lec 24: Colored Gaussian Noise
Rate-distortion theory

Recap:

* Converse of the coding theorem for Gaussian channels.
* bandlimited Gaussian noise channels
* parallel Gaussian channels

Bandlimited Gaussian noise channels

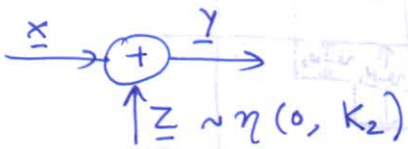
Gallager p 406.

fns time limited to $\left(-\frac{T}{2}, \frac{T}{2}\right)$

$\Rightarrow \approx 2WT$ degrees of freedom

(see also Wozencraft & Jacobs)

Colored Gaussian-noise setting



$$\underline{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \frac{\sum E[x_i^2]}{n} \le P$$

Define

$$K_x = E[\underline{x}\,\underline{x}^T]$$

$$\Rightarrow \text{trace}(K_x) \le nP$$

$$\text{tr}(A) = \sum_i A_{ii}$$

= sum of diagonal elements

$$C = \max_{P(x):\, Tr(K_x) \le nP} I(x;y)$$

$$I(\underline{x};\underline{y}) = h(\underline{y}) - h(\underline{y}|\underline{x})$$

$$= h(\underline{y}) - h(\underline{Z} + \underline{x}|\underline{x})$$

$$= h(\underline{y}) - h(\underline{z})$$

$$K_y = E[\underline{y}\,\underline{y}^T] = E\{(\underline{x}+\underline{z})(\underline{x}+\underline{z})^T\}$$

$$= E\{x x^T + z z^T + z x^T + x z^T\}$$

$$= E[x x^T] + E[z z^T] = K_x + K_z$$

$$\therefore \; I(X;Y) \le \frac{1}{2}\log(2\pi e)^n |K_X + K_Z| - \frac{1}{2}\log(2\pi e)^n |K_Z|$$

Thus our goal is to maximize

$$\frac{1}{2}\log(2\pi e)^n |K_X + K_Z| \text{ subject to } Tr(K_X) \le nP.$$

We can factor K_Z in the form $K_Z = Q \Lambda Q^t$ where Λ is a diagonal matrix of eigenvalues of K_Z and where Q is a ~~unitary~~ orthogonal matrix. $QQ^t = Q^tQ = I_n$ [as K_Z is a real symmetric matrix].

$$\therefore |K_X + K_Z| = |K_X + Q\Lambda Q^t| = |Q(Q^t K_X Q + \Lambda)Q^t|$$

$$= |Q|\,|Q^t K_X Q + \Lambda|\,|Q^t|$$

$$= |\underbrace{Q^t K_X Q}_{\triangleq A} + \Lambda|$$

$$= |A + \Lambda|$$

Note that $Tr(A) = Tr(Q^t K_X Q)$

Now if U, V are $n \times n$ matrices then it $tr(UV) = tr(VU)$

$$\Rightarrow Tr(A) = Tr(Q^t K_X Q) = Tr(K_X Q^T Q) = Tr(K_X) \le nP.$$

Goal is to ~~minimize~~ maximize $|A+\Lambda|$ subject to $tr(A) \le nP$

Now by Hadamard's inequality, (shown earlier)

$$|A+\Lambda| \le \prod_{i=1}^{n} (A+\Lambda)_{ii} = \prod_{i=1}^{n}(a_{ii} + \lambda_i)$$

$$\therefore \max \prod_{i=1}^{n}(a_{ii} + \lambda_i) \text{ subject to } \sum_{i=1}^{n} a_{ii} \le nP.$$

Gallager: Info Theory & Reliable communication (1968 version) ③

p (406) (Ch 8)

" Thus if we use set of functions over $\left(-\frac{T}{2}, \frac{T}{2}\right)$ with $2WT(1+\epsilon)$ degrees of freedom, some of these functions will have vanishingly small fraction of their energy in $-W \le f \le W$ for WT arbitrarily large.

Conversely with $2WT(1-\epsilon)$ degrees of freedom, the minimum fraction of energy in the band $[-W, W]$ over all ~~class~~ functions in the class will approach 1 if WT is arbitrarily large".

Continuation of colored Guassian noise channel capacity

$$\max \prod_{i=1}^{n} (a_{ii} + \lambda_i)$$

$$\sum_{i=1}^{n} a_{ii} \le nP$$

$$\Leftrightarrow \left(\max \underbrace{\sum_{i=1}^{n} \ln (a_{ii} + \lambda_i)}_{= f(a_{11}, \ldots a_{nn})} \text{ subject to } \sum_{i=1}^{n} a_{ii} \le nP \right.$$

NASC (Necessary & Sufficient condition)

$$\frac{\partial}{\partial a_{ii}} f \begin{cases} = v & a_{ii} \ne 0 \\ \le v & a_{ii} = 0. \end{cases} \qquad \therefore \begin{cases} \frac{1}{a_{ii} + \lambda_i} = v & a_{ii} \ne 0. \\ \frac{1}{a_{ii} + \lambda_i} \le v & a_{ii} = 0. \end{cases}$$

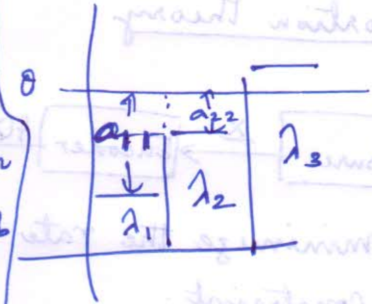
$$\Rightarrow \boxed{\begin{array}{ll} a_{ii} + \lambda_i = \frac{1}{v} = \theta \text{ (say)} & \text{when } a_{ii} \ne 0 \\ a_{ii} + \lambda_i \ge \theta & a_{ii} = 0 \end{array}}$$

Satisfies by setting $\boxed{a_{ii} = (\theta - \lambda_i)^+} \to$ ① .

with θ choosen to ensure that $\sum_{i=1}^{n} a_{ii} = nP$.

Thm (4.4.1, P.87 Gallager)

Let $f(\underline{\alpha})$ be a concave function of $\underline{\alpha} = [\alpha_1, \ldots \alpha_k]^t$ over a region R, $\alpha_i \geq 0$, $\sum_{i=1}^{k} \alpha_i = 1$ ($\underline{\alpha}$ is a prob vector)



Assume that the partial derivatives $\left\{ \frac{d f(\alpha)}{\partial \alpha_i} \right\}$ are defined and continuos over the region R. Then a NASC on a prob vector $\underline{\alpha}$ to maximize $f(\underline{\alpha})$ over R is that

$$\frac{\partial f(\alpha)}{\partial \alpha_i} = \lambda \quad \text{ist } \alpha_i > 0$$

$$\leq \lambda \quad \text{all } i \text{ st } \alpha_i = 0.$$

λ st $\sum_{i=1}^{k} \alpha_i = 1$

Here $\sum_{i=1}^{n} \ln(a_{ii} + \lambda_i)$ is a concave function of the probability vector $\left(\frac{a_{11}}{nP}, \frac{a_{22}}{nP}, \ldots \frac{a_{nn}}{nP}\right)$. This proves the waterfilling result.

$$\underset{\substack{\text{information} \\ \text{capacity}}}{C} = \prod_{i=1}^{n} (a_{ii} + \lambda_i)$$

where $\{a_{ii}\}$ are chosen according to ①.

$$Q^t K_x Q = A = \begin{bmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{bmatrix}.$$

and $K_x = Q A Q^t$

Rate distortion theory

$$\boxed{\text{Source}} \xrightarrow{\underline{X}} \boxed{\text{Encoder}} \xrightarrow{f(\hat{\underline{X}})}$$

Goal: minimize the rate of $f(\hat{X})$, given a distortion constraint.

Assume $\underline{X}$ is a RV taking on values in alphabet $\mathcal{X}_0$ (assume finite for now)

$$f(\underline{X}) \in \{1, 2, \ldots 2^{nR}\}$$
(index)

$$f(\underline{x}) \longrightarrow \boxed{\text{Decoder}} \xrightarrow{\hat{\underline{x}}}$$

Distortion function d:

$$d : \mathcal{X}_0 \times \hat{\mathcal{X}}_0 \longrightarrow R^+$$

$d(x, \hat{x})$ is a measure of the distortion associated with replacing x by $\hat{x}$.

Hamming distance function

$\mathcal{X}_0 = \{0, 1\}$

$$d(x, \hat{x}) = \begin{cases} 1 & x \neq \hat{x} \\ 0 & \text{else} \end{cases}$$

What is $E[d(\mathbf{x}, \hat{\mathbf{x}})]$
$= P(X \neq \hat{X}) = P_e$

Continuous sources

$$d(x, \hat{x}) = (x - \hat{x})^2.$$

Extension to sequences

$$d(\underline{x}, \hat{\underline{x}}) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) \qquad \underline{x} \in \mathcal{X}_0^n$$

A distortion measure is said to be bounded if

$$d_{max} = \max_{x \in \mathcal{X}_0, \hat{x} \in \hat{\mathcal{X}}_0} d(x, \hat{x}) < \infty.$$

Typically $\mathcal{X} = \hat{\mathcal{X}}$.

Defn: A $(2^{nR}, n)$ rate distortion code consists of an encoding function $f_n : \mathcal{X}^n \longrightarrow \{1, 2, \ldots 2^{nR}\}$ and a decoding function

$$g_n : \{1, 2, \ldots 2^{nR}\} \longrightarrow \mathcal{X}^n$$

The distortion associated with the code is given by

$$D = E\left[d\left[\underline{X}, \underbrace{g_n(f_n(\underline{X}))}_{\hat{\underline{X}}} \right]\right]$$

Codebook:

$$g_n(1), \quad \ldots \quad g_n(2^{nR})$$
$$\hat{\underline{X}}(1) \quad - - - \quad \hat{\underline{X}}(2^{nR})$$

Assignment regions:

$$f_n^{-1}(1), \quad f_n^{-1}(2), \ldots \quad f_n^{-1}(2^{nR})$$

$$\hat{\underline{X}}(w) \Rightarrow \begin{cases} \text{vector quantization} \\ \text{Source code} \\ \text{estimate etc} \end{cases}$$