Additive Gaussian Chanel

- Recal that the entropy of a continuous random variable x with pdf $p(x)$ was

$$H(X) = -\int_{-\infty}^{+\infty} f(x) \log f(x)\, dx$$

- We can show that the normal r.v. maximizes the entropy among all densities with finite variance

Maximize $-\int_{-\infty}^{+\infty} f(x) \log f(x)$ under condition $\int_{-\infty}^{+\infty} x^2 f(x)\, dx = \sigma^2 < \infty$

Using variation calculus — we can assume the $\overline{X} = 0$ $\int_{-\infty}^{+\infty} x f(x)\, dx = 0$

$$-\frac{\partial}{f} f \log f + \lambda_1 \frac{\partial}{\partial f} f + \lambda_2 \frac{\partial}{\partial f} (x^2 f) = 0$$

$$-\log_2 e + \lambda_1 + \lambda_2 x^2 = \log_2 f = \log_2 e \cdot \ln f$$

$$f = e^{-1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2}{\log_2 e} x^2}$$

$$\lambda_1 = \frac{1}{2} \log_2 \left(\frac{e^2}{2\pi\sigma^2} \right) \qquad \lambda_2 = -\frac{\log_2 e}{2\sigma^2}$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

$$H(X)_{max} = \frac{1}{2} \log_2 (2\pi e \sigma^2)$$

(another proof: $-\int_{-\infty}^{+\infty} f(x) \log f(x)\, dx \leq -\int_{-\infty}^{\infty} f(x) \log g(x)\, dx$

common reference, the information will be the same as the difference between the corresponding differential entropy terms. We are therefore perfectly justified in using the term $h(X)$, defined in Eq. 2.68, as the differential entropy of the continuous random variable X.

When we have a continuous random vector $\mathbf{X}$ consisting of n random variables $X_1, X_2, \ldots, X_n$, we define the differential entropy of $\mathbf{X}$ as the *n-fold integral*

$$h(\mathbf{X}) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) \log_2 \left[\frac{1}{f_{\mathbf{X}}(\mathbf{x})}\right] d\mathbf{x} \tag{2.70}$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the *joint probability density function* of $\mathbf{X}$.

EXAMPLE 8 MAXIMUM DIFFERENTIAL ENTROPY FOR SPECIFIED VARIANCE

In this example, we solve an important *constrained optimization problem*. We determine the form that the probability density function of a random variable X must have for the differential entropy of X to assume its largest value for some prescribed variance. In mathematical terms, we may restate the problem as follows:

With the differential entropy of a random variable X defined by

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x) dx,$$

find the probability density function $f_X(x)$ for which $h(X)$ is maximum, subject to the two constraints

$$\int_{-\infty}^{\infty} f_x(x) dx = 1 \tag{2.71}$$

and

$$\int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)\, dx = \sigma^2 = \text{constant} \tag{2.72}$$

where μ is the mean of X and σ^2 is its variance.

The formula for $h(X)$ is that of Eq. 2.68, reproduced here with a minor modification. The first constraint, Eq. 2.71 simply states that the area under $f_X(x)$, a probability density function, must equal unity. The second constraint, Eq. 2.72, recognizes that the variance of X has a prescribed value. The second constraint is significant, because σ^2 is a measure of average power, and so maximization of the differential entropy $h(X)$ is performed subject to a constraint of constant power. The result of this constrained optimization will be exploited later on in Section 2.9.

We use the *method of Lagrange multipliers** to solve this constrained optimization problem. Specifically, the differential entropy $h(X)$ will attain its maximum value only when the integral

* The method of Lagrange multipliers is described in Kaplan (1952, pp. 128–130).

$$\int_{-\infty}^{\infty} [-f_X(x) \log_2 f_X(x) + \lambda_1 f_X(x) + \lambda_2 (x - \mu)^2 f_X(x)] dx$$

is *stationary*. The parameters λ_1 and λ_2 are known as *Lagrange multipliers*. That is to say, $h(X)$ is maximum only when the derivative of the integrand

$$-f_X(x) \log_2 f_X(x) + \lambda_1 f_X(x) + \lambda_2 (x - \mu)^2 f_X(x)$$

with respect to $f_X(x)$ is zero. This yields the result

$$\begin{aligned} -\log_2 e + \lambda_1 + \lambda_2 (x - \mu)^2 &= \log_2 f_X(x) \\ &= (\log_2 e) \ln f_X(x) \end{aligned}$$

where e is the base of the natural logarithm. Solving for $f_X(x)$, we get

$$f_X(x) = \exp\left[-1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2}{\log_2 e} (x - \mu)^2\right] \tag{2.73}$$

Note that λ_2 has to be negative if the integrals of $f_X(x)$ and $(x - \mu)^2 f_X(x)$ with respect to x are to converge. Substituting Eq. 2.73 in Eqs. 2.71 and 2.72, and then solving for λ_1 and λ_2, we get

$$\lambda_1 = \frac{1}{2} \log_2 \left(\frac{e}{2\pi\sigma^2}\right)$$

and

$$\lambda_2 = -\frac{\log_2 e}{2\sigma^2}$$

The desired form for $f_X(x)$ is therefore described by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \tag{2.74}$$

which is recognized as the probability density of a *Gaussian random variable X of mean μ and variance σ^2*. The maximum value of the differential entropy of such a random variable is obtained by substituting Eq. 2.74 in Eq. 2.68. The result of this substitution is given by

$$h(X) = \frac{1}{2} \log_2 (2\pi e \sigma^2) \tag{2.75}$$

We may thus summarize the results of this example, as follows:

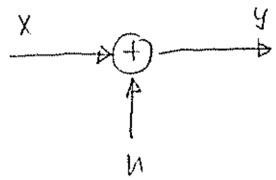
1. *For a given variance σ^2, the Gaussian random variable has the largest differential entropy attainable by any random variable.* That is, if X is a Gaussian random variable and Y is any other random variable with the same mean and variance, then for all Y

$$h(X) \geq h(Y) \tag{2.76}$$

   where the equality holds if, and only if, $Y = X$.

2. *The entropy of a Gaussian random variable X is uniquely determined by the variance of X* (i.e., it is independent of the mean of X).

Indeed, it is because of Property 1 that the Gaussian channel model is so widely used in the study of digital communication systems.



$$I(X;Y) \triangleq H(Y) - H(Y|X) \qquad\qquad n \quad N(0,\sigma_n^2)$$

$$C \triangleq \max_{f(x)} I(X;Y)$$

$$H(Y|X) = \frac{1}{2}\log(2\pi e\sigma_n^2) \quad = H(N)$$

$$\underset{\max}{H}(Y) = \frac{1}{2}\log(2\pi e(\sigma_x^2+\sigma_n^2))$$ — maximized when is gaussian

$$C = \frac{1}{2}\log(2\pi e(\sigma_x^2+\sigma_n^2)) - \frac{1}{2}\log(2\pi e\sigma_n^2) = \frac{1}{2}\log\left(1+\frac{\sigma_x^2}{\sigma_n^2}\right)$$

$$H(Y|X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x)\cdot p(y|x)\,\log\frac{1}{p(y|x)}\,dx\,dy$$

$$= -\int_{-\infty}^{+\infty} p(x)\int_{-\infty}^{+\infty}\frac{e^{-\frac{1}{2\sigma_n^2}(y-x)^2}}{\sqrt{2\pi\sigma^2}}\,\log\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma_n^2}(y-x)^2}\,dy$$

$$H(Y|X) = H(X+N|X) = H(N)$$

Information Measures for Continuous Sources

$$H(X) = -\int_{-\infty}^{+\infty} p(x) \log p(x)\, dx$$

$$H(X|Y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y) \log p(x|y)\, dx\, dy$$

$$I(X;Y) = H(X) - H(X|Y)$$

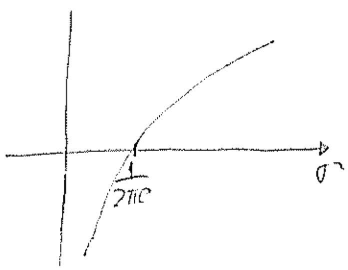
$$I(X;Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Example: $$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

$$H(X) = -\int_{-\infty}^{+\infty} p(x)\left(\log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{x^2}{2\sigma^2}\frac{1}{\ln 2}\right) dx$$

$$= +\frac{1}{2}\log 2\pi\sigma^2 + \frac{1}{2\sigma^2}\frac{\log e}{\log 2}\underbrace{\int_{-\infty}^{+\infty} p(x)\, x^2 dx}_{\sigma^2}$$

$$= \frac{1}{2}\log 2\pi e\sigma^2$$



Example $y = x + n$

$x: N(0, \sigma_x^2)$; $n: N(0, \sigma_n^2)$

$y: N(0, \sigma_x^2 + \sigma_n^2)$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$H(Y|X) = H(N) = \frac{1}{2}\log 2\pi e\sigma_n^2$$

$$I(X;Y) = \frac{1}{2}\log\left(1 + \frac{\sigma_x^2}{\sigma_n^2}\right)$$

Input is a descrete:

$$I(X;Y) = \sum_{i=1}^{m} P(x_i) \cdot \int_{-\infty}^{+\infty} p(y|x_i) \log \frac{p(y|x_i)}{p(y)}$$

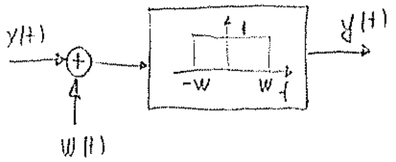
$$p(y) = \sum_{i=1}^{m} P(x_i) P(y|x_i)$$

Example

$X = \{-A, A\}$ $\quad Y \in R$ $\quad y = x + n$

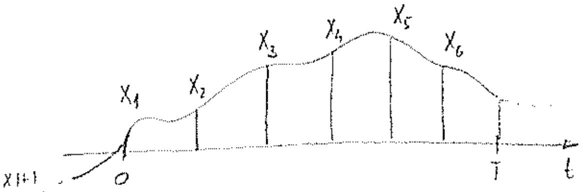
$$C = \frac{1}{2}\int_{-\infty}^{+\infty} p(y|A) \log \frac{p(y|A)}{p(y)} dy + \frac{1}{2}\int_{-\infty}^{+\infty} p(y|-A) \log \frac{p(y|-A)}{p(y)} dy$$

Waveform Channels



$$y(t) = x(t) + n(t)$$

$x(t), y(t), n(t)$ — band-limited signals



$\{\phi_i(t)\}_{1\le i<\infty}$ — set of orthogonal functions on $[0,T]$

$$\langle \phi_i(t), \phi_j(t)\rangle = \int_0^T \phi_i(t)\phi_j^*(t)\,dt = \delta_{ij}$$

$$x(t) = \sum_i x_i \phi_i(t)$$

$$y(t) = \sum_i y_i \phi_i(t)$$

$$n(t) = \sum_i n_i \phi_i(t)$$

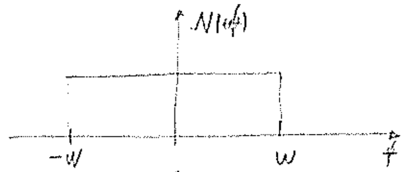
Котельников – Nyquist theorem
Shannon

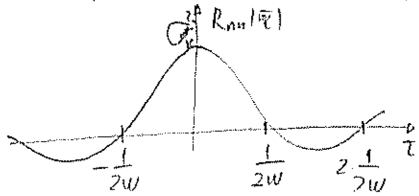
$$x(t) = \sum_i x\left(i\frac{1}{2W}\right)\frac{\sin 2\pi W\left(t-\frac{i}{2W}\right)}{2\pi W\left(t-\frac{i}{2W}\right)}$$

$$x_i = x\left(\frac{i}{2W}\right)$$

$$E\left(\left(x(t) - \sum_i x_i\phi_i(t)\right)^2\right) = 0$$

$$y_i = x_i + n_i$$





$$R_{nn}\left(i\frac{1}{2W}\right) = 0$$

$$p(n_i, n_j) = p(n_i)\,p(n_j)$$

$$p(y_1, y_2, \dots, y_N | x_1, x_2, \dots, x_N) = \prod_{i=1}^{N} p(y_i|x_i)$$

$$p(y_i|x_i) = p(n_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y_i-x_i)^2}{2\sigma_n^2}}$$

$$I(X_N;Y_N) = H(Y_N) - H(X_N|Y_N)$$

$$= \sum_{i=1}^{N} H(Y_i) - \sum_{i=1}^{N} H(N_i)$$

$$= N\left(H(Y_1) - H(N_1)\right)$$

$$= N\left(\frac{1}{2}\log 2\pi e(\sigma_x^2+\sigma_n^2) - \frac{1}{2}\log 2\pi e(\sigma_x^2\sigma_n^2)\right)$$

$$= N\frac{1}{2}\log\left(1+\frac{\sigma_x^2}{\sigma_n^2}\right)$$

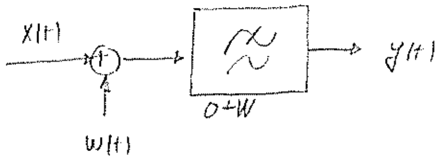
$$S = \frac{1}{T}\int_0^T E(x^2(t))\,dt = \frac{1}{T}\sum_{i=1}^{N} E(x_i^2) = \frac{1}{T}N\sigma_x^2$$

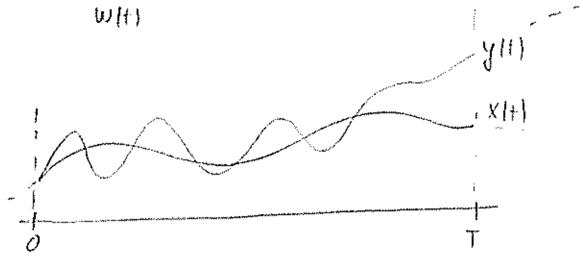
$$\sigma_x^2 = \frac{TS}{N} \qquad T = N\frac{1}{2W} \qquad N = 2WT$$

$$I(X_N, Y_N) = WT\log\left(1+\frac{TS}{N\cdot\frac{N_0}{2}}\right)$$

$W\cdot T$

Capacity of a Waveform Channel





$x(t) = \sum_i x_i \phi_i(t)$

$x_i = \langle x(t), \phi_i(t) \rangle = \int_0^T x(t)\phi_i^*(t)dt$

$\{\phi_i(t)\}$ – set of orthonormal functions

$\langle \phi_i(t) \cdot \phi_j(t) \rangle = \delta_{ij}$

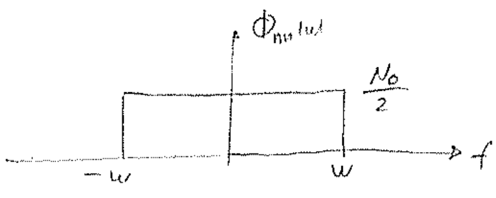
Котельников – Nyquist – Shannon theorem

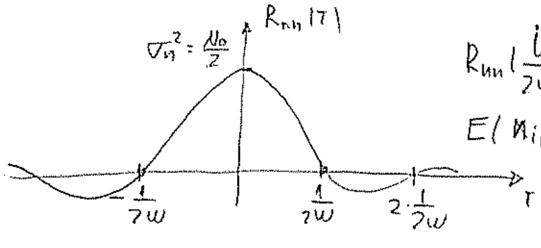
$$E\left(x(t) - \sum_{i=0}^{N} x\left(\frac{i}{2w}\right) \frac{\sin 2\pi w(t - \frac{i}{2w})}{2\pi w(t - \frac{i}{2w})} \right) = 0$$

$$N = \frac{T}{\frac{1}{2w}} = T \cdot 2W$$

$x_i = x\left(i\frac{1}{2w}\right) \qquad n_i = n\left(i\frac{1}{2w}\right)$

$y_i = y\left(i\frac{1}{2w}\right)$

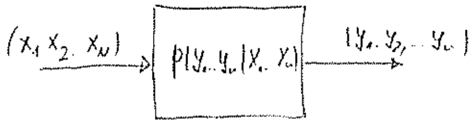




$R_{nn}\left(\frac{i}{2w}\right) = 0$

$E(n_i n_j) = 0$

$n_i, n_j \quad i \neq j$ are independent



$$p(y_1, y_2, \ldots y_N | x_1, x_2 \ldots x_N) = \prod_{i=1}^{N} p(y_i | x_i)$$

$$p(y_i | x_i) = p(n_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y_i - x_i)^2}{2\sigma^2}}$$

$$I(X_N ; Y_N) = H(Y_N) - H(Y_N | X_N)$$
$$= H(Y_N) - H(N_N)$$

x_i – Gaussian $\Rightarrow$ $I(X_N, Y_N)$ is maximized

$$H(N_N) = \sum_{i=1}^{N} H(N_i) = N \cdot \frac{1}{2} \log 2\pi e \sigma_n^2$$

$$H(Y_N) = \sum_{i=1}^{N} H(Y_i) = N \frac{1}{2} \log 2\pi e (\sigma_x^2 + \sigma_n^2)$$

$$I^*(X_N; Y_N) = N \cdot \frac{1}{2} \log\left(1 + \frac{\sigma_x^2}{\sigma_n^2}\right)$$

$$\sigma_n^2 = \frac{N_0}{2}$$

$$\sigma_x^2 = E(x_i^2) \qquad S = \frac{1}{T}\int_0^T E(x^2(t))dt = \frac{1}{T}\sum_{i=1}^{N} E(x_i^2)$$

$$S = \frac{1}{T} \cdot N \cdot \sigma_x^2$$

$$I^*(X_N; Y_N) = T \cdot W \log\left(1 + \frac{ST}{\frac{N \cdot N_0}{2}}\right)$$

$$= TW \log\left(1 + \frac{ST}{TW N_0}\right)$$

$$= TW \log\left(1 + \frac{S}{N_0 W}\right)$$

$$C = \lim_{T\to\infty} \frac{1}{T} I^{*}(X_N; Y_N) = W \log\left(1+\frac{S}{N_0 W}\right)$$

$$S = \mathcal{E}_b \cdot C$$

$$\frac{C}{W} = \log\left(1+\frac{\mathcal{E}_b C}{N_0 W}\right) = \frac{\log_e\left(1+\frac{\mathcal{E}_b}{N_0}\frac{C}{W}\right)}{\log_e 2}$$

$$e^{\frac{C}{W}\ln 2} = 1 + \frac{\mathcal{E}_b}{N_0}\frac{C}{W}$$

$$\frac{\mathcal{E}_b}{N_0} = \frac{e^{\frac{C}{W}\ln 2} - 1}{\frac{C}{W}}$$

Shannon - channel coding theorem.

For $R<C$ it is possible to communicate reliably through a channel with arbitrary low probability of error

$R>C$ the same code would produce error with probability 1

$$C = W\frac{\ln\left(1+\frac{S}{N_0 W}\right)}{\ln 2} \le W\frac{1}{\ln 2}\frac{S}{N_0 W} = \frac{S}{N_0 \ln 2}$$

$$\lim_{\frac{C}{W}\to 0}\frac{\mathcal{E}_b}{N_0} = \frac{\sum_{i=0}^{+\infty}\left(\frac{C}{W}\ln 2\right)^i\frac{1}{i!} - 1}{\frac{C}{W}} = \ln 2 = -1.6\,\text{dB}$$

$$\frac{C}{W} = 1 \Rightarrow \frac{\mathcal{E}_b}{N_0} = 1$$