb{E}[X_D^2]} \\
d_{\min}(X_D) &= \inf_{x_i, x_j \in \mathrm{supp}(X_D): i \neq j} |x_i - x_j|.
\end{aligned}
$$

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.

Take X_D to be PAM with $N = \sqrt{1 + \mathrm{snr}}$ then

$$\frac{1}{2}\log(1 + \mathrm{snr}) - I(X_D; Y) = \mathrm{gap} \approx 0.75 \text{ bit}$$

Ozarow-Wyner Bound

(Ozarow-Wyner Bound.) Let X_D be a discrete random variable then

$$
\begin{aligned}
H(X_D) - \mathrm{gap} \le I(X_D;Y) &\le H(X_D), \\
\mathrm{gap} &= \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \mathrm{lmmse}(X_D|Y)}{d_{\min}^2(X_D)}\right), \\
\mathrm{lmmse}(X_D|Y) &= \frac{\mathbb{E}[X_D^2]}{1 + \mathrm{snr} \cdot \mathbb{E}[X_D^2]} \\
d_{\min}(X_D) &= \inf_{x_i, x_j \in \mathrm{supp}(X_D): i \neq j} |x_i - x_j|.
\end{aligned}
$$

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.

Can this be improved

Take X_D to be PAM with $N = \sqrt{1 + \mathrm{snr}}$ then

$$\frac{1}{2}\log(1 + \mathrm{snr}) - I(X_D;Y) = \mathrm{gap} \approx 0.75 \text{ bit}$$

Ozarow-Wyner Bound

(Ozarow-Wyner Bound.) Let X_D be a discrete random variable then

$$
\begin{aligned}
H(X_D) - \text{gap} \leq I(X_D;Y) \leq H(X_D), \\
\text{gap} &= \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \text{lmmse}(X_D|Y)}{d_{\min}^2(X_D)}\right), \\
\text{lmmse}(X_D|Y) &= \frac{\mathbb{E}[X_D^2]}{1 + \text{snr} \cdot \mathbb{E}[X_D^2]} \\
d_{\min}(X_D) &= \inf_{x_i, x_j \in \text{supp}(X_D): i \neq j} |x_i - x_j|.
\end{aligned}
$$

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.

Improvement

$$
\begin{aligned}
\text{gap} &= \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \text{mmse}(X_D|Y)}{d_{\min}^2(X_D)}\right), \\
\text{mmse}(X_D|Y) &= \mathbb{E}\left[(X_D - \mathbb{E}[X_D|Y])^2\right].
\end{aligned}
$$

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.

Ozarow-Wyner Bound

(Ozarow-Wyner Bound.) Let X_D be a discrete random variable then

$$H(X_D) - \text{gap} \le I(X_D;Y) \le H(X_D),$$

$$\text{gap} = \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \text{lmmse}(X_D|Y)}{d_{\min}^2(X_D)}\right),$$

$$\text{lmmse}(X_D|Y) = \frac{\mathbb{E}[X_D^2]}{1 + \text{snr}\cdot\mathbb{E}[X_D^2]}$$

$$d_{\min}(X_D) = \inf_{x_i,x_j \in \text{supp}(X_D): i \neq j} |x_i - x_j|.$$

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.

$$\text{mmse}(X_D|Y) \le \text{lmmse}(X_D|Y),$$

$$\text{lmmse}(X_D|Y) = O\left(\frac{1}{\text{snr}}\right),$$

$$\text{mmse}(X_D|Y) = O\left(\mathrm{e}^{-\frac{\text{snr}\, d_{\min}^2}{4}}\right).$$

Improvement

$$\text{gap} = \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \text{mmse}(X_D|Y)}{d_{\min}^2(X_D)}\right),$$

$$\text{mmse}(X_D|Y) = \mathbb{E}\left[(X_D - \mathbb{E}[X_D|Y])^2\right].$$

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.

iTw 2016 CAMBRIDGE

PRINCETON UNIVERSITY

Ozarow-Wyner Bound

(Ozarow-Wyner Bound.) Let X_D be a discrete random variable then

$$
\begin{aligned}
H(X_D) - \text{gap} &\le I(X_D;Y) \le H(X_D), \\
\text{gap} &= \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12\cdot \text{lmmse}(X_D|Y)}{d_{\min}^2(X_D)}\right), \\
\text{lmmse}(X_D|Y) &= \frac{\mathbb{E}[X_D^2]}{1+\text{snr}\cdot\mathbb{E}[X_D^2]} \\
d_{\min}(X_D) &= \inf_{x_i,x_j\in\text{supp}(X_D):i\neq j} |x_i - x_j|.
\end{aligned}
$$

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.

$$\frac{1}{2}\log(1+\text{snr}) - I(X_D;Y) \approx 0.75 \text{ bit.}$$

Improvement

$$
\begin{aligned}
\text{gap} &= \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12\cdot \text{mmse}(X_D|Y)}{d_{\min}^2(X_D)}\right), \\
\text{mmse}(X_D|Y) &= \mathbb{E}\left[(X_D - \mathbb{E}[X_D|Y])^2\right].
\end{aligned}
$$

$$\frac{1}{2}\log(1+\text{snr}) - I(X_D;Y) \approx 0.65 \text{ bit.}$$

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.

PRINCETON UNIVERSITY

Ozarow-Wyner Bound

(Ozarow-Wyner Bound.) Let X_D be a discrete random variable then

$$H(X_D) - \text{gap} \le I(X_D;Y) \le H(X_D),$$
$$\text{gap} = \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \text{lmmse}(X_D|Y)}{d_{\min}^2(X_D)}\right),$$
$$\text{lmmse}(X_D|Y) = \frac{\mathbb{E}[X_D^2]}{1 + \text{snr}\cdot\mathbb{E}[X_D^2]}$$
$$d_{\min}(X_D) = \inf_{x_i,x_j \in \text{supp}(X_D): i \neq j} |x_i - x_j|.$$

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.

$$\frac{1}{2}\log(1+\text{snr}) - I(X_D;Y) \approx 0.75 \text{ bit.}$$

Improvement

$$\frac{1}{2}\log(1+\text{snr}) - I(X_D;Y) \approx 0.65 \text{ bit.}$$

$$\text{gap} = \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \text{mmse}(X_D|Y)}{d_{\min}^2(X_D)}\right),$$
$$\text{mmse}(X_D|Y) = \mathbb{E}\left[(X_D - \mathbb{E}[X_D|Y])^2\right].$$

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.

Can we do better?

PRINCETON UNIVERSITY

MMPE

The p-norm of a random vector $\mathbf{U} \in \mathbb{R}^n$ is given by

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

For $n = 1$, $\|U\|_p = \mathrm{E}[|U|^p]$.

Example: $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

$$n\|\mathbf{U}\|_p^p = 2^{\frac{p}{2}} \frac{\Gamma\left(\frac{n+p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)},$$

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt.$$

MMPE

We define the Minimum Mean p-th Error (MMPE) of estimating $\mathbf{X}$ from $\mathbf{Y}$ as

$$\boxed{\begin{gathered}\mathrm{mmpe}(\mathbf{X}|\mathbf{Y};p) := \inf_f \|\mathbf{X} - f(\mathbf{Y})\|_p^p = \inf_f \mathbb{E}\left[\mathsf{Err}^{\frac{p}{2}}(\mathbf{X}, f(\mathbf{Y}))\right], \\ \mathsf{Err} = \sqrt{(\mathbf{X} - f(\mathbf{Y}))^T(\mathbf{X} - f(\mathbf{Y}))}.\end{gathered}}$$

We denote the optimal estimator as $\boxed{f_p(\mathbf{X}|\mathbf{Y}).}$
For $n = 1$:

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

Example: for $p = 2$

$$\mathrm{mmpe}(\mathbf{X}|\mathbf{Y};p=2) = \mathrm{mmse}(\mathbf{X}|\mathbf{Y}),$$
$$f_{p=2}(\mathbf{X}|\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}].$$

We shall denote

Study Properties

$$\boxed{\mathrm{mmpe}(\mathbf{X}|\mathbf{Y};p) = \mathrm{mmpe}(\mathbf{X}, \mathrm{snr}, p).}$$

PRINCETON UNIVERSITY

Properties of the MMPE

Optimal Estimator

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

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

Orthogonality-like Property

Theorem. For any $1 \leq p < \infty$ optimal estimator $f_p(\mathbf{X}|\mathbf{Y})$ satisfies

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

where $\mathbf{W} = \mathbf{X} - f_p(\mathbf{X}|\mathbf{Y})$ for any deterministic function $g(\cdot)$.

Properties of Optimal Estimators

E[X|Y]

1. if $0 \le X \in \mathbb{R}^1$ then $0 \le \mathrm{E}[X|Y]$,
2. (Linearity) $\mathrm{E}[a\mathbf{X} + b|\mathbf{Y}] = a\mathrm{E}[\mathbf{X}|\mathbf{Y}] + b$,
3. (Stability) $\mathrm{E}[g(\mathbf{Y})|\mathbf{Y}] = g(\mathbf{Y})$ for any function $g(\cdot)$,
4. (Idempotent) $\mathrm{E}[\mathrm{E}[\mathbf{X}|\mathbf{Y}]|\mathbf{Y}] = \mathrm{E}[\mathbf{X}|\mathbf{Y}]$,
5. (Degradedness)0 $\mathrm{E}\left[\mathbf{X}|\mathbf{Y}_{\mathrm{snr}}, \mathbf{Y}_{\mathrm{snr}_0}\right] = \mathrm{E}\left[\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0}\right]$, for $\mathbf{X} \to \mathbf{Y}_{\mathrm{snr}_0} \to \mathbf{Y}_{\mathrm{snr}}$,
6. (Shift Invariance) $\mathrm{mmse}(\mathbf{X} + a, \mathrm{snr}) = \mathrm{mmse}(\mathbf{X}, \mathrm{snr})$,
7. (Scaling) $\mathrm{mmse}(a\mathbf{X}, \mathrm{snr}) = a^2 \mathrm{mmse}(\mathbf{X}, a^2\mathrm{snr})$.

$$\mathrm{E}\left[\mathrm{E}\left[X|Y\right]\right] = \mathrm{E}\left[X\right]$$

f_p(X|Y)

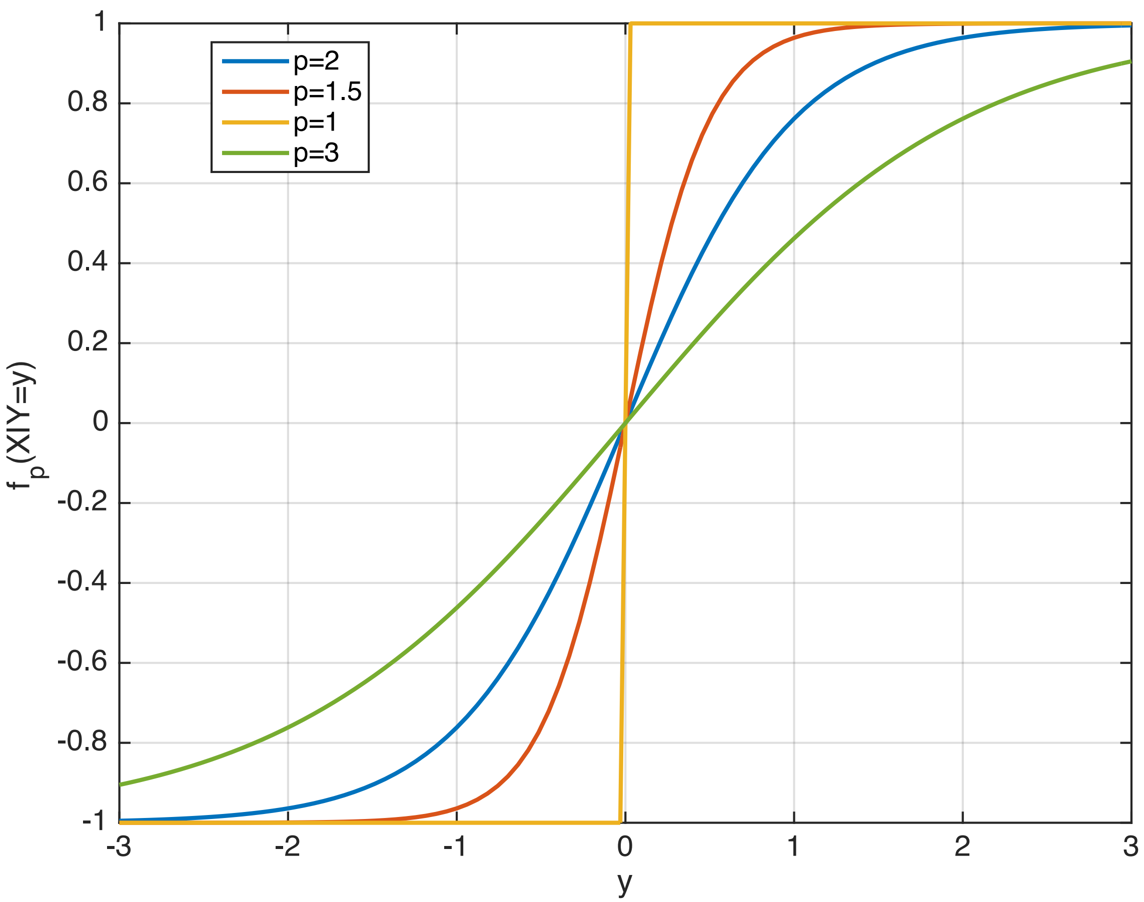
1. if $0 \le X \in \mathbb{R}^1$ then $0 \le f_p(X|Y)$,
2. (Linearity) $f_p(a\mathbf{X} + b|\mathbf{Y}) = a f_p(\mathbf{X}|\mathbf{Y}) + b$,
3. (Stability) $f_p(g(\mathbf{Y})|\mathbf{Y}) = g(\mathbf{Y})$ for any function $g(\cdot)$,
4. (Idempotent) $f_p(f_p(\mathbf{X}|\mathbf{Y})|\mathbf{Y}) = f_p(\mathbf{X}|\mathbf{Y})$,
5. (Degradedness) $f_p\left(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0}, \mathbf{Y}_{\mathrm{snr}}\right) = f_p\left(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0}\right)$, for $\mathbf{X} \to \mathbf{Y}_{\mathrm{snr}_0} \to \mathbf{Y}_{\mathrm{snr}}$,
6. (Shift) $\mathrm{mmpe}(\mathbf{X} + a, \mathrm{snr}, p) = \mathrm{mmpe}(\mathbf{X}, \mathrm{snr}, p)$,
7. (Scaling) $\mathrm{mmpe}(a\mathbf{X}, \mathrm{snr}, p) = a^p \mathrm{mmpe}(\mathbf{X}, a^2\mathrm{snr}, p)$.

$$\cancel{\mathrm{E}\left[f_p\left(X|Y\right)\right] = \mathrm{E}\left[X\right]}$$

PRINCETON UNIVERSITY

Examples of Optimal Estimators

- if $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ then for $p \geq 1$ $f_p(\mathbf{X}|\mathbf{Y}=\mathbf{y}) = \frac{\sqrt{\text{snr}}}{1+\text{snr}}\mathbf{y}$ $\boxed{\text{(i.e., linear);}}$
- if $X = \{\pm 1\}$ equally likely (BPSK) then for $p \geq 1$ $f_p(X|Y) = \tanh\left(\frac{\sqrt{\text{snr}}}{p-1} y\right)$.



Function of p

S. Sherman, "Non-mean-square error criteria," *IRE Transactions on Information Theory*, vol. 3, no. 4, pp. 125–126, 1958.

iTw 2016 CAMB

PRINCETON UNIVERSITY

Bounds on the MMPE

$$\mathrm{mmse}(\mathbf{X}, \mathrm{snr}) \le \min\left(\frac{1}{\mathrm{snr}}, \|\mathbf{X}\|_2^2\right)$$

Theorem. For $\mathrm{snr} \ge 0$, $0 < q \le p$, and input $\mathbf{X}$

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

$$n^{\frac{p}{q}-1}\mathrm{mme}^{\frac{p}{q}}(\mathbf{X}, \mathrm{snr}, q) \le \mathrm{mme}(\mathbf{X}, \mathrm{snr}, p) \le \|\mathbf{X} - \mathrm{E}[\mathbf{X}|\mathbf{Y}]\|_p^p.$$

iTW 2016 CAMBRIDGE

PRINCETON UNIVERSITY

Bounds on the MMPE

Interpolation Bounds

Theorem. For any $0 < p < q < r \leq \infty$ and $\alpha \in (0,1)$ such that

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

where $\bar{\alpha} = 1 - \alpha$,

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

In particular,

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

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

MMPE with Gaussian Input

(Estimation of Gaussian Input.) For $\mathbf{X}_{\mathrm{G}} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ and $p \geq 1$

$$\mathrm{mope}(\mathbf{X}_{\mathrm{G}}, \mathrm{snr}, p) = \frac{\sigma^p \|\mathbf{Z}\|_p^p}{(1 + \mathrm{snr}\sigma^2)^{\frac{p}{2}}},$$

with the optimal estimator given by

$$f_p(\mathbf{X}_{\mathrm{G}} | \mathbf{Y} = \mathbf{y}) = \frac{\sigma^2 \sqrt{\mathrm{snr}}}{1 + \mathrm{snr}\sigma^2} \mathbf{y}.$$

(Asymptotic Optimality of Gaussian Input.) For every $p \geq 1$, and a random variable $\mathbf{X}$ such that $\|\mathbf{X}\|_p^p \leq \sigma^p \|\mathbf{Z}\|_p^p$, we have

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

where

$$\begin{cases} k_{p,\sigma^2 \mathrm{snr}} = 1 & p = 2, \\ 1 \leq k_{p,\mathrm{snr}}^{\frac{1}{p}} = \frac{1+\sqrt{\mathrm{snr}}}{\sqrt{1+\mathrm{snr}}} \leq 1 + \frac{1}{\sqrt{1+\mathrm{snr}}} & p \neq 2. \end{cases}$$

Conditional MMPE

We define the conditional MMPE as

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

Additional Information

(Conditioning Reduces MMPE.) For every $\mathrm{snr} \geq 0$, $p \geq 0$, and random variable $\mathbf{X}$, we have

$$\mathrm{mmpe}(\mathbf{X}, \mathrm{snr}, p) \geq \mathrm{mmpe}(\mathbf{X}, \mathrm{snr}, p|\mathbf{U}).$$

PRINCETON UNIVERSITY

Bounds on the MMPE

(Extra Independent Noisy Observation.) 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

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

Consequences:

- MMPE is decreasing function of SNR
- New Proof of the Single-Crossing-Point Property

MMSE Bounds: Single Crossing Point Property

(**SCPP.**) For any fixed $\mathbf{X}$, suppose that $\mathrm{mmse}(\mathbf{X}, \mathrm{snr}_0) = \frac{\beta}{1+\beta \mathrm{snr}_0}$, for some fixed $\beta \geq 0$. Then

$$\mathrm{mmse}(\mathbf{X}, \mathrm{snr}) \leq \frac{\beta}{1+\beta \mathrm{snr}}, \quad \mathrm{snr} \in [\mathrm{snr}_0, \infty)$$
$$\mathrm{mmse}(\mathbf{X}, \mathrm{snr}) \geq \frac{\beta}{1+\beta \mathrm{snr}}, \quad \mathrm{snr} \in [0, \mathrm{snr}_0).$$

SCPP+I-MMSE important tool for derive converses:

- Version of EPI
- BC
- Wiretap
- MIMO channels

D. Guo, Y. Wu, S. Shamai, and S. Verdu, "Estimation in Gaussian noise: Properties of the minimum mean-square error," IEEE Trans. Inf. Theory, vol. 57, no. 4, pp. 2371–2385, April 2011.

R. Bustin, R. Schaefer, H. Poor, and S. Shamai, "On MMSE properties of optimal codes for the Gaussian wiretap channel," in Proc. IEEE Inf. Theory Workshop, April 2015, pp. 1–5.

R. Bustin, M. Payaro, D. 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.

PRINCETON UNIVERSITY

New Proof of the SCPP

$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathrm{snr}_0, p) = \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \mathrm{snr}_0}$ for some $\beta \geq 0$, $\Leftrightarrow$, $\beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \mathrm{snr}_0 m}$.

Proof: $\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}, p) = \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0 + \Delta, p)$

New Proof of the SCPP

$$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathrm{snr}_0, p) = \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \mathrm{snr}_0} \text{ for some } \beta \geq 0, \Leftrightarrow, \beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \mathrm{snr}_0 m}.$$

Proof:

$$\begin{aligned} \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}, p) &= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0 + \Delta, p) \\ &= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0, p | \mathbf{Y}_\Delta), \text{ where } \mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta \end{aligned}$$

iTw 2016 CAMBRIDGE

PRINCETON UNIVERSITY

New Proof of the SCPP

$$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathrm{snr}_0, p) = \tfrac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \mathrm{snr}_0} \text{ for some } \beta \geq 0, \Leftrightarrow, \beta = \tfrac{m}{\|\mathbf{Z}\|_p^2 - \mathrm{snr}_0 m}.$$

Proof:

$$\begin{aligned} \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}, p) &= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0 + \Delta, p) \\ &= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0, p|\mathbf{Y}_\Delta), \text{ where } \mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta \\ &= \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_\Delta, \mathbf{Y}_{\mathrm{snr}_0})\|_p \end{aligned}$$

New Proof of the SCPP

$$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathrm{snr}_0, p) = \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \mathrm{snr}_0} \text{ for some } \beta \geq 0, \Leftrightarrow, \beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \mathrm{snr}_0 m}.$$

Proof:

$$\begin{aligned}
\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}, p) &= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0 + \Delta, p) \\
&= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0, p | \mathbf{Y}_\Delta), \ \text{where } \mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta \\
&= \|\mathbf{X} - f_p(\mathbf{X} | \mathbf{Y}_\Delta, \mathbf{Y}_{\mathrm{snr}_0})\|_p \\
&\leq \|\mathbf{X} - \hat{\mathbf{X}}\|_p, \ \text{where } \hat{\mathbf{X}} = \frac{(1-\gamma)}{\sqrt{\Delta}} \mathbf{Y}_\Delta + \gamma f_p(\mathbf{X} | \mathbf{Y}_{\mathrm{snr}_0}), \gamma \in [0,1]
\end{aligned}$$

New Proof of the SCPP

$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathrm{snr}_0, p) = \frac{\beta\|\mathbf{Z}\|_p^2}{1+\beta\mathrm{snr}_0}$ for some $\beta \geq 0$, $\Leftrightarrow$, $\beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \mathrm{snr}_0 m}$.

Proof:

$$\begin{aligned}
\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}, p) &= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0 + \Delta, p) \\
&= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0, p|\mathbf{Y}_\Delta), \text{ where } \mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta \\
&= \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_\Delta, \mathbf{Y}_{\mathrm{snr}_0})\|_p \\
&\leq \|\mathbf{X} - \hat{\mathbf{X}}\|_p, \text{ where } \hat{\mathbf{X}} = \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Y}_\Delta + \gamma f_p(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0}), \gamma \in [0,1] \\
&= \left\| \gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Z}_\Delta \right\|_p
\end{aligned}$$

New Proof of the SCPP

$$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathrm{snr}_0, p) = \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \mathrm{snr}_0} \text{ for some } \beta \geq 0, \Leftrightarrow, \beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \mathrm{snr}_0 m}.$$

Proof:

$$\begin{aligned}
\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}, p) &= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0 + \Delta, p) \\
&= \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}_0, p | \mathbf{Y}_\Delta), \text{ where } \mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta \\
&= \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_\Delta, \mathbf{Y}_{\mathrm{snr}_0})\|_p \\
&\leq \|\mathbf{X} - \hat{\mathbf{X}}\|_p, \text{ where } \hat{\mathbf{X}} = \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Y}_\Delta + \gamma f_p(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0}), \gamma \in [0,1] \\
&= \left\| \gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Z}_\Delta \right\|_p \\
&= \frac{\left\| \|\mathbf{Z}\|_p^2(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0})) - \sqrt{\Delta}\cdot m \cdot \mathbf{Z}_\Delta \right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta \cdot m}, \text{ choose } \gamma = \frac{\|\mathbf{Z}\|_p^2}{\|\mathbf{Z}\|_p^2 + \Delta \cdot m}
\end{aligned}$$

New Proof of the SCPP

$m := \text{mmpe}^{\frac{2}{p}}(\mathbf{X}, \text{snr}_0, p) = \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \text{snr}_0}$ for some $\beta \geq 0$, $\Leftrightarrow$, $\beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \text{snr}_0 m}$.

Proof:

$$
\begin{aligned}
\text{mmpe}^{\frac{1}{p}}(\mathbf{X}, \text{snr}, p) &= \text{mmpe}^{\frac{1}{p}}(\mathbf{X}, \text{snr}_0 + \Delta, p) \\
&= \text{mmpe}^{\frac{1}{p}}(\mathbf{X}, \text{snr}_0, p | \mathbf{Y}_\Delta), \ \text{where } \mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta \\
&= \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_\Delta, \mathbf{Y}_{\text{snr}_0})\|_p \\
&\leq \|\mathbf{X} - \hat{\mathbf{X}}\|_p, \ \text{where } \hat{\mathbf{X}} = \frac{(1-\gamma)}{\sqrt{\Delta}} \mathbf{Y}_\Delta + \gamma f_p(\mathbf{X}|\mathbf{Y}_{\text{snr}_0}), \gamma \in [0,1] \\
&= \left\| \gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\text{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}} \mathbf{Z}_\Delta \right\|_p \\
&= \frac{\left\| \|\mathbf{Z}\|_p^2 (\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\text{snr}_0})) - \sqrt{\Delta} \cdot m \cdot \mathbf{Z}_\Delta \right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta \cdot m}, \ \text{choose } \gamma = \frac{\|\mathbf{Z}\|_p^2}{\|\mathbf{Z}\|_p^2 + \Delta \cdot m} \\
&\leq c_p \cdot \frac{\sqrt{m} \|\mathbf{Z}\|_p}{\sqrt{\|\mathbf{Z}\|_p^2 + \Delta \cdot m}}, c_p = \begin{cases} 2 & p \geq 1, \ \text{triangle inequality} \\ 1 & p = 2, \ \text{expand} \end{cases} \\
&= c_p \cdot \frac{\beta \|\mathbf{Z}\|_p^2}{1 + \beta \ \text{snr}}
\end{aligned}
$$

iTW 2016 CAMBRIDGE

PRINCETON UNIVERSITY

New Proof of the SCPP

(**SCPP Bound for MMPE.**) Suppose $\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathrm{snr}_0, p) = \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \mathrm{snr}_0}$ for some $\beta \geq 0$. Then

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

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

Observations:

- Previous proof relied on the knowing the derivative of the MMSE
- New proof relies on simple estimation observations
- Extends to the MMPE for all p>1

Connections to IT

Bounds on Conditional Entropy

$$h(\mathbf{U}|\mathbf{V}) \leq \frac{n}{2} \log(2\pi \mathrm{e}\ \mathrm{mmse}(\mathbf{U}|\mathbf{V})),$$

Theorem. If $\|\mathbf{U}\|_p < \infty$ for some $p \in (0, \infty)$ then for any $\mathbf{V} \in \mathbb{R}^n$, we have

$$h(\mathbf{U}|\mathbf{V}) \leq \frac{n}{2} \log \left(k_{n,2p}^2 \cdot n^{\frac{1}{p}} \cdot \mathrm{mmpe}^{\frac{1}{p}}(\mathbf{U}|\mathbf{V}; p) \right),$$

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

- Sharper version of Ozarow-Wyner bound
- Bounds on the derivative of the MMSE

PRINCETON UNIVERSITY

Ozarow-Wyner Bound

(Ozarow-Wyner Bound.) Let X_D be a discrete random variable then

$$
\begin{aligned}
H(X_D) - \mathrm{gap} \leq I(X_D; Y) &\leq H(X_D), \\
\mathrm{gap} &= \frac{1}{2}\log\left(\frac{\pi e}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12 \cdot \mathrm{lmmse}(X_D|Y)}{d_{\min}^2(X_D)}\right), \\
\mathrm{lmmse}(X_D|Y) &= \frac{\mathbb{E}[X_D^2]}{1 + \mathrm{snr} \cdot \mathbb{E}[X_D^2]} \\
d_{\min}(X_D) &= \inf_{x_i, x_j \in \mathrm{supp}(X_D) : i \neq j} |x_i - x_j|.
\end{aligned}
$$

Generalized Ozarow-Wyner Bound

Let $\mathbf{X}_D$ be a discrete random vector, then for any $p \geq 1$

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

where

$$\mathsf{gap}_p = \inf_{\mathbf{U} \in \mathcal{K}} \mathrm{gap}(\mathbf{U}),$$

$$n^{-1} \cdot \mathrm{gap}(\mathbf{U}) = \log\left(\frac{\|\mathbf{U} + \mathbf{X}_D - f_p(\mathbf{X}_D|\mathbf{Y})\|_p}{\|\mathbf{U}\|_p}\right) + \log\left(\frac{k_{n,p} \cdot n^{\frac{1}{p}} \cdot \|\mathbf{U}\|_p}{\mathrm{e}^{\frac{1}{n} h_e(\mathbf{U})}}\right).$$

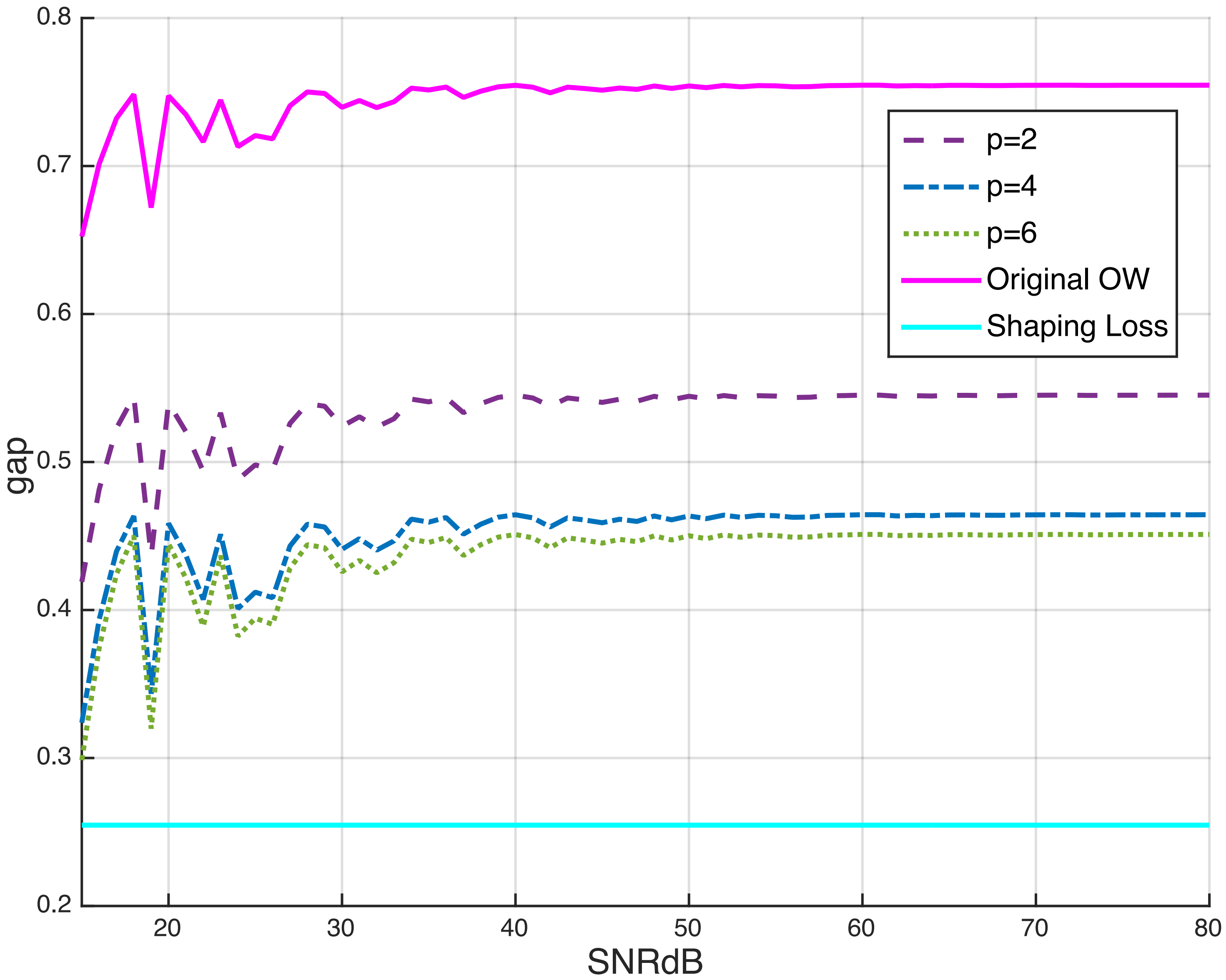
where $\mathcal{K} = \{\mathbf{U} : \mathrm{diam}(\mathrm{supp}(\mathbf{U})) \leq 2 \cdot d_{\min}(\mathbf{X}_D)\}$.
Moreover,

$$\log\left(\frac{\|\mathbf{U} + \mathbf{X}_D - f_p(\mathbf{X}_D|\mathbf{Y})\|_p}{\|\mathbf{U}\|_p}\right) \overset{\text{for } p \geq 1}{\leq} \log\left(1 + \frac{\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathrm{snr}, p)}{\|\mathbf{U}\|_p}\right).$$

Generalized Ozarow-Wyner Bound

Take X_D to be PAM with $N = \sqrt{1+\text{snr}}$ then

$$\frac{1}{2}\log(1+\text{snr}) - I(X_D;Y) = \text{gap}$$



PRINCETON UNIVERSITY

Concluding Remarks

- Properties of the MMPE functional
- Conditional MMPE
- New Proof of the SCPP and its extension
- Bounds on differential entropy
- Generalized Ozarow-Wyner bound

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/pdf/1603.07628, 2016.

Thank you

arXiv:1607.01461

adytso@princeton.edu

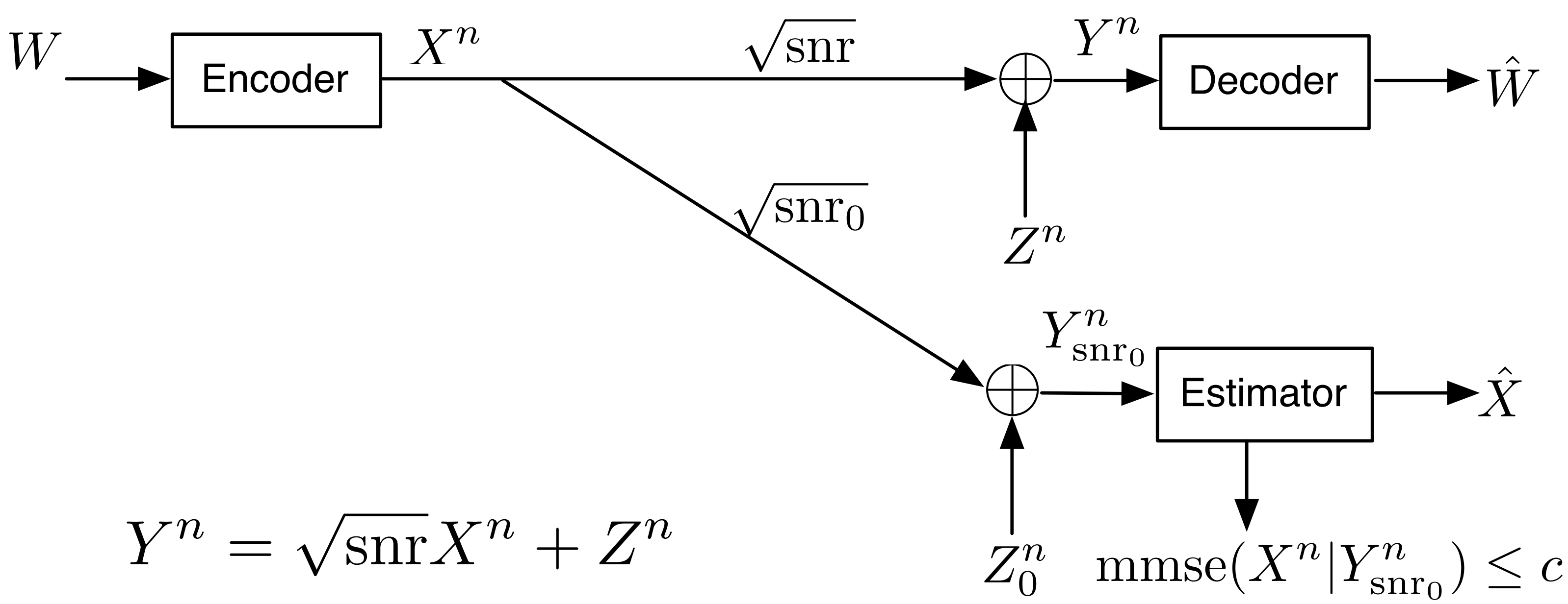
PRINCETON UNIVERSITY

Back Up Slides

arXiv:1607.01461

adytso@princeton.edu

Problem



$$Y^n = \sqrt{\mathrm{snr}}X^n + Z^n$$
$$Y^n_{\mathrm{snr}_0} = \sqrt{\mathrm{snr}_0}X^n + Z^n_0$$
$$\text{where } Z^n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), Z^n_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

max-I problem

$$\mathcal{C}_n(\mathrm{snr}, \mathrm{snr}_0, \beta) := \sup_{X^n} \frac{1}{n} I(X^n, Y^n_{\mathrm{snr}}),$$
$$\text{s.t.} \operatorname{Tr}\left(\mathbb{E}\left[X^n (X^n)^{\mathrm{T}}\right]\right) \leq n,$$
$$\text{and } \mathrm{mmse}(X^n, \mathrm{snr}_0) \leq \frac{\beta}{1 + \beta\mathrm{snr}_0}.$$

iTw 2016 CAMBRIDGE

PRINCETON UNIVERSITY

max-MMSE problem

max-MMSE problem

$$
\begin{aligned}
&\mathrm{M}_n(\mathrm{snr}, \mathrm{snr}_0, \beta) := \sup_{\mathbf{X}} \mathrm{mmse}(\mathbf{X}, \mathrm{snr}),\\
&\text{s.t.} \;\; \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}}]\right) \le 1,\\
&\text{and } \; \mathrm{mmse}(\mathbf{X}, \mathrm{snr}_0) \le \frac{\beta}{1+\beta\mathrm{snr}_0}.
\end{aligned}
$$

Relationship to max-I problem

$$
\frac{\mathrm{snr}}{2}\mathrm{M}_n(\mathrm{snr}, \mathrm{snr}_0, \beta) \le \mathcal{C}_n(\mathrm{snr}, \mathrm{snr}_0, \beta) \le \frac{1}{2}\int_0^{\mathrm{snr}} \mathrm{M}_n(t, \mathrm{snr}_0, \beta) dt.
$$