Overview and New Advances on the Interplay Between Estimation and Information Measures

Alex Dytso
Hosted by: Salim El Rouayheb

ILLINOIS INSTITUTE OF TECHNOLOGY
UIC UNIVERSITY OF ILLINOIS AT CHICAGO Department of Electrical and Computer Engineering COLLEGE OF ENGINEERING
PRINCETON UNIVERSITY

Outline

- Review of Estimation Theory and Information Theory Results
- Applications of the I-MMSE
- New Directions

ILLINOIS INSTITUTE OF TECHNOLOGY

Notation

Channel Model

For scalar channels

$$Y = \sqrt{\text{snr}}X + Z$$

where $X \in \mathbb{R}$ is independent of $Z \in \mathbb{R}$ and $Z \sim \mathcal{N}(0,1)$.

For vector channels

$$\mathbf{Y} = \sqrt{\text{snr}}\mathbf{X} + \mathbf{Z}$$

where $\mathbf{X} \in \mathbb{R}^n$ is independent of $\mathbf{Z} \in \mathbb{R}^n$ and $Z \sim \mathcal{N}(0,\mathbf{I})$.

Main Functionals

$$I(\mathbf{X},\text{snr}) := I(\mathbf{X};\mathbf{Y}) = \mathbb{E}\left[\log\left(\frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{Y})}\right)\right]$$

$$\text{mmse}(\mathbf{X},\text{snr}) := \text{mmse}(\mathbf{X}|\mathbf{Y}) = \mathbb{E}\left[\left(\mathbf{X} - \mathbb{E}\left[\mathbf{X}|\mathbf{Y}\right]\right)^T\left(\mathbf{X} - \mathbb{E}\left[\mathbf{X}|\mathbf{Y}\right]\right)\right]$$

I-MMSE

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

D. Guo, S. Shamai, and S. Verdu´, ``The Interplay Between Information and Estimation Measures''. now Publishers Incorporated, 2013.

S. Shamai, "From constrained signaling to network interference alignment via an information-estimation perspective," IEEE Information Theory Society Newsletter, vol. 62, no. 7, pp. 6–24, September 2012.

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.

Part 1(a)

Estimation Theory

ILLINOIS INSTITUTE OF TECHNOLOGY

Conditional Expectation

Definition

$$\mathbb{E}\left[X\right] = \int X dP_X,$$
$$\mathbb{E}\left[X|Y = y\right] = \int X dP_{X|Y=y}.$$

Properties

1. if $0 \leq X \in \mathbb{R}^1$ then $0 \leq \mathrm{E}[X|Y]$,
2. (Linearity) $\mathrm{E}[a\mathbf{X}_1 + b\mathbf{X}_2|\mathbf{Y}] = a\mathrm{E}[\mathbf{X}_1|\mathbf{Y}] + b\mathrm{E}[\mathbf{X}_2|\mathbf{Y}]$,
3. (Total Expectation) $\mathrm{E}[\mathrm{E}[\mathbf{X}|\mathbf{Y}]] = \mathrm{E}[\mathbf{X}]$,
4. (Stability) $\mathrm{E}[g(\mathbf{Y})|\mathbf{Y}] = g(\mathbf{Y})$ for any function $g(\cdot)$,
5. (Idempotent) $\mathrm{E}[\mathrm{E}[\mathbf{X}|\mathbf{Y}]|\mathbf{Y}] = \mathrm{E}[\mathbf{X}|\mathbf{Y}]$,
6. (Product Rule) $\mathrm{E}[g(\mathbf{Y})\mathbf{X}|\mathbf{Y}]] = g(\mathbf{Y})\mathrm{E}[\mathbf{X}|\mathbf{Y}]]$,
7. (Degradedness) $\mathrm{E}\left[\mathbf{X}|\mathbf{Y}_{\mathrm{snr}}, \mathbf{Y}_{\mathrm{snr}_0}\right] = \mathrm{E}\left[\mathbf{X}|\mathbf{Y}_{\mathrm{snr}_0}\right]$,
   for $\mathbf{X} \to \mathbf{Y}_{\mathrm{snr}_0} \to \mathbf{Y}_{\mathrm{snr}}$,

There are more properties.

ILLINOIS INSTITUTE OF TECHNOLOGY

Orthogonality Principle

(Orthogonality Principle.) For all $g(Y) \in L^2$

$$\mathbb{E}[(X - \mathbb{E}[X|Y])g(Y)] = 0.$$

Proof:

$$\begin{aligned}
\mathbb{E}[(X - E[X|Y])g(Y)] &= \mathbb{E}[g(Y)X] - \mathbb{E}[g(Y)\mathbb{E}[X|Y]], \text{ linearity} \\
&= \mathbb{E}[g(Y)X] - \mathbb{E}[\mathbb{E}[g(Y)X|Y]], \text{ product rule} \\
&= \mathbb{E}[g(Y)X] - \mathbb{E}[g(Y)X], \text{ law of total expectation} \\
&= 0
\end{aligned}$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Minimum Mean Square Error

Minimum Mean Square Error (MMSE)

(MMSE.) For any $X \in L^2$ then

$$\inf_{f(Y) \in L^2} \mathbb{E}[(X - f(Y))^2] = \mathbb{E}[(X - \mathbb{E}[X|Y])^2]$$

Proof:

$$\begin{aligned}
\mathbb{E}[(X - f(Y))^2] &= \mathbb{E}[(X - \mathbb{E}[X|Y] + \mathbb{E}[X|Y] - f(Y))^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X|Y])^2] \\
&+ 2\mathbb{E}[(X - \mathbb{E}[X|Y])(\mathbb{E}[X|Y] - f(Y))] \\
&+ \mathbb{E}[(\mathbb{E}[X|Y] - f(Y))^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X|Y])^2] + \mathbb{E}[(\mathbb{E}[X|Y] - f(Y))^2]
\end{aligned}$$

General Cost Function

More Generally:

$$\inf_{f(Y)\in\mathcal{F}} \mathbb{E}\left[\mathsf{Err}(X, f(Y))\right]$$

Naturally, to say more we have to make assumptions on either Err, $\mathcal{F}$, (X,Y).

ILLINOIS INSTITUTE OF TECHNOLOGY

General Cost Function

More Generally:

$$\inf_{f(Y) \in \mathcal{F}} \mathbb{E}\left[\mathsf{Err}(X, f(Y))\right]$$

We will look at the class of cost function given by:

$$\mathsf{Err}(X, f(Y)) = |X - f(Y)|^p, \ p > 0.$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Optimality of Linear Estimation

Linear estimator: $f(Y) = aY + b$

Linear Minimum Mean Square Error (LMMSE)

(LMMSE.) If $Y = \sqrt{\text{snr}}X + Z$, $\mathbb{E}[X^2] = 1$ and $\mathbb{E}[X] = 0$ then

$$\text{lmmse}(X|Y) = \inf_{f(Y):f(Y)=aY+b} \mathbb{E}[(X - f(Y))^2] = \frac{1}{1+\text{snr}},$$

where optimal coefficient are given by

$$a = \frac{\sqrt{\text{snr}}}{1+\text{snr}}, \ \ b = 0.$$

ILLINOIS INSTITUTE OF TECHNOLOGY

When is linear estimation optimal?

(Gaussian is the 'hardest' to estimate.)

$$\mathrm{mmse}(X|Y) \leq \mathrm{lmmse}(X|Y),$$

with equality iff X is Gaussian.

ILLINOIS INSTITUTE OF TECHNOLOGY

When is linear estimation optimal?

(LMMSE Optimality for General Cost Function.) Let X be Gaussian and

$$e = X - f(Y), \text{ s.t.} f(Y) = aY + b,$$
$$\mathsf{E}rr(X, f(Y)) = g(e),$$

where the function g satisfies

$$0 \leq g(e), \text{ non-negative},$$
$$g(e) = g(-e), \text{ symmetric},$$
$$0 \leq e_1 \leq e_2, \Rightarrow g(e_1) \leq g(e_2), \text{ increasing}.$$

Then

$$\mathsf{arg} \inf_{f(Y): f(Y) = aY+b} \mathbb{E}[\mathsf{Err}(X, f(Y))] = \mathsf{arg} \inf_{f(Y): f(Y) = aY+b} \mathbb{E}[(X - f(Y))^2].$$

In other words, the optimal estimator is the LMMSE estimator.

ILLINOIS INSTITUTE OF TECHNOLOGY

When is linear estimation optimal?

(LMMSE Optimality for General Cost Function.) Let X be Gaussian and

$$e = X - f(Y), \text{ s.t.} f(Y) = aY + b,$$
$$\mathsf{E}rr(X, f(Y)) = g(e),$$

where the function g satisfies

$$0 \leq g(e), \text{ non-negative},$$
$$g(e) = g(-e), \text{ symmetric},$$
$$0 \leq e_1 \leq e_2, \Rightarrow g(e_1) \leq g(e_2), \text{ increasing}.$$

Then

$$\mathsf{arg} \inf_{f(Y): f(Y) = aY+b} \mathbb{E}[\mathsf{Err}(X, f(Y))] = \mathsf{arg} \inf_{f(Y): f(Y) = aY+b} \mathbb{E}[(X - f(Y))^2].$$

In other words, the optimal estimator is the LMMSE estimator.

Very general class

ILLINOIS INSTITUTE OF TECHNOLOGY

When is linear estimation optimal?

Proof:

J. Brown, “Asymmetric non-mean-square error criteria,” *IRE Transactions on Automatic Control*, vol. 7, no. 1, pp. 64–66, Jan 1962.

Let

$$\sigma^2 = \mathbb{E}[e^2] = \mathbb{E}[(X - f(Y))^2].$$

Now using properties of $g(\cdot)$ we have that

$$\begin{aligned}
\mathbb{E}[\mathsf{Err}(X, f(Y))] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) e^{-\frac{x^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\sigma t) e^{-\frac{t^2}{2}}\, dt \\
&= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} g(\sigma t) e^{-\frac{t^2}{2}}\, dt.
\end{aligned}$$

So, if $\sigma_1 \leq \sigma_2$ then

$$\frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} g(\sigma_1 t) e^{-\frac{t^2}{2}}\, dt \leq \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} g(\sigma_2 t) e^{-\frac{t^2}{2}}\, dt.$$

ILLINOIS INSTITUTE OF TECHNOLOGY

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

When is linear estimation optimal?

Proof:

J. Brown, "Asymmetric non-mean-square error criteria," *IRE Transactions on Automatic Control*, vol. 7, no. 1, pp. 64–66, Jan 1962.

Let

$$\sigma^2 = \mathbb{E}[e^2] = \mathbb{E}[(X - f(Y))^2].$$

Now using properties of $g(\cdot)$ we have that

$$\begin{aligned}\mathbb{E}[\mathsf{Err}(X, f(Y))] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) e^{-\frac{x^2}{2\sigma^2}}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\sigma t) e^{-\frac{t^2}{2}}\, dt \\ &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} g(\sigma t) e^{-\frac{t^2}{2}}\, dt.\end{aligned}$$

So, if $\sigma_1 \leq \sigma_2$ then

$$\frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} g(\sigma_1 t) e^{-\frac{t^2}{2}}\, dt \leq \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} g(\sigma_2 t) e^{-\frac{t^2}{2}}\, dt.$$

Minimized by LMMSE optimal estimator

ILLINOIS INSTITUTE OF TECHNOLOGY

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

Part 1(b)

Information Theory

ILLINOIS INSTITUTE OF TECHNOLOGY

Entropy Power Inequality (EPI)

(Entropy Power Inequality.) Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent continuous random vectors then

$$e^{\frac{2}{n}h(\mathbf{X}_1+\mathbf{X}_2)} \geq e^{\frac{2}{n}h(\mathbf{X}_1)} + e^{\frac{2}{n}h(\mathbf{X}_2)}.$$

EPI is equivalent to Lieb Inequality.

(Lieb Inequality.)

$$h(\sqrt{\alpha}\mathbf{X}_1 + \sqrt{1-\alpha}\mathbf{X}_2) \geq \alpha h(\mathbf{X}_1) + (1-\alpha)h(\mathbf{X}_2),$$

for all $\alpha \in [0, 1]$.

C.Shannon,"A mathematical theory of communication,"*BellSyst.Tech.J.*,vol.27,no. 379-423,623-656,Jul.,Oct.1948.

A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. Conor.*, vol.2, pp. 101-112, 1959.

Capacity of the Gaussian Wiretap Channel

Let

$$Y = \sqrt{\text{snr}_1}X + Z_1,$$
$$Y_e = \sqrt{\text{snr}_2}X + Z_2,$$

the the secrecy capacity is given by

$$C = \max_{\mathbb{E}[X^2]\leq 1} (I(X;Y) - I(X;Y_e)) = \frac{1}{2}\log\left(\frac{1+\text{snr}_1}{1+\text{snr}_2}\right)$$

Proof uses EPI.

ILLINOIS INSTITUTE OF TECHNOLOGY

A. El Gamal and Y.-H. Kim, *Network Information Theory*. Camrbidge, U.K.: Camrbidge Univ. Press, 2012.

Capacity scalar Gaussian BC

(Scalar Gaussian BC.) Let

$$Y_1 = \sqrt{\mathrm{snr}_1} X + Z_1$$
$$Y_2 = \sqrt{\mathrm{snr}_2} X + Z_2$$

then

$$C = \bigcup_{P_{UX}:\mathbb{E}[X^2]\le 1, U-X-(Y_1,Y_2)} \left\{ \begin{array}{ll} R_1 & \le I(X;Y_1|U) \\ R_2 & \le I(U;Y_2) \end{array} \right\}$$
$$= \bigcup_{\alpha\in[0,1]} \left\{ \begin{array}{ll} R_1 & \le \frac{1}{2}\log(1+\alpha\mathrm{snr}_1) \\ R_2 & \le \frac{1}{2}\log\left(\frac{1+\mathrm{snr}_2}{1+\alpha\mathrm{snr}_2}\right) \end{array} \right\}$$

where capacity achieving distribution P_{UX} is Gaussian with $\mathbb{E}[UX] = \sqrt{1-\alpha}$.

Classical proof is by using EPI.

ILLINOIS INSTITUTE OF TECHNOLOGY

A. El Gamal and Y.-H. Kim, *Network Information Theory*. Camrbidge, U.K.: Camrbidge Univ. Press, 2012.

Part 2

Applications

Differential Entropy

(Differential Entropy.)

$$h(X) = \frac{1}{2}\int_0^{\infty} \mathrm{mmse}(X,t) - \frac{1}{2\pi e + t}dt$$

Proof:

$$\begin{aligned} h(X + aZ) &= I(X; X + aZ) + h(aZ) \\ &= \frac{1}{2}\int_0^{a^{-2}} \mathrm{mmse}(X,t)dt + \frac{1}{2}\log(2\pi e a^2) \\ &= \frac{1}{2}\int_0^{a^{-2}} \mathrm{mmse}(X,t) - \frac{1}{2\pi e + t}dt \end{aligned}$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Proof of EPI

(Entropy Power Inequality.) Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent continuous random vectors then

$$e^{\frac{2}{n}h(\mathbf{X}_1+\mathbf{X}_2)} \geq e^{\frac{2}{n}h(\mathbf{X}_1)} + e^{\frac{2}{n}h(\mathbf{X}_2)}.$$

EPI is equivalent to Lieb Inequality.

(Lieb Inequality.)

$$h(\sqrt{\alpha}\mathbf{X}_1 + \sqrt{1-\alpha}\mathbf{X}_2) \geq \alpha h(\mathbf{X}_1) + (1-\alpha)h(\mathbf{X}_2),$$

for all $\alpha \in [0,1]$.

We prove Lieb Inequality

ILLINOIS INSTITUTE OF TECHNOLOGY

Proof of EPI

(Lieb Inequality.)

$$h(\sqrt{\alpha}\mathbf{X}_1 + \sqrt{1-\alpha}\mathbf{X}_2) \geq \alpha h(\mathbf{X}_1) + (1-\alpha)h(\mathbf{X}_2),$$

for all $\alpha \in [0,1]$.

(MMSE version of Lieb Inequality.) For any independent vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ and $\alpha \in [0,1]$,

$$\mathrm{mmse}(\sqrt{\alpha}\mathbf{X}_1 + \sqrt{1-\alpha}\mathbf{X}_2, \mathrm{snr}) \geq (1-\alpha)\mathrm{mmse}(\mathbf{X}_1, \mathrm{snr}) + \alpha\mathrm{mmse}(\mathbf{X}_2, \mathrm{snr})$$

Proof: Integrate the MMSE version to get the entropy version.

Proof of EPI

(MMSE version of Lieb Inequality.) For any independent vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ and $\alpha \in [0,1]$,

$$\mathrm{mmse}(\sqrt{1-\alpha}\mathbf{X}_1 + \sqrt{\alpha}\mathbf{X}_2, \mathrm{snr}) \geq (1-\alpha)\mathrm{mmse}(\mathbf{X}_1, \mathrm{snr}) + \alpha \mathrm{mmse}(\mathbf{X}_2, \mathrm{snr})$$

Proof:

Define

D. Guo, S. Shamai, and S. Verdu´, ``The Interplay Between Information and Estimation Measures". now Publishers Incorporated, 2013.

$$\begin{aligned}
\mathbf{X} &= \sqrt{1-\alpha}\mathbf{X}_1 + \sqrt{\alpha}\mathbf{X}_2, \\
\mathbf{Y}_1 &= \sqrt{\mathrm{snr}}\mathbf{X}_1 + \mathbf{Z}_1, \\
\mathbf{Y}_2 &= \sqrt{\mathrm{snr}}\mathbf{X}_2 + \mathbf{Z}_2, \\
\mathbf{Y} &= \mathbf{Y}_1\sqrt{1-\alpha} + \mathbf{Y}_2\sqrt{\alpha} = \sqrt{\mathrm{snr}}\mathbf{X} + (\sqrt{1-\alpha}\mathbf{Z}_1 + \sqrt{\alpha}\mathbf{Z}_2)
\end{aligned}$$

where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent.

$$\begin{aligned}
\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) &= \mathbb{E}[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|_2] \\
&\geq \mathbb{E}[\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}_1, \mathbf{Y}_2]\|_2] \\
&= (1-\alpha)\mathbb{E}[\|\mathbf{X}_1 - \mathbb{E}[\mathbf{X}_1|\mathbf{Y}_1]\|_2] + \alpha\mathbb{E}[\|\mathbf{X}_2 - \mathbb{E}[\mathbf{X}_2|\mathbf{Y}_2]\|_2] \\
&= (1-\alpha)\mathrm{mmse}(\mathbf{X}_1, \mathrm{snr}) + \alpha\mathrm{mmse}(\mathbf{X}_2, \mathrm{snr})
\end{aligned}$$

Capacity of the Gaussian Wiretap Channel

Let

$$
\begin{aligned}
Y &= \sqrt{\mathrm{snr}_1} X + Z_1, \\
Y_e &= \sqrt{\mathrm{snr}_2} X + Z_2,
\end{aligned}
$$

the the secrecy capacity is given by

$$
C = \max_{\mathbb{E}[X^2] \le 1} \left(I(X;Y) - I(X;Y_e) \right).
$$

The capacity achieving distribution is Gaussian. The classical proof uses EPI.

Simpler Proof:

$$
\begin{aligned}
I(X;Y) - I(X;Y_e) &= \frac{1}{2} \int_{\mathrm{snr}_2}^{\mathrm{snr}_1} \mathrm{mmse}(X,t)\,dt \\
&\le \frac{1}{2} \int_{\mathrm{snr}_2}^{\mathrm{snr}_1} \mathrm{lmmse}(X,t)\,dt \\
&= I(X_G;Y) - I(X_G;Y_e)
\end{aligned}
$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Single-Crossing Point Property (SCPP)

(SCPP.) For every given random variable X (which is not necessarily unit variance), the curve of $\mathrm{mmse}(X,\gamma)$ crosss the curve of $\frac{1}{1+\gamma}$ which is the MMSE of standard Gaussian distribution, at most once on $\gamma \in (0,\infty)$. Precisely, denote their difference as

$$f(\gamma) = \frac{1}{1+\gamma} - \mathrm{mmse}(X,\gamma).$$

Then

1. $f(\gamma)$ is strictly increasing at every γ with $f(\gamma) < 0$;
2. if $f(\mathrm{snr}_0) = 0$, then $f(\gamma) \geq 0$ at every $\gamma > \mathrm{snr}_0$;
3. $\lim_{\gamma\to 0} f(\gamma) = 0$.

ILLINOIS INSTITUTE OF TECHNOLOGY

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

Capacity scalar Gaussian BC

(Scalar Gaussian BC.) Let

$$
\begin{aligned}
Y_1 &= \sqrt{\mathrm{snr}_1} X + Z_1 \\
Y_2 &= \sqrt{\mathrm{snr}_2} X + Z_2
\end{aligned}
$$

then

$$
\begin{aligned}
C &= \bigcup_{P_{UX}:\mathbb{E}[X^2]\leq 1, U-X-(Y_1,Y_2)} \left\{ \begin{array}{ll} R_1 & \leq I(X;Y_1|U) \\ R_2 & \leq I(U;Y_2) \end{array} \right\} \\
&= \bigcup_{\alpha\in[0,1]} \left\{ \begin{array}{ll} R_1 & \leq \frac{1}{2}\log(1+\alpha \mathrm{snr}_1) \\ R_2 & \leq \frac{1}{2}\log\left(\frac{1+\mathrm{snr}_2}{1+\alpha \mathrm{snr}_2}\right) \end{array} \right\}
\end{aligned}
$$

where capacity achieving distribution P_{UX} is Gaussian with $\mathbb{E}[UX] = \sqrt{1-\alpha}$.

Classical proof is by using EPI.

ILLINOIS INSTITUTE OF TECHNOLOGY

Capacity scalar Gaussian BC

Bound on R2

Since $I(X;Y_2|U) \leq I(X:Y_2) \leq \frac{1}{2}\log(1+\text{snr}_2)$ then there exists some $\alpha \in [0,1]$ and

$$I(X;Y_2|U) = \frac{1}{2}\log(1+\alpha\text{snr}_2).$$

So,

$$\begin{aligned}
R_2 = I(U;Y_2) &= I(X,U;Y_2) - I(X:Y_2|U) \\
&= I(X;Y_2) - I(X;Y_2|U) \\
&\leq \frac{1}{2}\log(1+\text{snr}_2) - \frac{1}{2}\log(1+\alpha\text{snr}_2) \\
&= \frac{1}{2}\log\left(\frac{1+\text{snr}_2}{1+\alpha\text{snr}_2}\right)
\end{aligned}$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Capacity scalar Gaussian BC

Bound on R1

We know that there exists $\alpha \in [0,1]$ such that

$$I(X;Y_2|U) = \frac{1}{2}\log(1+\alpha\mathrm{snr}_2) = \frac{1}{2}\int_0^{\mathrm{snr}_2} \frac{\alpha}{1+\alpha t}dt,$$

No the other hand,

$$I(X;Y_2|U) = \frac{1}{2}\int_0^{\mathrm{snr}_2} \mathrm{mmse}(X_U, t|U)dt,$$

So, there must exists $\mathrm{snr}_0 \leq \mathrm{snr}_2$ such that

$$\mathrm{mmse}(X_U, \mathrm{snr}_0|U) = \frac{\alpha}{1+\alpha\mathrm{snr}_0}.$$

So, by SCPP for all $t \geq \mathrm{snr}_2 \geq \mathrm{snr}_0$ we have that

$$\mathrm{mmse}(X_U, t|U) \leq \frac{\alpha}{1+\alpha t}.$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Capacity scalar Gaussian BC

Bound on R1

We know that there exists $\alpha \in [0,1]$ such that

$$I(X;Y_2|U) = \frac{1}{2}\log(1+\alpha\text{snr}_2).$$

For $t \geq \text{snr}_2 \geq \text{snr}_0$ we have that

$$\text{mmse}(X_U, t|U) \leq \frac{\alpha}{1+\alpha t}.$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Capacity scalar Gaussian BC

Bound on R1

We know that there exists $\alpha \in [0,1]$ such that

$$I(X;Y_2|U) = \frac{1}{2}\log(1+\alpha\text{snr}_2).$$

For $t \geq \text{snr}_2 \geq \text{snr}_0$ we have that

$$\text{mmse}(X_U, t|U) \leq \frac{\alpha}{1+\alpha t}.$$

$$\begin{aligned}
I(X;Y_1|U) &= \frac{1}{2}\int_0^{\text{snr}_1} \text{mmse}(X_U, t|U)dt \\
&= \frac{1}{2}\int_0^{\text{snr}_2} \text{mmse}(X_U, t|U)dt + \frac{1}{2}\int_{\text{snr}_2}^{\text{snr}_1} \text{mmse}(X_U, t|U)dt \\
&= \frac{1}{2}\log(1+\alpha\text{snr}_2) + \frac{1}{2}\int_{\text{snr}_2}^{\text{snr}_1} \text{mmse}(X_U, t|U)dt \\
&\leq \frac{1}{2}\log(1+\alpha\text{snr}_2) + \frac{1}{2}\int_{\text{snr}_2}^{\text{snr}_1} \frac{\alpha}{1+\alpha t}dt \\
&= \frac{1}{2}\log(1+\alpha\text{snr}_1)
\end{aligned}$$

ILLINOIS INSTITUTE OF TECHNOLOGY

Part 3

Some New Results

ILLINOIS INSTITUTE OF TECHNOLOGY

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)

UIC Department of Electrical and Computer Engineering
UNIVERSITY OF ILLINOIS AT CHICAGO
COLLEGE OF ENGINEERING

PRINCETON UNIVERSITY

TECHNION
Israel Institute of Technology

MMPE

The p-norm of a random vector $\mathbf{U} \in \mathbb{R}^n$ is give 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: $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

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

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

ILLINOIS INSTITUTE OF TECHNOLOGY

MMPE

We define the minimum mean p-th error (MMPE) of estimating $\mathbf{X}$ from $\mathbf{Y}$ as

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

We donate the optimal estimator by $\boxed{f_p(\mathbf{X}|\mathbf{Y}).}$
Example: for $p = 2$

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

Study Properties

ILLINOIS INSTITUTE OF TECHNOLOGY

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 $g(\cdot)$.

ILLINOIS INSTITUTE OF TECHNOLOGY

Properties of Optimal Estimators

E[X|Y]

1. if $0 \leq X \in \mathbb{R}^1$ then $0 \leq \mathrm{E}[X|Y]$,
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)

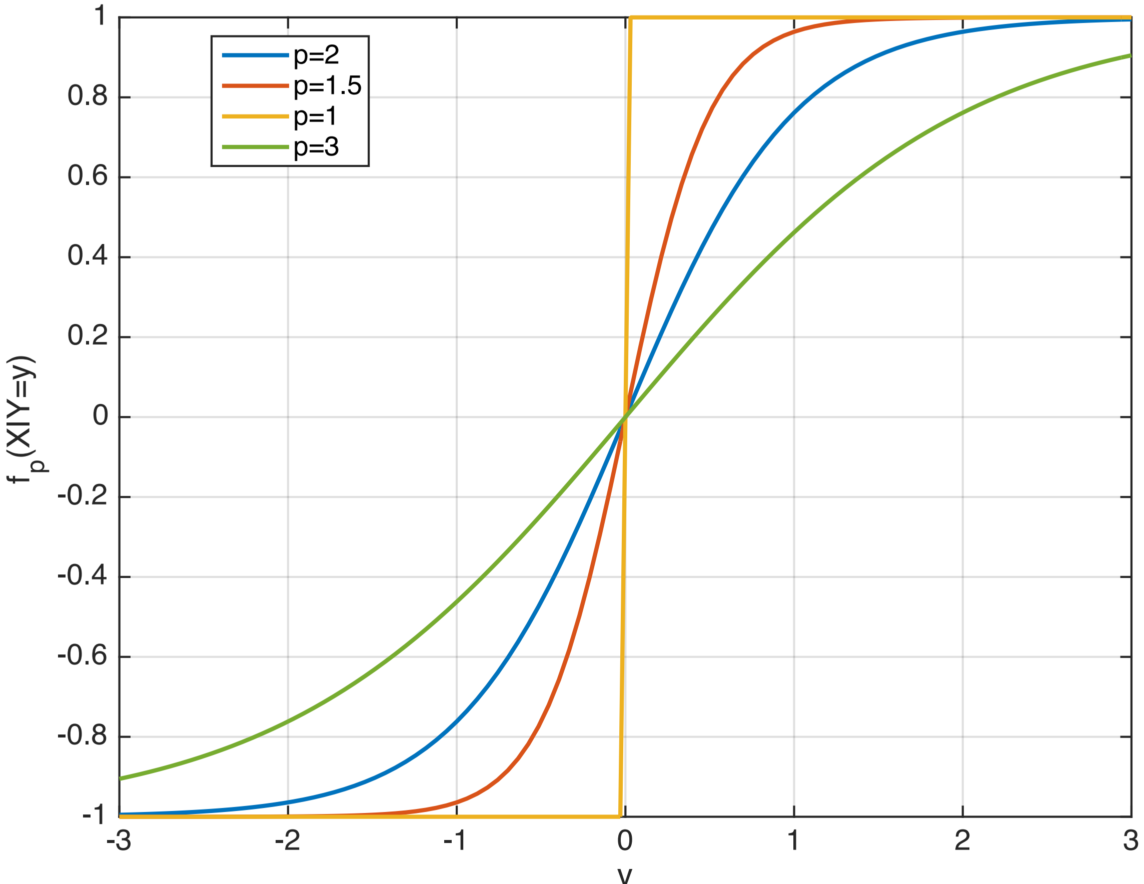
1. if $0 \leq X \in \mathbb{R}^1$ then $0 \leq f_p(X|Y)$,
2. (Linearity) $f_p(a\mathbf{X} + b|\mathbf{Y}) = af_p(\mathbf{X}|\mathbf{Y}) + b$,
3. (Stability) $f_p(g(\mathbf{Y})|\mathbf{Y}) = g(\mathbf{Y})$ for any function $g(\cdot)$,
4. (Idempotent) $f_p(f_p(\mathbf{X}|\mathbf{Y})|\mathbf{Y}) = f_p(\mathbf{X}|\mathbf{Y})$,
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) $\mathrm{mmpe}(a\mathbf{X}, \mathrm{snr}, p) = a^p \mathrm{mmpe}(\mathbf{X}, a^2\mathrm{snr}, p)$.

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

ILLINOIS INSTITUTE OF TECHNOLOGY

Examples of Optimal Estimators

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



ILLINOIS INSTITUTE OF TECHNOLOGY

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

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 $\mathbf{X}$

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

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

ILLINOIS INSTITUTE OF TECHNOLOGY

Bounds on The MMPE

Interpolation Bounds

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

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

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,

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

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

ILLINOIS INSTITUTE OF TECHNOLOGY

Estimation of Gaussian Inputs

(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 snr ≥ 0, $p \geq 0$, and random variable $\mathbf{X}$, we have

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

ILLINOIS INSTITUTE OF TECHNOLOGY

Conditional MMPE (Con't)

(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

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

Consequences:

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

ILLINOIS INSTITUTE OF TECHNOLOGY

New Proof SCPP

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

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

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

Observations:

- Previous proof relied on the notion of the derivative of the MMSE
- New proof relies on simple estimation observations
- Extends to the MMPE

ILLINOIS INSTITUTE OF TECHNOLOGY

New Proof 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 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:

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

New Proof 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:

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

ILLINOIS INSTITUTE OF TECHNOLOGY

New Proof 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:

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

ILLINOIS INSTITUTE OF TECHNOLOGY

New Proof 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:

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

ILLINOIS INSTITUTE OF TECHNOLOGY

New Proof 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:

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

ILLINOIS INSTITUTE OF TECHNOLOGY

New Proof SCPP

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

Proof:

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

ILLINOIS INSTITUTE OF TECHNOLOGY

SCPP

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.

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

- Constant is sharp
- Sharper version of Ozarow-Wyner Bound

ILLINOIS INSTITUTE OF TECHNOLOGY

OW Bound

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

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

Generalized OW 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

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

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

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

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

ILLINOIS INSTITUTE OF TECHNOLOGY

Generalized OW Bound

Proof:

$$\begin{aligned} I(\mathbf{X}_D, \mathbf{Y}) &\geq I(\mathbf{X}_D + \mathbf{U}, \mathbf{Y}), \text{ data processing} \\ &= h(\mathbf{X}_D + \mathbf{U}) - h(\mathbf{X}_D + \mathbf{U}|\mathbf{Y}) \\ &= H(\mathbf{X}_D) + h(\mathbf{U}) - h(\mathbf{X}_D + \mathbf{U}|\mathbf{Y}), \text{ compact support of } \mathbf{U} \text{ ,} \end{aligned}$$

Next, observe

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

By combining all the bounds

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

ILLINOIS INSTITUTE OF TECHNOLOGY

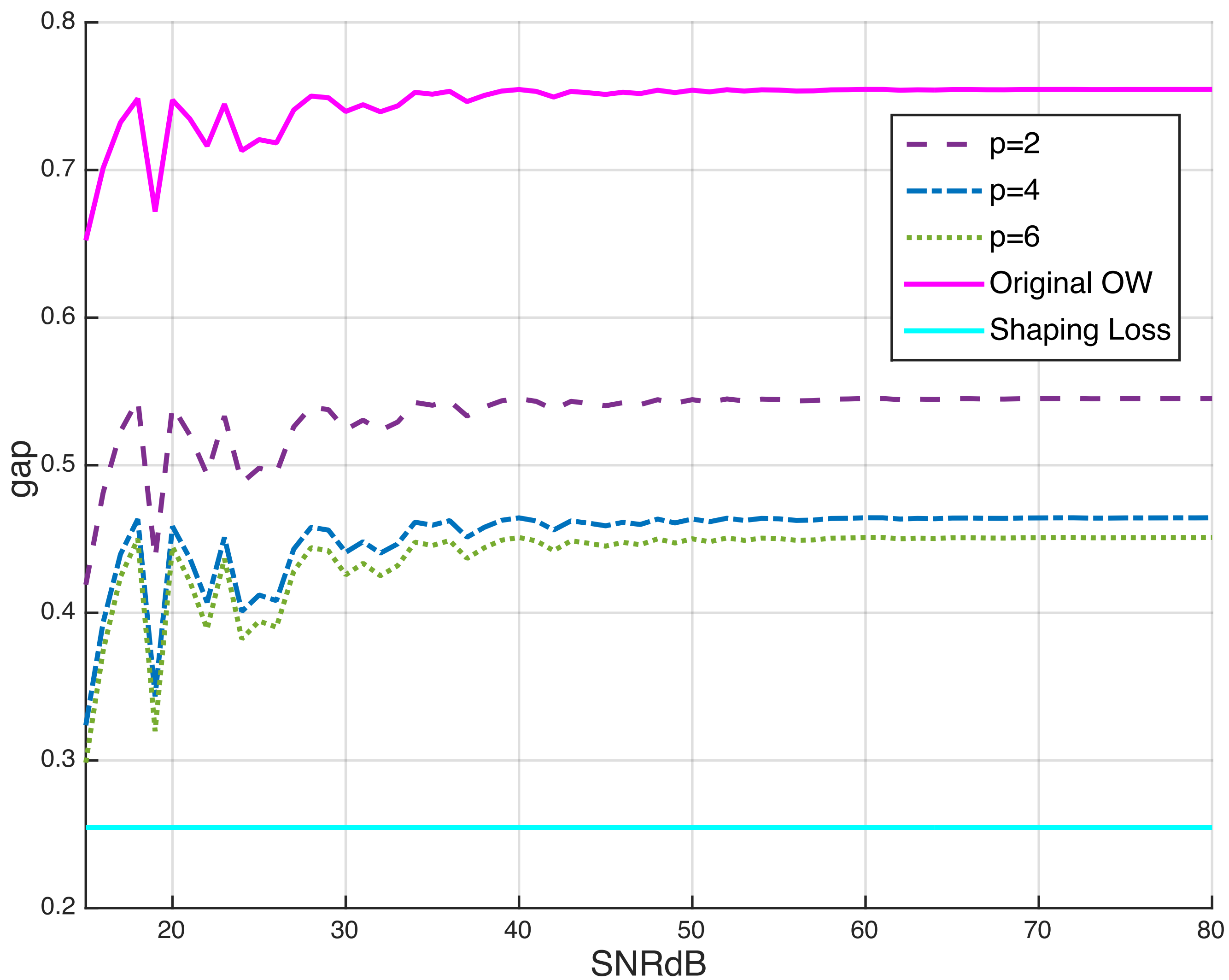
Generalized OW Bound

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

$$H(X_D) \approx \frac{1}{2}\log(1+\text{snr})$$

and we are interested in

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



ILLINOIS INSTITUTE OF TECHNOLOGY

Concluding Remarks

- Properties of the MMPE functional
- Conditional MMPE
- New Proof of the SCPP and it's extension
- Bounds on differential entropy
- Generalized OW bound

Open Problems

- Generalizations to vector inputs
- Generalization to general channel matrices
- Non-Gaussian Channels
- Statistical Physics Applications

Thank you

arXiv:1607.01461

odytso2@uic.edu

ILLINOIS INSTITUTE OF TECHNOLOGY

Abstract:

In this talk, we overview advances on the subject of the interplay between estimation and information measures. Recently, this area has received considerable attention due to the seminal work of Guo-Verdu-Shamai that established the so-called I-MMSE relationship. The I-MMSE relationship shows that the derivative of mutual information with respect to signal-to-noise ratio (SNR) is given by one-half times Minimum Means Square Error (MMSE).

In the first part of the talk, we focus on the necessary estimation and information theoretic background. Specifically, we will review estimation theoretical notions such as orthogonality principle, optimal estimation for a given cost function, properties of conditional expectation and optimality of linear estimators. For the information theoretic concepts, we will review basic properties of the mutual information, entropy power inequality (EPI) and converses of scalar broadcast and wiretap channels.

In the second part of the talk, we will focus on the I-MMSE relationship and its applications. We will demonstrate simple, few line, proof of the EPI. Next, we will review an estimation concept of Single-Crossing Point Property (SCPP), which is a powerful tool in showing information theoretic converses for additive white Gaussian noise channels. Next, by using SCPP and I-MMSE, we will demonstrate alternative proofs of the converses for scalar broadcast and wiretap channels.

In the third part of the talk, we discuss some more recent research results. One such area is estimation under the higher orders errors. Specifically, we study the notion of the Minimum Mean p-th Error (MMPE) of which the MMSE is the special case. We derive and discuss several bounds and properties of the MMPE. We also focus on properties of the MMPE optimal estimators in terms of input distribution, such as linearity, stability, degradedness, average bias, etc.

Next, we discuss applications of the MMPE functional. As the first application, the notion of the MMPE is applied to derive bounds on the conditional differential entropy as well as to generalize the Ozarow-Wyner mutual information lower bound for a discrete input on AWGN channel. As the second application, we show that the MMPE can be used to bound the MMSE of finite length codes improving on previous characterizations of the phase transition phenomenon for capacity-achieving codes of infinite length. Interestingly, this leads to some non-trivial results in statistical physics.

The final part of the presentation will focus on some future application and major open problems.

This work is in collaboration with H. Vincent Poor, Ronit Bustin, Daniela Tuninetti, Natasha Devroye and Shlomo Shamai.

ILLINOIS INSTITUTE OF TECHNOLOGY