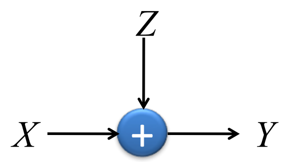
EE 231A Information Theory
Lecture 12:

Gaussian Channel Results
and
Sufficient Statistics

A. Binary Input Gaussian Channel
B. Capacity of a Bandlimited Gaussian Channel
C. Capacity of Parallel Gaussian Channels
D. Sufficient Statistics

Part 12 A:
Binary Input Gaussian Channel

Gaussian Channel



- $Z \sim \mathrm{N}(0, N)$
- Z independent of X
- Power constraint on X: $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \le P$ or $EX^2 \le P$
- In general, X is continuous.

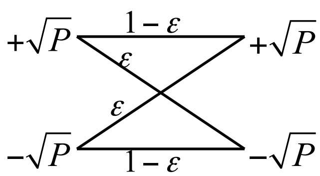
BSC from Gaussian Channel

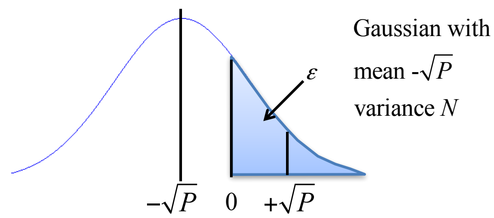
- Now suppose we construct a BSC from the Gaussian channel.
- $X = \pm\sqrt{P}$ which guarantees $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \le P.$

Quantization

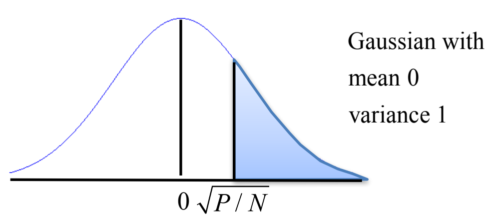
- Quantize the Gaussian channel so that

$$Y' = \operatorname{sgn}(Y)\sqrt{P}$$





Q-function



- Q-function
  - tail of a unit variance zero mean Gaussian.
  - $Q(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$
  - $\varepsilon = Q(\sqrt{P/N})$

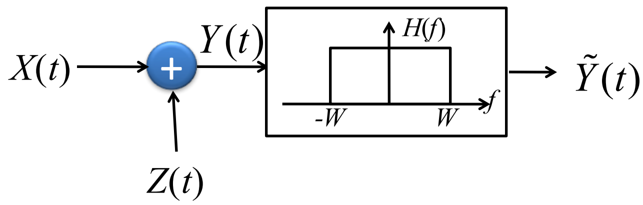
Limitation of Binary Input

- Recall capacity is $1\text{-}H(\varepsilon)$ for BSC, so this use of the Gaussian channel will never support more than 1 bit per channel use.

Part 12 B: Capacity of a Bandlimited Gaussian Channel

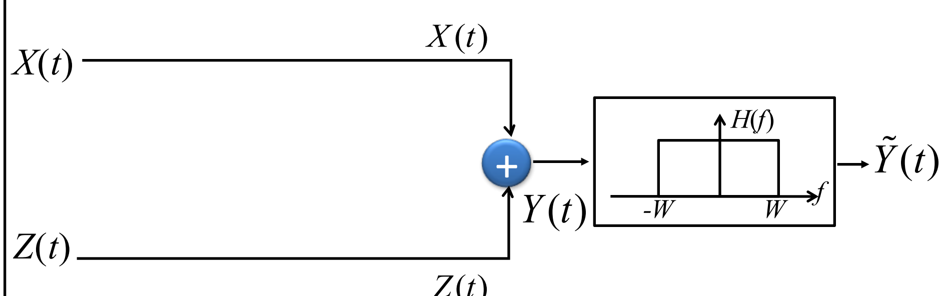
We have extended our discussion from discrete to continuous alphabet. Now we extend from discrete to continuous time.

Continuous-Time Band-limited Gaussian Channel

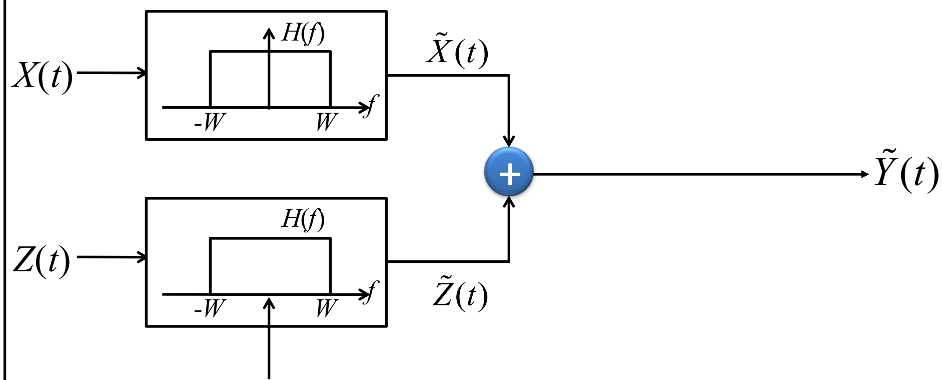


$$S_Z(f) = \frac{N_0}{2}$$

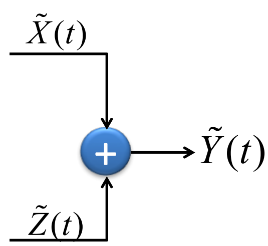
Redrawing the system…

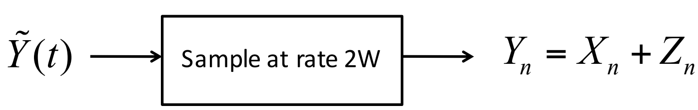


We can move the filter to before the addition.



Equivalent Sampled Inputs and Outputs





Capacity of discrete time channel

- $\lim_{n\to\infty}\frac{1}{2n}I\left(\left(X(t)\Big|_{-nT}^{nT}\right);\ \left(Y(t)\Big|_{-nT}^{nT}\right)\right) = \lim_{n\to\infty}\frac{1}{2n}I\left(X^{2n};Y^{2n}\right)$
- Capacity of resulting discrete time channel is
  - $C = \frac{1}{2}\log(1+SNR)$ per symbol

  or
  - $C = 2W\cdot\frac{1}{2}\log(1+SNR)$ bits per second

$$\boxed{C = W\log(1+SNR)\ \text{bits per second}}$$

What is SNR?

Evaluating Noise Power

- $E\left[\tilde{Z}^2(t)\right] = R_{\tilde{Z}}(0)$

$$= \int_{-\infty}^{\infty} S_{\tilde{Z}}(f)df$$

$$= \int_{-W}^{W}\frac{N_0}{2}df$$

$$= N_0 W$$

- Let $P = E[X_n^2]$

$$\boxed{C = W\log\left(1+\frac{P}{N_0W}\right)}$$

Limit of infinite bandwidth

$$\ln(1+x) \approx x$$
$$(= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots)$$

$$\lim_{W\to\infty} W \log\left(1+\frac{P}{N_0 W}\right) = \lim_{W\to\infty} W\left(\log_2 e\right)\ln\left(1+\frac{P}{N_0 W}\right)$$
$$= W \log_2 e \frac{P}{N_0 W}$$
$$= \log_2 e \frac{P}{N_0}$$

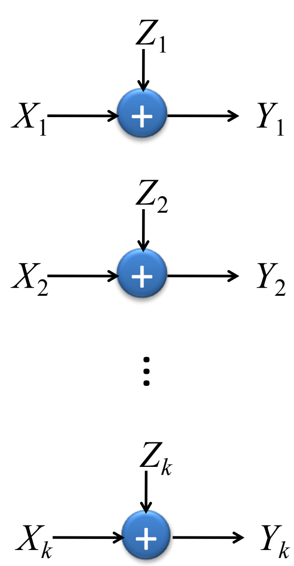
Infinite Bandwidth Conclusions

$$\lim_{W\to\infty} W \log\left(1+\frac{P}{N_0 W}\right) = \log_2 e \frac{P}{N_0}$$

- Capacity grows linearly with SNR for extremely large bandwidths.
- For fixed P/N_0 capacity increases monotonically to the constant $\log_2 e \frac{P}{N_0}$.

Part 12 C: Capacity of Parallel Gaussian Channels

Parallel Gaussian Channels



Capacity of parallel channels

$$C = \max_{f(x_1, x_2, \ldots, x_k), \sum EX_i^2 \le P} I(X_1, X_2, \ldots, X_k; Y_1, Y_2, \ldots, Y_k)$$

Capacity of Parallel Gaussian Channels

$$\begin{aligned} I(X^k; Y^k) &= h(Y^k) - h(Y^k \mid X^k) \\ &= h(Y^k) - h(Z^k \mid X^k) \\ &= h(Y^k) - h(Z^k) \\ &= h(Y^k) - \sum_i h(Z_i) \\ &\le \sum_i \left(h(Y_i) - h(Z_i) \right) \\ &\le \sum_i \frac{1}{2} \log(1 + \frac{P_i}{N_i}) \end{aligned}$$

Achievability

- Where $P_i = EX_i^2, \sum P_i = P$ and equality is achieved when

$$X^k \sim N\left(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & P_k \end{bmatrix}\right)$$

- What are the optimal P_i's?

Convert to convex optimization

$$\begin{aligned} &\text{maximize} && \sum_i \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right) && \text{(Concave function)} \\ &\text{subject to} && \sum_i P_i \le P && \text{(linear constraint)} \end{aligned}$$

- Convex optimization techniques may be applied.

Lagrange multipliers (Duality)

$$J(P_1,\ldots,P_k) = \sum_i \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right) + \lambda \sum_i P_i$$

$$\frac{\partial J}{\partial P_i} = \frac{1}{2}\frac{\partial}{\partial P_i}\log_2 e\left[\ln\left(\frac{1}{N_i}\right)+\ln(N_i+P_i)\right]+\lambda \quad = \frac{\frac{1}{2}\log_2 e}{N_i+P_i}+\lambda$$

$$\boxed{\text{Set } \frac{\partial J}{\partial P_i} = 0}$$

$$\frac{\frac{1}{2}\log_2 e}{N_i+P_i}+\lambda = 0 \Longrightarrow \frac{1}{2}\log_2 e = -\lambda\left(N_i+P_i\right) \Longrightarrow P_i = \underbrace{\frac{-1}{2\lambda}\log_2 e}_{\nu} - N_i$$

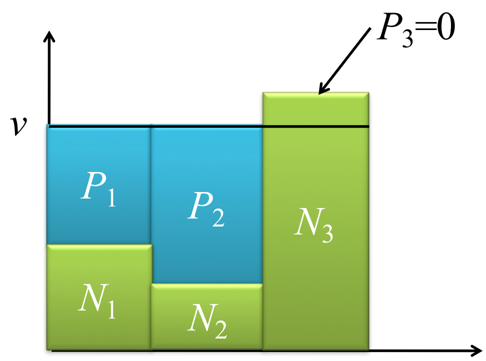
Optimal solution

- Setting $\frac{\partial J}{\partial P_i} = 0$ is equivalent to setting $P_i = \nu - N_i$ for some constant ν. (except where $\nu - N_i$ is negative.)
- Choose ν to meet power constraint.

$$\sum (\nu - N_i)^+ = P$$

$$(x)^+ = \begin{cases} x & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Water filling



Part 12 D: Sufficient Statistics

Section 2.10 "Sufficient Statistics"

- $\{P_\theta(x)\}$ or $\{f_\theta(x)\}$ is a family of pmf's or pdf's on x indexed by θ.
- x^n is a random sampling of $x_1,\ldots,x_n$ from a distribution in the family.

Statistics

- Guessing or estimating θ from observation of X^n.
- A "statistic" $T(X^n)$ is a function of X^n that is used to prepare our estimate.

Bernoulli Example

- $0 \le \theta \le 1$, X is Bernoulli $\quad P_\theta(1) = \theta$, $P_\theta(0) = 1 - \theta$
- X^n is the result of n flips
  - 1 1 0 1 0 ….
- $T(X^n)$ might be $K = \sum_{i=1}^{n} x_i$
  - i.e. the number of heads.

Sufficient Statistics

- A Sufficient Statistic contains all the information in X^n about θ.
- Formally, $T(X^n)$ is a sufficient statistic of X^n for θ if $\theta \to T(X^n) \to X^n$ or equivalently if

$$I(\theta; X^n) = I\left(\theta; T(X^n)\right)$$

Sufficient Statistics (cont.)

- Check if $T(X^n)$ is sufficient by checking if $\theta \to T(X^n) \to X^n$.
- i.e. Check if $P\left(X^n \mid \theta, T(X^n)\right) = P\left(X^n \mid T(X^n)\right)$.

K is a Suff. Stat. in Bernoulli Example.

- Consider the Bernoulli example

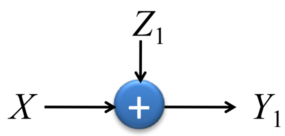
$$P(X^n \mid \theta, k) = \frac{P(X^n, k \mid \theta)}{P(k \mid \theta)} = \frac{\theta^k (1-\theta)^{n-k}}{\binom{n}{k} \theta^k (1-\theta)^{n-k}} = \frac{1}{\binom{n}{k}}$$

$$P(X^n \mid k) = \frac{1}{\binom{n}{k}}$$

- Yes, k is a sufficient statistic of X^n for θ.

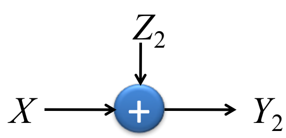
Gaussian Example

- Now consider



$$Y_1 = X + Z_1, \quad Y_2 = X + Z_2$$

$$Z_i \sim N(0, \sigma^2)$$



$$\boxed{T(Y_1, Y_2) = Y_1 + Y_2}$$

Showing Sum is a Suff. Stat. for Gauss.

- Show that $T(Y_1, Y_2) = Y_1 + Y_2$ is a sufficient statistic of (Y_1, Y_2) for X.

$$X \to T = \sum_{i=1}^{2} Y_i \to Y_1, Y_2$$

$$Y_1 \sim N(X, \sigma^2)$$
$$Y_2 \sim N(X, \sigma^2)$$
$$T = Y_1 + Y_2$$
$$T \sim N(2X, 2\sigma^2)$$

We need to show that
$$f(y_1, y_2 \mid t, x) = f(y_1, y_2 \mid t)$$

- X is a random variable.

The relevant conditional distributions

- Given X, we have the following conditional distributions

$$f(y_1, y_2 \mid x) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(y_1 - x)^2 + (y_2 - x)^2}{2\sigma^2}\right]$$

$$f(t \mid x) = \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left[-\frac{(t - 2x)^2}{4\sigma^2}\right]$$

$$f(y_1, y_2, t \mid x) = f(y_1, y_2 \mid x)\underbrace{\delta(y_1 + y_2 - t)}_{\text{Dirac delta function}}$$

$$f(y_1, y_2 \mid t, x) = \frac{f(y_1, y_2, t \mid x)}{f(t \mid x)}$$

$$= \frac{f(y_1, y_2 \mid x)\delta(y_1 + y_2 - t)}{f(t \mid x)}$$

$$= \frac{\frac{1}{2\pi\sigma^2} \exp\left[-\frac{(y_1 - x)^2 + (y_2 - x)^2}{2\sigma^2}\right]\delta(y_1 + y_2 - t)}{\frac{1}{\sqrt{4\pi\sigma^2}} \exp\left[-\frac{(t - 2x)^2}{4\sigma^2}\right]}$$

$$= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{2(y_1 - x)^2 + 2(y_2 - x)^2 - (t - 2x)^2}{4\sigma^2}\right]\delta(y_1 + y_2 - t)$$

$$= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{2(y_1 - x)^2 + 2(y_2 - x)^2 - ((y_1 - x) + (y_2 - x))^2}{4\sigma^2}\right]\delta(y_1 + y_2 - t)$$

$$= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{2(y_1-x)^2+2(y_2-x)^2-((y_1-x)+(y_2-x))^2}{4\sigma^2}\right]\delta(y_1+y_2-t)$$

$$= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{(y_1-x)^2+(y_2-x)^2-2(y_1-x)(y_2-x)}{4\sigma^2}\right]\delta(y_1+y_2-t)$$

$$= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{((y_1-x)-(y_2-x))^2}{4\sigma^2}\right]\delta(y_1+y_2-t)$$

$$= \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{(y_1-y_2)^2}{4\sigma^2}\right]\delta(y_1+y_2-t)$$

Strangeness

- Our expression looks strange because the variances in the coefficient and exponent don't seem to match.

- Let's see if it integrates to 1.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sqrt{\pi\sigma^2}}\exp\left[-\frac{(y_1-y_2)^2}{4\sigma^2}\right]\delta(y_1+y_2-t)dy_1dy_2$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{\pi\sigma^2}}\exp\left[-\frac{(t-2y_2)^2}{4\sigma^2}\right]dy_2$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{\pi\sigma^2}}\exp\left[-\frac{(y_2-t/2)^2}{\sigma^2}\right]dy_2$$

$=1$ –Since we are integrating a Gaussian.

What we know about Y's given T

- Given $T=t,\ Y_2 \sim N(\frac{t}{2},\frac{\sigma^2}{2})$, by symmetry $Y_1 \sim N(\frac{t}{2},\frac{\sigma^2}{2})$ as well and $Y_1 = T - Y_2$.

Conclusion

We need to show that

$$f(y_1, y_2 \mid t, x) = f(y_1, y_2 \mid t)$$

- This turns out to be easy since $f(y_1, y_2 \mid t, x)$ does not depend on x.

$$\boxed{f(y_1, y_2 \mid t, x) = \frac{1}{\sqrt{\pi\sigma^2}} \exp\left[-\frac{(y_1 - y_2)^2}{4\sigma^2}\right] \delta(y_1 + y_2 - t)}$$

$$\begin{aligned} f(y_1, y_2 \mid t) &= \int_x f(y_1, y_2, x \mid t) dx \\ &= \int_x f(y_1, y_2 \mid t, x) f(x) dx \\ &= f(y_1, y_2 \mid t, x) \end{aligned}$$