The Gaussian Channel

Information Theory Lecture 8b

Let's consider a channel

- Which delivers real (continuous) values
- In real time
- And has real noise

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

- Real channels have real costs, in terms of maximum *power*

$$\frac{1}{T}\int_0^T dt\left[x(t)\right]^2 \leq P$$

How to send information

- Think of the signal being comprised of a superposition of orthonormal basis functions

$$x(t) = \sum_{n=1}^{N} x_n \phi(t)$$

- We can encode a set of *n* real numbers by setting the amplitudes of these basis functions at the sending end
- And taking a Fourier transform at the other end

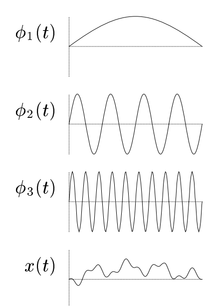
The common basis functions

- Let's assume we get white noise (0 mean Gaussian noise) in the channel

$$n_0 = \text{Normal}\left(0, \sigma^2\right)$$

- Then, the transform at the output yields

$$y_0 = \int_0^T dt\phi_n(t)\, y(t) = x_0 + \int_0^T dt\phi_n(t)\, n(t)$$
$$= x_n + n_0$$



Limited Power

- Recall

$$\frac{1}{T}\int_0^T dt\left[x(t)\right]^2 \leq P$$

$$\sum_n x_n^2 \leq PT$$

$$\frac{x_n^2}{N} \leq \frac{PT}{N}$$

Limited *Bandwidth*

- If the highest frequency present in a signal is *W*, then we can figure the amplitude of that frequency by a Nyquist *sampling* interval of

$$\Delta t = \frac{1}{2W}$$

- This means for our time interval of *T*, the maximum number of orthogonal basis functions is

$$N^{\max} = 2WT$$

$$W = \frac{N^{\max}}{2T}$$

The relationship

- Between a real, continuous time channel and a *discrete time* Gaussian Channel
- The use of a real continuous channel is equivalent to $2W$ uses per second of a Gaussian Channel with the same noise level and power constraint

$$\frac{x_n^2}{N} \le \frac{P}{2W}$$

Discrete time

- Let's assume we are using a code where we send a vector of real numbers to represent a single bit in one time step

$$P(\mathbf{n}) = \left[\det\left(\frac{\mathbf{A}}{2\pi}\right)\right]^{1/2} \exp\left(-\frac{1}{2}\mathbf{n}^{\mathbf{T}}\mathbf{A}\mathbf{n}\right)$$

$$P(\mathbf{y} \mid s) = \left[\det\left(\frac{\mathbf{A}}{2\pi}\right)\right]^{1/2} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{x}_s)^{\mathbf{T}}\mathbf{A}(\mathbf{y}-\mathbf{x}_s)\right)$$

Consider a discriminator

- Based on the a posteriori probability ratio

$$\frac{P(s=1 \mid \mathbf{y})}{P(s=0 \mid \mathbf{y})} = \frac{P(\mathbf{y} \mid s=1)}{P(\mathbf{y} \mid s=0)} \frac{P(s=1)}{P(s=0)}$$

$$= \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{x}_1)^T \mathbf{A}(\mathbf{y}-\mathbf{x}_1) + \frac{1}{2}(\mathbf{y}-\mathbf{x}_0)^T \mathbf{A}(\mathbf{y}-\mathbf{x}_0) + \ln\frac{P(s=1)}{P(s=0)}\right)$$

$$= \exp\left(\mathbf{y}^T \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_0) + \theta\right)$$

$$\theta = -\frac{1}{2}\mathbf{x}_1^T \mathbf{A}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_0^T \mathbf{A}\mathbf{x}_0 + \ln\frac{P(s=1)}{P(s=0)}$$

We can use a *linear* discriminator

- Based on the log of the a posteriori ratio

$$a(\mathbf{y}) = \mathbf{w}^{\mathbf{T}}\mathbf{y} + \theta$$

$$\mathbf{w} = \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_0)$$

$$a(\mathbf{y}) > 0 \rightarrow \text{guess } s = 1$$

$$a(\mathbf{y}) < 0 \rightarrow \text{guess } s = 0$$

$$a(\mathbf{y}) = 0 \rightarrow \text{guess either}$$

Channel Capacity

- Remember that the channel capacity is found by maximizing the mutual information between the input and the output

$$I(X;Y) = \int dxdy P(x,y) \log\frac{P(x,y)}{P(x)P(y)}$$

$$= \int dxdy P(x \mid y) \log\frac{P(x \mid y)}{P(y)}$$

- But we also have to consider the power constraint

Lagrange Multipliers

- A trick to maximize constrained functions
- Basically, you subtract a term to represent the constraint
- Since the gradient of both terms has to be zero and the maximum, this generally works
- The λ is included to balance the effects of both terms, and is itself a variable

$$\max\ f(x)$$

$$\text{subject to } g(x) = C$$

$$\frac{df(x)}{dx} = 0$$

$$\frac{dg(x)}{dx} = 0$$

$$\therefore$$

$$\max\ f(x) - \lambda g(x)$$

$$0 = \frac{df(x)}{dx} - \lambda\frac{dg(x)}{dx}$$

Therefore

- To enforce the power constraint and keep *P(x)* a probability distribution

$$F = I(X;Y) - \lambda \int dxP(x)x^2 - \mu \int dxP(x)$$
$$= \int dxP(x)\left[\int dyP(y|x)\ln\frac{P(y|x)}{P(y)} - \lambda x^2 - \mu\right]$$
$$\frac{\partial F}{\partial P(x^*)} = \int dyP(y|x^*)\ln\frac{P(y|x^*)}{P(x)} - \lambda x^{*2} - \mu - 1$$
$$= 0$$

Given that this is a Gaussian Channel

$$P(y|x) = \frac{\exp\left(\frac{-(y-x)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$$
$$\int dy \frac{\exp\left(-(y-x)^2/2\sigma^2\right)}{\sqrt{2\pi\sigma^2}} \ln\left[P(y)\sigma\right] = -\lambda x^2 - \mu'$$
$$\ln\left[P(y)\sigma\right] = a + by + cy^2 \ldots$$

Therefore, *P(y)* and thus *P(x)*, are Gaussian for the maximum mutual information

Given the best input distribution

- In terms of maximum mutual information is Gaussian
- Let's say it has variance *v*, and therefore the output has standard deviation *v+σ²*

$$I(X;Y) = \int dxdyP(x)P(y|x)\log P(y|x) - \int dyP(y)\log P(y)$$
$$= \frac{1}{2}\log\frac{1}{\sigma^2} - \frac{1}{2}\log\frac{1}{v+\sigma^2}$$

The Capacity of a Gaussian Channel

- Is therefore given by

$$C = \frac{1}{2}\log\left(1 + \frac{v}{\sigma^2}\right)$$

- The last term, the ratio of the signal and noise variances, is also the ratio of their power, and is called *the signal-to-noise ratio* (SNR)
- This important quantity is usually measured in decibels

Decibels

- A few useful conversions

$$\text{dB} = 10\log\frac{P_2}{P_1}$$
$$\text{dB} = 20\log\frac{V_2}{V_1}$$

- And standards:
  - For audio, P_1 is for 1mW in 600 Ohms
  - For Radio/TV, P_1 is for 1mW rms in 75 Ohms
  - For Radio frequency, P_1 is for 1mW in 50 Ohms or 1μV/m field strength
  - Radio engineers use dBm (1mW) or dBu (1μV)

Back to continuous time

- Recall that he use of a real continuous channel is equivalent to 2*W* uses per second of a Gaussian Channel with the same noise level and power constraint

$$\frac{x_n^2}{N} \le \frac{P}{2W}$$

- Substituting into the channel capacity

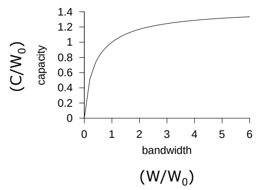
$$C = W\log\left(1 + \frac{P}{2\sigma^2 W}\right) = W\log\left(1 + \frac{P}{N_0 W}\right)$$

How to transmit

- Let's look at bandwidths and relative to a standard SNR

$$W_0 = P/N_0$$

- So, it'd be good to take tons of bandwidth, at low power
- However, it will tick off your neighbours!



Take Home Messages

- Real channels send information in orthonormal basis functions
- This transmission is limited by power and bandwidth
- Looking at the discrete time Gaussian Channel, the power limited channel capacity is defined in terms of SNR
- We can relate this back to the continuous time channel