I-MMSE Relationship

Recall that the capacity of the point-to-point channel $p_{Y|X}$ is

$$C = \max_{p_X} I(X;Y)$$

Suppose we have an additive white Gaussian noise (AWGN) channel

$Y = \sqrt{SNR}\, X + Z$, where Z is a standard Gaussian noise and

there is also an input power constraint $E[X^2] \leq 1$.

Then the capacity is

$$C = \frac{1}{2} \log(1 + SNR).$$

The converse part of the AWGN capacity is via a maximum entropy argument.

Consider slightly more general setting with noise variance N and power constraint P.

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z), \quad \text{since } Z \text{ independent of } X. \end{aligned}$$

Now $h(Z) = \frac{1}{2}\log 2\pi e N$. Also, we have

$$E[Y^2] = E[(X+Z)^2] = E[X^2] + 2E[X]E[Z] + E[Z^2] = P + N,$$

since $X \perp Z$ and $E[Z] = 0$.

Since $E[Y^2] = P+N$, the entropy of Y is bounded (by the entropy-power inequality) by $\frac{1}{2}\log(2\pi e(P+N))$, achieved by Gaussian,

$$\begin{aligned} I(X;Y) &= h(Y) - h(Z) \\ &\leq \frac{1}{2}\log(2\pi e(P+N)) - \frac{1}{2}\log(2\pi e N) \\ &= \frac{1}{2}\log\left(1 + \frac{P}{N}\right) = \frac{1}{2}\log(1 + SNR). \end{aligned}$$

Minimum mean-square estimation is a standard problem in estimation theory.

$E[X|Y]$ is conditional mean estimate of X given Y.

Corresponding conditional variance is

$$\mathrm{var}[X|Y] = E[(X - E[X|Y])^2 \mid Y].$$

Since it is well-known the optimal MMSE estimator is the conditional mean, it follows the estimation error is just the average conditional variance:

$$\mathrm{mmse}(X|Y) = E[\mathrm{var}[X|Y]].$$

when $Y = \sqrt{SNR}\, X + Z$ where $Z \sim \mathcal{N}(0,1)$, we get

$$\mathrm{mmse}(X, SNR) = \mathrm{mmse}(X \mid \sqrt{SNR}\, X + Z)$$
$$= E[(X - E[X \mid \sqrt{SNR}\, X + Z])^2].$$

For a Gaussian input with mean m and variance σ_X^2 denoted $X \sim \mathcal{N}(m, \sigma_X^2)$

$$\mathrm{mmse}(X, SNR) = \frac{\sigma_X^2}{1 + \sigma_X^2\, SNR}.$$

Notice that if $\sigma_X^2 = 1$, we can directly see a relationship between

$I(SNR) = \frac{1}{2}\log(1 + SNR)$ and $\mathrm{mmse}(SNR) = \frac{1}{1 + SNR}$

namely $\frac{d}{dSNR} I(SNR) = \frac{1}{2}\,\mathrm{mmse}(SNR)$.

What is striking is that Guo, Shamai, and Verdu (2005) showed

$\frac{d}{dSNR} I(SNR) = \frac{1}{2}\,\mathrm{mmse}(SNR)$ relationship holds not just

for Gaussian inputs, but for any input ~~[illegible]~~ distribution.

Integral form: $I(X; Y_{snr}) = \frac{1}{2}\int_0^{SNR} \mathrm{mmse}(\gamma)\, d\gamma$

The I-MMSE relationship can be proved using an "incremental channel" approach which gauges the decrease in mutual information from infinitesimally small additional Gaussian noise.

→ mutual information is essentially half the SNR in the vicinity of zero SNR but insensitive to the shape of the input distribution otherwise.

The I-MMSE relation and similar information-estimation identities are expectation identities

Venkat and Weissman (2012) explicitly characterize the random quantities involved in a pointwise sense. Then they draw on Ito calculus to characterize their properties.

Recall the mutual information can be expressed in terms of the expected value of information density:

$$I(X;Y) = E\left[\log \frac{dP_{Y|X}}{dP_Y}\right]$$

$i(X,Y)$.

Consider additive Gaussian channel at signal-to-noise γ:

$$Y_\gamma = \gamma X + W_\gamma$$

for $\gamma \in [0, \text{SNR}]$ where W is Brownian motion independent of X, where $W_\gamma \sim N(0,\gamma)$.

I-MMSE can be expressed as

$$E\left[\log \frac{dP_{Y_{\text{snr}}|X}}{dP_{Y_{\text{snr}}}}\right] = E\left[\frac{1}{2}\int_0^{\text{snr}} (X - E[X|Y_\gamma])^2 d\gamma\right].$$

So I-MMSE can be restated as

$$E[Z]=0 \quad \text{where} \quad Z = \log \frac{dP_{Y_{snr}|X}}{dP_{Y_{snr}}} - \frac{1}{2}\int_0^{snr} (X-E[X|Y_\gamma])^2\, d\gamma$$

is a tracking error between information density and half squared error

integrated over snr.

One can characterize this Z further beyond just zero-mean:

Thm $\mathrm{var}(Z) = \int_0^{snr} \mathrm{mmse}(\gamma)\, d\gamma = 2I(X;Y_{snr})$.

Information-Theoretic Diffusion [Kong, Brekelmans, and Ver Steeg, 2023].

Diffusion models can be thought of / analyzed as denoising autoencoders, VAEs, or through stochastic differential equations (SDEs).

SDEs are fairly complicated mathematically: is there an information-theoretic approach?

- Adding noise is like communication
- Removing noise is like MMSE estimation.

↳ called denoising function, so optimal denoising is conditional expectation.

Develop another pointwise generalization of the I-MMSE relation

(Use machine learning notation).

$$z_\gamma = \sqrt{\gamma}\, x + \epsilon \quad \text{where } \epsilon \sim \mathcal{N}(0,1)$$ where γ is SNR and $p(x)$ is input distribution.

Thm $\frac{d}{d\gamma} D_{KL}\left[p(z_\gamma \mid x) \,\|\, p(z_\gamma) \right] = \frac{1}{2} \, \text{mmse}(x, \gamma).$

where marginal $p(z_\gamma) = \int p(z_\gamma \mid x)\, p(x)\, dx$ and pointwise MMSE is

$$\text{mmse}(x, \gamma) = \mathbb{E}_{p(z_\gamma \mid x)} \left[\| x - \hat{x}^*(z_\gamma, \gamma) \|_2^2 \right].$$