

Gaussian Channel

The Most Important Continuous Channel

Radu Trîmbițaș

November 2012

Outline

Contents

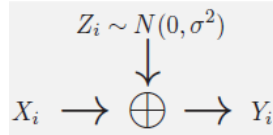
1	Introduction	1
2	Capacity of Gaussian Channel	4
3	The Coding Theorem for Gaussian Channels	5
3.1	Coding Theorem	5
4	Parallel Gaussian Channels	7
4.1	Uncorrelated noise	7
4.2	Colored Noise	11

1 Introduction

Continuous Channels

- So far, we have considered discrete channels which are modeled by conditional probability distributions $p(y|x)$.
- That is, for a given $x \in \mathcal{X}$, $p(y|x)$ models the form of distortion that x undergoes when it is being sent from source to receiver.
- Real channels are continuous as are real signals. What really happens to a continuous random variable X is that we have $Y = v(X)$ where v is a random function that may or may not be dependent on X .
- This is quite hard to analyze so we may consider only additive noise $Y = X + v$ where v is a random variable.
- We further simplify by saying that v and X are independent and moreover that v is Gaussian, leading to the Gaussian channel.

Gaussian Channel



- Above is our model, where $Y_i = X_i + Z_i$, with $Z_i \sim N(0; \sigma^2)$ and Z_i, X_i independent.
- If $\sigma^2 = 0$, what is the capacity of this channel?
- If $\sigma^2 = 0$, capacity is infinite since one can perfectly send an arbitrarily precise real number (consider arithmetic coding, it sends a number all within $[0, 1)$).
- If $\sigma^2 > 0$, what is the capacity?
- If $\sigma^2 > 0$, capacity is still infinite, since we can make input power as large as we want, effectively removing a finite strict subinterval within $[0, 1)$.
- If input power is constrained as well (which is also more practical and realistic), then the problem becomes interesting.

Power constraint

- Average power constraint: for any codeword (x_1, \dots, x_n) of length n , we require that

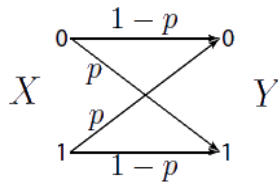
$$\frac{1}{n} \sum_{i=1}^n x_i^2 \leq P. \quad (1)$$

where P is the average power $\approx E(X^2)$

- This communication channel models many practical channels, including radio and satellite links.
- The additive noise in such channels may be due to a variety of causes.
- However, by the central limit theorem, the cumulative effect of a large number of small random effects will be approximately normal, so the Gaussian assumption is valid in a large number of situations.

Example

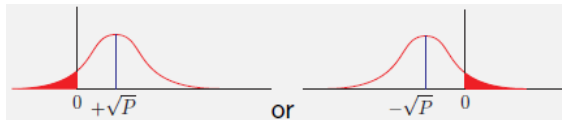
- Send 1 bit over channel at a time (obviously sub-optimal use of the channel).
- $X \in \{-\sqrt{P}, \sqrt{P}\}$ means that $E(X^2) = P$, so this satisfies the constraint.



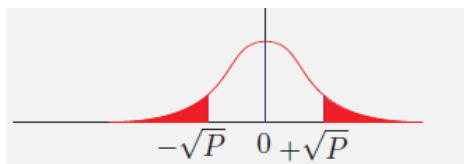
- For a uniform source distribution, decode as $+\sqrt{P}$ if $Y > 0$ and $-\sqrt{P}$ if $Y < 0$.
- Error:

$$\begin{aligned}
 P_e &= \frac{1}{2}P(Y < 0|X = \sqrt{P}) + \frac{1}{2}P(Y > 0|X = -\sqrt{P}) \\
 &= \frac{1}{2}P(Z < \sqrt{P}|X = \sqrt{P}) + \frac{1}{2}P(Z > -\sqrt{P}|X = -\sqrt{P}) \\
 &= P(Z > \sqrt{P}).
 \end{aligned}$$

- The two separate error types



- Lead to total error



- We have that

$$P(Z > \sqrt{P}) = 1 - \Phi\left(\frac{\sqrt{P}}{\sigma^2}\right).$$

- In fact, we have essentially just turned a Gaussian channel into a discrete BSC:
where $p = P_e$ for the Gaussian.
- This will be the common idea. We convert continuous channels into discrete ones with appropriate encodings.
- This is essentially a process of vector quantization (where under different quantization schemes, we study the tradeoffs that exist when coding). Tradeoffs take the form of rate vs. distortion under the power constraints.

2 Capacity of Gaussian Channel

Capacity of Gaussian Channel

We need a capacity notion, but here under a power constraint.

Definition 1. The information capacity of the Gaussian channel with power constraint P is

$$C = \max_{f(x): EX^2 \leq P} I(X; Y). \quad (2)$$

- $I(X; Y)$ has a nice form in this case, as

$$\begin{aligned} I(X; Y) &= h(Y) - h(Y|X) = h(Y) - h(X + Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z) \end{aligned}$$

- But since Z is Gaussian, $h(Z) = \frac{1}{2} \log(2\pi e \sigma^2)$ where σ^2 is the noise power, $E(Z^2) = \sigma^2 = N$, with $E(Z) = 0$.
- We also saw earlier, since Gaussians have maximum entropy for a given 2nd moment, that if $EX = 0$, $var(X) = K$, then

$$h(X) \leq \frac{1}{2} \log [2\pi e^2 |K|]$$

- Also,

$$\begin{aligned} E(Y^2) &= E(X + Z)^2 = E(X^2) + 2E(X)E(Z) + E(Z^2) \\ &= \underbrace{P}_{\text{signal power}} + \underbrace{\sigma^2}_{\text{noise power}} \end{aligned}$$

- Thus, we can upper bound the mutual information as follows:

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

where SNR is the signal to noise ratio.

- We can achieve the bound on $h(Y)$ by ensuring Y is Gaussian, and this is the case if X is Gaussian (sums of Gaussians are Gaussian).
- The capacity of the Gaussian channel is

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

- Makes sense: the maximum transmission rate obtained when $X \sim N(0; P)$.
- Rate depends on SNR - if signal level is allowed to be much larger than noise, then rate should increase (log when information measured in bits).
- In fact, from this we get the standard 6.02 dB SNR/bit for audio. I.e., $16 = 1/2 \log(1 + SNR)$, or $2^{32} = 1 + SNR$ or $SNR = 2^{32} - 1$.
- $10 \log_{10}(SNR) = 10 \times 32 \log_{10}(2) = 96.33$ dB.
- And $96.33/16 = 6.02$ dB/bit.
- Every additional bit (on an audio CD) adds 6.02 dB of SNR.

Capacity of the channel definitions

Definition 2. An (M, n) code for the Gaussian channel, with power constraint P , consist of

1. An index set $\{1, 2, \dots, M\}$
2. An encoding function $x : \{1, 2, \dots, M\} \rightarrow \mathcal{X}^n$ giving codewords $X^n(1), X^n(2), \dots, X^n(M)$ with

$$\sum_{i=1}^n x_i^2(w) \leq nP, \quad w = 1, 2, \dots, n.$$

3. A decoding function $g : \mathcal{Y}^n \rightarrow \{1, 2, \dots, M\}$.

Definition 3. The *rate* is

$$R = \frac{\log M}{n} \text{ bits per channel use.}$$

Definition 4. The average probability of error is

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum \lambda_i.$$

Definition 5. A rate R is *achievable* if \exists a sequence of $(2^{nR}, n)$ codes satisfying the power constraint P s.t. $\lambda^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. The capacity of the channel is the supremum of the achievable rates.

3 The Coding Theorem for Gaussian Channels

3.1 Coding Theorem

The Coding Theorem for Gaussian Channels

Theorem 6. The capacity of a Gaussian channel with input power constraint P and noise variance $\sigma^2 = N$ is

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right) \text{ bits per transmission.} \quad (3)$$

Proof sketch...

- Typical X set $A_\epsilon^{(n)}$ has volume $\leq 2^{n(h(X)+\epsilon)}$
- Conditional typical Y volume $\leq 2^{n(h(Y|X)+\epsilon)} = 2^{n(h(Z)+\epsilon)}$
- Unconditional typical set volume of $Y \leq 2^{n(h(Y)+\epsilon)}$, but $h(Y) \leq \frac{1}{2} \log [2\pi e (P + \sigma^2)]$ and $h(Z) = \frac{1}{2} \log (2\pi e \sigma^2)$.

□

The Coding Theorem for Gaussian Channels

... proof sketch.

- How many X -conditional volumes can we pack into total available volume?

$$\leq \frac{2^{nh(Y)}}{2^{nh(Z)}} = \frac{2^{n \frac{1}{2} \log [2\pi e (P + \sigma^2)]}}{2^{n \frac{1}{2} \log [2\pi e \sigma^2]}} = 2^{\frac{n}{2} \log \left(\frac{P + \sigma^2}{\sigma^2} \right)} = \left(\frac{P + \sigma^2}{\sigma^2} \right)^{\frac{n}{2}}$$

- The above is measured in counts for n channel usages. To convert it into bits per channel use, we take log and divide by n to get

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

- Assuming no overlap of volumes which is best we can do.

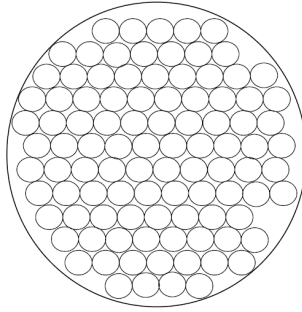
□

The Coding Theorem for Gaussian Channels

- Moreover, if everything is jointly Gaussian, and i.i.d., then the typical volumes will be spheres, and we can relate the volume of a sphere to the typical set to get the radius (in n -Dim)

$$r_{\sigma^2} = \Gamma^{1/2} \left(\frac{n}{2} + 1 \right) (2e\sigma^2)^{\frac{1}{2}}.$$

- figure



proof diff. from discrete proof

proof Δ from discrete proof. • We need to show that if $R < C$, \exists a code with $P_e^n \rightarrow 0$ when $n \rightarrow \infty$.

- We do random codeword generation (like in discrete case) but in this case from Gaussians with $E(X^2) = P - \epsilon$ so that

$$\frac{1}{n} \sum_{i=1}^n x_i^2 \rightarrow P - \epsilon, \quad \text{as } n \rightarrow \infty$$

- Also have an additional source of possible error

$$E_0 = \left\{ \frac{1}{n} \sum x_i^2(1) > P \right\}$$

□

proof diff. from discrete proof

proof Δ from discrete proof. • We add E_0 to the other errors we previously had in discrete case (note again we use same trick to show that considering only message 1 is sufficient due to symmetry).

- By the weak law of large numbers, $E_0 \rightarrow 0$ also as $n \rightarrow \infty$ as do the other sources of errors.

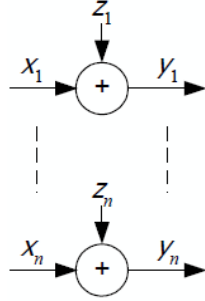
□

4 Parallel Gaussian Channels

4.1 Uncorrelated noise

Parallel Gaussian Channels

- Suppose we have k independent Gaussian channels with a common power constraint.



$Y_j = X_j + Z_j, j = 1, 2, \dots, k, Z_j \sim N(0, N_j),$
 Z_i, Z_j independent for $i \neq j$

$$Z \sim N \left(0, \begin{bmatrix} N_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & N_k \end{bmatrix} \right)$$

- power constraint

$$E \left(\sum_{i=1}^n X_j^2 \right) \leq P.$$

- Goal: to distribute the total power among the channel in order to maximize the total capacity.

Parallel Channels

- Find maximum capacity

$$C = \max_{f(x_{1:k}) : \sum E(X_j^2) \leq P} I(X_{1:k}, Y_{1:k}) \quad (4)$$

- We have

$$I(X_{1:k}, Y_{1:k}) = h(Y_{1:k}) - h(Y_{1:k} | X_{1:k}) = h(Y_{1:k}) - h(Z_{1:k} | X_{1:k}) \quad (5)$$

$$= h(Y_{1:k}) - h(Z_{1:k}) = h(Y_{1:k}) - \sum_i h(Z_i) \quad (6)$$

$$\leq \sum_j [h(Y_j) - h(Z_j)] \leq \sum_j \frac{1}{2} \log \left(1 + \frac{P_i}{N_i} \right) \quad (7)$$

- with $P_i = EX_i^2$ and $\sum_i P_i = P.$
- The way to solve this is to solve the optimization problem

$$\max_{(P_{1:k})} \sum_i \frac{1}{2} \log \left(1 + \frac{P_i}{N_i} \right) \quad (8)$$

$$\text{subject to} \quad \sum_i P_i = P \quad (9)$$

- Or in Lagrangian form

$$J(P_{1:k}) = \sum_i \frac{1}{2} \log \left(1 + \frac{P_i}{N_i} \right) + \lambda \left(\sum_i P_i - P \right)$$

Lagrangian optimization

$$\begin{array}{ll} \text{find} & \min_x f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

Lagrangian form

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

and define dual objective

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu).$$

Strong duality

- Notation: $\lambda_{opt} = \lambda^*, \nu_{opt} = \nu^*$
- Strong duality means $f(x_{opt}) = g(\lambda_{opt}, \nu_{opt})$
- Karush-Kuhn-Tucker conditions for optimality,

$$\begin{array}{ll} f_i(x_{opt}) \leq 0, & i = 1, \dots, m \\ h_i(x_{opt}) = 0, & i = 1, \dots, p \\ \lambda_i^* \geq 0, & i = 1, \dots, m \\ \lambda_i^* f_i(x_{opt}) = 0, & i = 1, \dots, m \end{array}$$

and

$$\nabla_x L|_{x=x_{opt}} = 0$$

Back to our problem

- We get Lagrangian

$$L(x, \lambda, \nu) = - \sum_{i=1}^k \frac{1}{2} \log \left(1 + \frac{P_i}{N_i} \right) - \sum_{i=1}^k \lambda_i P_i + \nu \left(\sum_{i=1}^k P_i - P \right).$$

- KKT conditions are:

$$\forall i P_i \geq 0, \sum_i P_i^* = P, \forall i \lambda_i^* \geq 0, \lambda_i^* P_i^* = 0$$

and also $\forall i$

$$-\frac{1}{(1 + P_i/N_i)} \frac{1}{N_i} - \lambda_i^* + v^* = 0.$$

KKT Conditions

- From the Lagrangian gradient conditions we can further get

$$-\frac{1}{N_i + P_i} - \lambda_i^* + v^* = 0 \implies -\frac{1}{N_i + P_i} + v^* = \lambda_i^* \geq 0$$

- We then eliminate λ_i^* to get KKT conditions in form

$$\begin{aligned} P_i^* &\geq 0, & \sum_i P_i^* &= P \\ \left(v^* - \frac{1}{N_i + P_i} \right) P_i^* &= 0, & v^* &\geq \frac{1}{N_i + P_i^*} \end{aligned}$$

Final solution

- So, P_i^* must have form

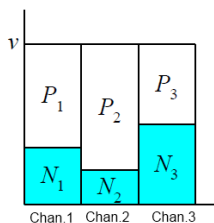
$$P_i^* = \left(\frac{1}{v^*} - N_i \right)_+$$

where $a_+ = \max\{0, a\}$.

- With the last constraint, we have that

$$\sum_i \left(\frac{1}{v^*} - N_i \right)_+ = P$$

- This leads to the famous water filling idea for parallel channels.



Capacity of Parallel Channels

- Hence, the final capacity is

$$\begin{aligned} C_n &= \frac{1}{2} \sum_{i=1}^k \log \left(1 + \frac{P_i}{N_i} \right) \\ &= \frac{1}{2} \sum_{i=1}^k \log \left(1 + \frac{\left(\frac{1}{v^*} - N_i \right)_+}{N_i} \right) \text{ bits per parallel channel use} \end{aligned}$$

- In units of bits per transmission (bits per single channel transmission, take the average):

$$C_n = \frac{1}{2n} \sum_{i=1}^k \log \left(1 + \frac{(v^* - N_i)_+}{N_i} \right)$$

4.2 Colored Noise

Colored Noise

- Consider the case when the noise is dependent
- Suppose $Y = X + Z$ where $E(ZZ^T) = K_Z$ and $E(XX^T) = K_X$
- We want to find K_X to maximize capacity subject to power constraint:

$$\sum E(X_i^2) \leq nP \iff \text{tr}(K_X) \leq nP$$

- Find noise eigenvectors: $K_Z = Q\Lambda Q^T$ with $QQ^T = I$.

- Now

$$Q^T Y = Q^T X + Q^T Z = Q^T X + W$$

where

$$E(WW^T) = E(Q^T Z Z^T Q) = E(Q^T K_Z Q) = \Lambda$$

is diagonal;

- W_i are now independent (so previous result on Parallel Gaussian Channels applies)
- Power constraint is unchanged $\text{tr}(Q^T K_X Q) = \text{tr}(K_X Q Q^T) = \text{tr}(K_X)$
- Use water-filling and indep. messages $Q^T K_X Q + \Lambda = \nu I$
- Choose $Q^T K_X Q = \nu I - \Lambda$ where $\nu = P + n^{-1} \text{tr}(\Lambda) \implies K_X = Q(\nu I - \Lambda) Q^T$.

Colored Noise

- If Z is from a stationary process then $\text{diag}(\Lambda) \rightarrow$ power spectrum $N(f)$, as $k \rightarrow \infty$
- To achieve capacity use waterfilling on noise power spectrum

$$P = \int_{-W}^W (v - N(f))_+ df$$
$$C = \frac{1}{2} \int_{-W}^W \log \left(1 + \frac{(v - N(f))_+}{N(f)} \right) df$$

- Waterfilling on spectral domain

