Chapter 11
Continuous-Valued Channels

Department of Information Engineering
The Chinese University of Hong Kong

Preamble

- In a physical communication system, the input and output of a channel often take continuous real values.
- A waveform channel is one which takes transmission in continuous time.

11.1 Discrete-Time Channels

Definition 11.1 Let $f(y|x)$ be a conditional pdf defined for all x, where

$$-\int_{\mathcal{S}_Y(x)} f(y|x) \log f(y|x) dy < \infty$$

for all x. A discrete-time continuous channel $f(y|x)$ is a system with input random variable X and output random variable Y such that Y is related to X through $f(y|x)$ (cf. Definition 10.22).

Remark The integral in Definition 11.1 is precisely the conditional differential entropy $h(Y|X = x)$, which is required to be finite.

Definition 11.2 Let $\alpha : \Re \times \Re \to \Re$, and Z be a real random variable, called the noise variable . A discrete-time continuous channel (α, Z) is a system with a real input and a real output. For any input random variable X, the noise random variable Z is independent of X, and the output random variable Y is given by

$$Y = \alpha(X, Z).$$

Definition 11.3 Two continuous channels $f(y|x)$ and (α, Z) are equivalent if for every input distribution $F(x)$,

$$\Pr\{\alpha(X,Z) \leq y, X \leq x\} = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{Y|X}(v|u) dv \, dF_X(u)$$

for all x and y.

Remarks

1. Definition 11.2 is more general than Definition 11.1 because the former does not require the existence of $f(y|x)$.
2. We confine our discussion to channels defined by Definition 11.1.

Definition 11.4 (CMC I) A continuous memoryless channel (CMC) $f(y|x)$ is a sequence of replicates of a generic continuous channel $f(y|x)$. These continuous channels are indexed by a discrete-time index i, where $i \geq 1$, with the ith channel being available for transmission at time i. Transmission through a channel is assumed to be instantaneous. Let X_i and Y_i be respectively the input and the output of the CMC at time i, and let T_{i-} denote all the random variables that are generated in the system before X_i. The Markov chain $T_{i-} \to X_i \to Y_i$ holds, and

$$\Pr\{Y_i \leq y, X_i \leq x\} = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{Y|X}(v|u) dv \, dF_{X_i}(u).$$

Definition 11.5 (CMC II) A continuous memoryless channel (α, Z) is a sequence of replicates of a generic continuous channel (α, Z). These continuous channels are indexed by a discrete-time index i, where $i \geq 1$, with the ith channel being available for transmission at time i. Transmission through a channel is assumed to be instantaneous. Let X_i and Y_i be respectively the input and the output of the CMC at time i, and let T_{i-} denote all the random variables that are generated in the system before X_i. The noise variable Z_i for the transmission at time i is a copy of the generic noise variable Z, and is independent of (X_i, T_{i-}). The output of the CMC at time i is given by

$$Y_i = \alpha(X_i, Z_i).$$

Definition 11.6 Let κ be a real function. An average input constraint (κ, P) for a CMC is the requirement that for any codeword $(x_1, x_2, \cdots, x_n)$ transmitted over the channel,

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

Definition 11.7 The capacity of a continuous memoryless channel $f(y|x)$ with input constraint (κ, P) is defined as

$$C(P) = \sup_{F(x): E\kappa(X) \leq P} I(X;Y)$$

Theorem 11.8 $C(P)$ is non-decreasing, concave, and left-continuous.

Proof

1. Non-decreasing – Immediate.
2. Concave – A consequence of the the concavity of mutual information with respect to the input distribution.
3. Left-continuous – A consequence of concavity.

Remarks

1. $C(P)$ is also right-continous (a consequence of concavity) but requires a separate proof.
2. This property of $C(P)$ is not used in this chapter.

11.2 The Channel Coding Theorem

Definition 11.9 An (n, M) code for a continuous memoryless channel with input constraint (κ, P) is defined by an encoding function

$$e : \{1, 2, \cdots, M\} \rightarrow \Re^n$$

and a decoding function

$$g : \Re^n \rightarrow \{1, 2, \cdots, M\}.$$

The set $\{1, 2, \cdots, M\}$, denoted by $\mathcal{W}$, is called the message set. The sequences $e(1), e(2), \cdots, e(M)$ in $\Re^n$ are called codewords, and the set of codewords is called the codebook. Moreover,

$$\frac{1}{n}\sum_{i=1}^{n} \kappa(x_i(w)) \leq P \quad \text{for } 1 \leq w \leq M,$$

where $e(w) = (x_1(w), x_2(w), \cdots, x_n(w))$, i.e., each codeword satisfies the input power constraint.

Assumptions and Notations

- W is randomly chosen from the message set $\mathcal{W}$, so $H(W) = \log M$.
- $\mathbf{X} = (X_1, X_2, \cdots, X_n)$; $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)$
- Thus $\mathbf{X} = e(W)$.
- Let $\hat{W} = g(\mathbf{Y})$ be the estimate on the message W by the decoder.

Error Probabilities

Definition 11.10 For all $1 \leq w \leq M$, let

$$\lambda_w = \Pr\{\hat{W} \neq w | W = w\} = \int_{\{\mathbf{y} \in \mathcal{Y}^n : g(\mathbf{y}) \neq w\}} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|e(w)) d\mathbf{y}$$

be the conditional probability of error given that the message is w.

Definition 11.11 The maximal probability of error of an (n, M) code is defined as

$$\lambda_{max} = \max_w \lambda_w.$$

Definition 11.12 The average probability of error of an (n, M) code is defined as

$$P_e = \Pr\{\hat{W} \neq W\}.$$

Definition 11.13 A rate R is (asymptotically) achievable for a continuous memoryless channel if for any $\epsilon > 0$, there exists for sufficiently large n an (n, M) code such that

$$\frac{1}{n} \log M > R - \epsilon$$

and

$$\lambda_{max} < \epsilon$$

Theorem 11.14 A rate R is achievable for a continuous memoryless channel if and only if $R \leq C$, the capacity of the channel.

11.3.1 The Converse

- First establish the Markov chain $W \to \mathbf{X} \to \mathbf{Y} \to \hat{W}$.
- $W, \hat{W}$ – discrete
- $\mathbf{X}$ – real but discrete
- $\mathbf{Y}$ – real and continuous

Lemma 11.15 (Data Processing Theorem)

$$I(W; \hat{W}) \leq I(\mathbf{X}; \mathbf{Y})$$

Converse Proof

- Let R be an achievable rate, i.e., for any $\epsilon > 0$, there exists for sufficiently large n and (n, M) code such that

$$\frac{1}{n}\log M > R - \epsilon \quad \text{and} \quad \lambda_{max} < \epsilon.$$

- Consider

$$\begin{aligned}
\log M &= H(W) \\
&= H(W|\hat{W}) + I(W;\hat{W}) \\
&\leq H(W|\hat{W}) + I(\mathbf{X};\mathbf{Y}) \\
&= H(W|\hat{W}) + h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) \\
&\leq H(W|\hat{W}) + \sum_{i=1}^{n} h(Y_i) - h(\mathbf{Y}|\mathbf{X}) \\
&= H(W|\hat{W}) + \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Y_i|X_i) \\
&= H(W|\hat{W}) + \sum_{i=1}^{n} I(X_i;Y_i)
\end{aligned}$$

- Let V be a mixing random variable distributed uniformly on $\{1, 2, \cdots, n\}$ which is independent of X_i, $1 \leq i \leq n$.
- Let $X = X_V$ and Y be the output of the channel with X being the input.
- Then

$$\begin{aligned}
E\kappa(X) &= EE[\kappa(X)|V] \\
&= \sum_{i=1}^{n} \Pr\{V = i\} E[\kappa(X)|V = i] \\
&= \sum_{i=1}^{n} \Pr\{V = i\} E[\kappa(X_i)|V = i] \\
&= \sum_{i=1}^{n} \frac{1}{n} E\kappa(X_i) \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} \kappa(X_i)\right] \\
&\leq P
\end{aligned}$$

- By the concavity of mutual information with respect to the input distribution,
$$\frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i) \le I(X;Y) \le C$$
- The second inequality above follows because X satisfies $E\kappa(X) \le P$ as shown.
- It follows that
$$\log M \le H(W|\hat{W}) + nC$$
- The proof is completed by invoking Fano's inequality.

11.3.2 Achievability

Remarks

1. In the formula
$$C(P) = \sup_{F(x): E\kappa(X) \le P} I(X;Y)$$
X may not have a pdf, so it is difficult to consider sequences typical w.r.t. $F(x)$.

2. Need a new notion of joint typicality.

3. Recall that for any input distribution $F(x)$, $f(y)$ exists as long as $f(y|x)$ exists. Hence
$$I(X;Y) = E\left[\log \frac{f(y|x)}{f(y)}\right]$$

Mutual Typicality

Definition 11.16 The mutually typical set $\Psi^n_{[XY]\delta}$ with respect to $F(x,y)$ is the set of $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ such that

$$\left| \frac{1}{n} \log \frac{f(\mathbf{y}|\mathbf{x})}{f(\mathbf{y})} - I(X;Y) \right| \leq \delta,$$

where

$$f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} f(y_i|x_i)$$

and

$$f(\mathbf{y}) = \prod_{i=1}^{n} f(y_i),$$

and δ is an arbitrarily small positive number. A pair of sequences $(\mathbf{x},\mathbf{y})$ is called mutually δ-typical if it is in $\Psi^n_{[XY]\delta}$.

Proof

1.
$$\frac{1}{n}\log\frac{f(\mathbf{Y}|\mathbf{X})}{f(\mathbf{Y})} = \frac{1}{n}\log\prod_{i=1}^{n}\frac{f(Y_i|X_i)}{f(Y_i)} = \frac{1}{n}\sum_{i=1}^{n}\log\frac{f(Y_i|X_i)}{f(Y_i)}$$

2. By WLLN,
$$\frac{1}{n}\sum_{i=1}^{n}\log\frac{f(Y_i|X_i)}{f(Y_i)} \rightarrow E\log\frac{f(Y|X)}{f(Y)} = I(X;Y)$$
in probability.

Lemma 11.18 Let $(\mathbf{X}', \mathbf{Y}')$ be n i.i.d. copies of a pair of generic random variables (X', Y'), where X' and Y' are independent and have the same marginal distributions as X and Y, respectively. Then

$$\Pr\{(\mathbf{X}', \mathbf{Y}') \in \Psi^n_{[XY]\delta}\} \leq 2^{-n(I(X;Y)-\delta)}.$$

Proof

- $$\left|\frac{1}{n}\log\frac{f(\mathbf{y}|\mathbf{x})}{f(\mathbf{y})} - I(X;Y)\right| \leq \delta \quad \Rightarrow \quad \frac{1}{n}\log\frac{f(\mathbf{y}|\mathbf{x})}{f(\mathbf{y})} \geq I(X;Y) - \delta$$
$$\Rightarrow \quad f(\mathbf{y}|\mathbf{x}) \geq f(\mathbf{y})2^{n(I(X;Y)-\delta)}$$

- Then

$$\begin{aligned} 1 &\geq \Pr\{(\mathbf{X}, \mathbf{Y}) \in \Psi^n_{[XY]\delta})\} \\ &= \int \int_{\Psi^n_{[XY]\delta}} f(\mathbf{y}|\mathbf{x}) dF(\mathbf{x})\, d\mathbf{y} \\ &\geq 2^{n(I(X;Y)-\delta)} \int \int_{\Psi^n_{[XY]\delta}} f(\mathbf{y}) dF(\mathbf{x})\, d\mathbf{y} \\ &= 2^{n(I(X;Y)-\delta)} \Pr\{(\mathbf{X}', \mathbf{Y}') \in \Psi^n_{[XY]\delta}\} \end{aligned}$$

- Hence

$$\Pr\{(\mathbf{X}', \mathbf{Y}') \in \Psi^n_{[XY]\delta}\} \leq 2^{-n(I(X;Y)-\delta)}$$

Random Coding Scheme

- Since $C(P)$ is left-continuous, there exists $\gamma > 0$ such that

$$C(P-\gamma) > C(P) - \frac{\epsilon}{6}$$

- By the definition of $C(P-\gamma)$, there exists an input random variable X such that

$$E\kappa(X) \leq P - \gamma \quad \text{and} \quad I(X;Y) \geq C(P-\gamma) - \frac{\epsilon}{6}$$

- Choose for a sufficiently large n an even integer M satisfying

$$I(X;Y) - \frac{\epsilon}{6} < \frac{1}{n}\log M < I(X;Y) - \frac{\epsilon}{8}$$

- Then

$$\frac{1}{n}\log M > I(X;Y) - \frac{\epsilon}{6} \geq C(P-\gamma) - \frac{\epsilon}{3} > C(P) - \frac{\epsilon}{2}$$

The random coding scheme:

1. Construct the codebook $\mathcal{C}$ of an (n, M) code randomly by generating M codewords in $\Re^n$ independently and identically according to $F(x)^n$. Denote these codewords by $\tilde{\mathbf{X}}(1)$, $\tilde{\mathbf{X}}(2), \cdots, \tilde{\mathbf{X}}(M)$.

2. Reveal the codebook $\mathcal{C}$ to both the encoder and the decoder.

3. A message W is chosen from $\mathcal{W}$ uniformly.

4. The sequence $\mathbf{X} = \tilde{\mathbf{X}}(W)$ is transmitted through the channel.

5. The channel outputs a sequence $\mathbf{Y}$ according to

$$\Pr\{Y_i \leq y_i, 1 \leq i \leq n | \mathbf{X}(W) = \mathbf{x}\} = \prod_{i=1}^{n} \int_{-\infty}^{y_i} f(y|x_i) dy.$$

6. The sequence $\mathbf{Y}$ is decoded to the message w if $(\mathbf{X}(w), \mathbf{Y}) \in \Psi^n_{[XY]\delta}$ and there does not exist $w' \neq w$ such that $(\mathbf{X}(w'), \mathbf{Y}) \in \Psi^n_{[XY]\delta}$. Otherwise, $\mathbf{Y}$ is decoded to a constant message in $\mathcal{W}$. Denote by $\hat{W}$ the message to which $\mathbf{Y}$ is decoded.

Performance Analysis

- Let $\tilde{\mathbf{X}}(w) = (\tilde{X}_1(w), \tilde{X}_2(w), \cdots, \tilde{X}_n(w))$.
- Define the error event $Err = E_e \cup E_d$, where

$$E_e = \left\{ \frac{1}{n} \sum_{i=1}^{n} \kappa(\tilde{X}_i(W)) > P \right\} \quad \text{and} \quad E_d = \{\hat{W} \neq W\}$$

-

$$\begin{aligned} \Pr\{Err\} &= \Pr\{Err|W=1\} \\ &\leq \Pr\{E_e|W=1\} + \Pr\{E_d|W=1\} \end{aligned}$$

- Choose δ to be small to make

$$\Pr\{E_d|W=1\} \leq \frac{\epsilon}{4}$$

for sufficiently large n.

- By WLLN, for sufficiently large n,

$$\begin{aligned}
\Pr\{E_e|W=1\} &= \Pr\left\{\frac{1}{n}\sum_{i=1}^{n}\kappa(\tilde{X}_i(1)) > P \,\middle|\, W=1\right\} \\
&= \Pr\left\{\frac{1}{n}\sum_{i=1}^{n}\kappa(\tilde{X}_i(1)) > P\right\} \\
&= \Pr\left\{\frac{1}{n}\sum_{i=1}^{n}\kappa(\tilde{X}_i(1)) > (P-\gamma)+\gamma\right\} \\
&\le \Pr\left\{\frac{1}{n}\sum_{i=1}^{n}\kappa(\tilde{X}_i(1)) > E\kappa(X)+\gamma\right\} \\
&\le \frac{\epsilon}{4}
\end{aligned}$$

- So,

$$\Pr\{Err\} \le \frac{\epsilon}{2}$$

which implies for some codebook $\mathcal{C}^*$,

$$\Pr\{Err|\mathcal{C}^*\} \le \frac{\epsilon}{2}$$

- Rank the codewords in $\mathcal{C}^*$ in ascending order according to $\Pr\{Err|\mathcal{C}^*, W = w\}$.

- After discarding the worst half of the codewords in $\mathcal{C}^*$, if a codeword $\tilde{\mathbf{X}}(w)$ remains in $\mathcal{C}^*$, then
$$\Pr\{Err|\mathcal{C}^*, W = w\} \leq \epsilon$$

- However, it is not clear whether $\tilde{\mathbf{X}}(w)$ satisfies both $\lambda_w \leq \epsilon$ and the input power constraint.

- Since $Err = E_e \cup E_d$, we have
$$\lambda_w = \Pr\{E_d|\mathcal{C}^*, W = w\} \leq \epsilon$$
and
$$\Pr\{E_e|\mathcal{C}^*, W = w\} \leq \epsilon$$

- Observe that conditioning on $\{\mathcal{C}^*, W = w\}$, the codeword $\tilde{\mathbf{X}}(w)$ is deterministic, so either $\Pr\{E_e|\mathcal{C}^*, W = w\} = 0$ or 1. Therefore, $\Pr\{E_e|\mathcal{C}^*, W = w\} = 0$.

11.4 Memoryless Gaussian Channel

The Gaussian channel is the most commonly used model for a noisy channel with real input and output, because:

1. the Gaussian channel is highly analytically tractable
2. the Gaussian noise can be regarded as the worst kind of additive noise subject to a constraint on the noise power.

Definition 11.19 (Gaussian Channel) A Gaussian channel with noise energy N is a continuous channel with the following two equivalent specifications:

1. $f(y|x) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{(y-x)^2}{2N}}$.
2. $Z \sim \mathcal{N}(0, N)$ and $\alpha(X, Z) = X + Z$.

Definition 11.20 (Memoryless Gaussian Channel) A memoryless Gaussian channel with noise power N and input power constraint P is a memoryless continuous channel with the generic continuous channel being the Gaussian channel with noise energy N. The input power constraint P refers to the input constraint (κ, P) with $\kappa(x) = x^2$.

Theorem 11.21 (Capacity of a Memoryless Gaussian Channel) The capacity of a memoryless Gaussian channel with noise power N and input power constraint P is

$$\frac{1}{2}\log\left(1+\frac{P}{N}\right).$$

The capacity is achieved by the input distribution $\mathcal{N}(0, P)$.

Remarks

- The capacity of a memoryless Gaussian channel depends only on P/N, called the signal-to-noise ratio.
- The capacity is strictly positive no matter how small P/N is.
- The capacity is infinite if there is no input power constraint.

Lemma 11.22 Let $Y = X + Z$. Then $h(Y|X) = h(Z|X)$ provided that $f_{Z|X}(z|x)$ exists for all $x \in \mathcal{S}_X$.

Proof of Lemma 11.22

- $f_{Y|X}(y|x) = f_{Z|X}(y-x|x)$ exists.
- Then $h(Y|X=x)$ is defined, and

$$\begin{aligned} h(Y|X) &= \int h(Y|X=x)dF_X(x) \\ &= \int h(X+Z|X=x)dF_X(x) \\ &= \int h(x+Z|X=x)dF_X(x) \\ &= \int h(Z|X=x)dF_X(x) \\ &= h(Z|X) \end{aligned}$$

Proof of Theorem 11.21

- Let $F(x)$ be the CDF of the input random variable X such that $EX^2 \leq P$, where X is not necessarily continuous.
- Since $Z \sim \mathcal{N}(0, N)$, $f(y|x)$ and hence $f(y)$ exists.
- By Lemma 11.22,

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(Z|X) = h(Y) - h(Z)$$

- Since Z is independent of X and Z is zero-mean,

$$EY^2 = E(X+Z)^2 = EX^2 + EZ^2 \leq P + N$$

- By Theorem 10.43,

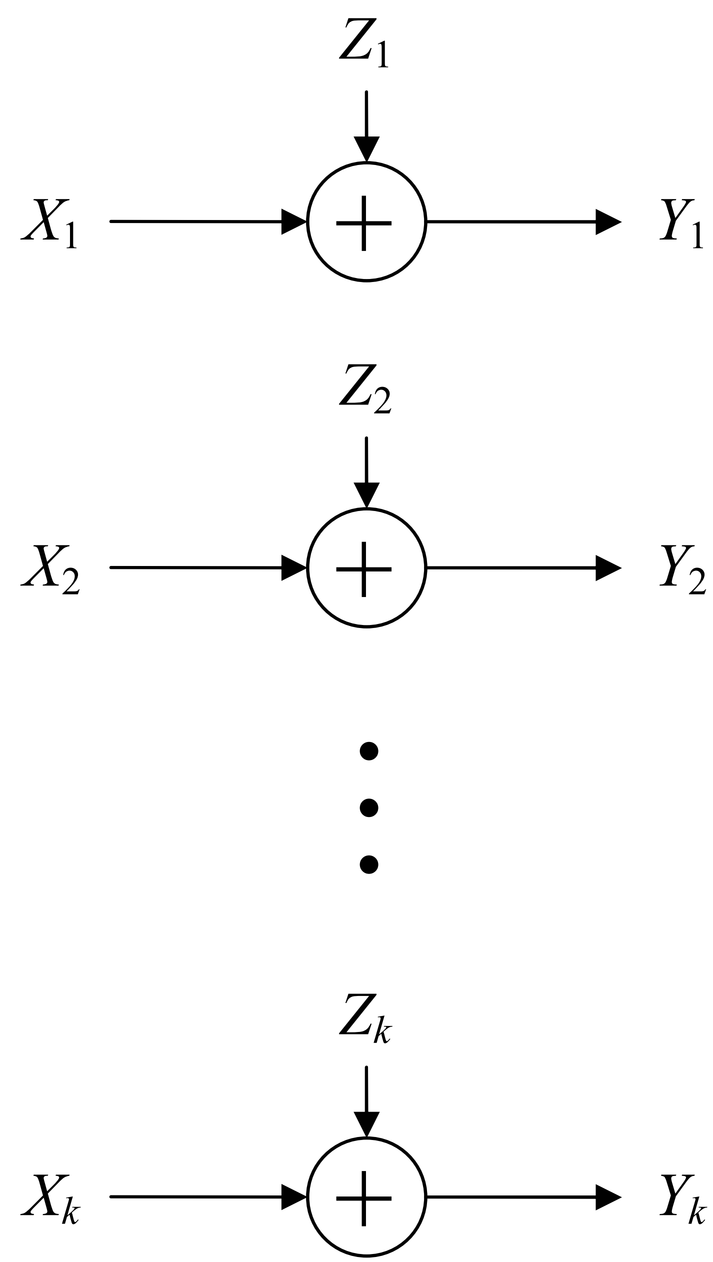
$$h(Y) \leq \frac{1}{2} \log[2\pi e(P+N)]$$

with equality if $Y \sim \mathcal{N}(0, P+N)$. This is achieved with $X \sim \mathcal{N}(0, P)$.

- Hence,

$$C = h(Y) - h(Z) = \frac{1}{2} \log[2\pi e(P+N)] - \frac{1}{2} \log(2\pi e N) = \frac{1}{2} \log\left(1 + \frac{P}{N}\right)$$

11.5 Parallel Gaussian Channels



- $Z_i \sim \mathcal{N}(0, N_i)$ and Z_i, $1 \le i \le k$ are independent.
- Total input power constraint: $E \sum_{i=1}^{k} X_i^2 \le P$.
- $$C(P) = \sup_{F(\mathbf{x}): E\sum_i X_i^2 \le P} I(\mathbf{X}; \mathbf{Y})$$
- Intuitively,

$$C(P) = \max_{P_1, P_2, \cdots, P_k : \sum_i P_i = P} \frac{1}{2} \sum_{i=1}^{k} \log\left(1 + \frac{P_i}{N_i}\right)$$

where $X_i \sim \mathcal{N}(0, P_i)$ and $X_1, X_2 \cdots, X_k$ are mutually independent.

Formal Justification:

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= h(\mathbf{Y}) - h(\mathbf{Z}) \\ &\le \sum_{i=1}^{k} h(Y_i) - \frac{1}{2}\sum_{i=1}^{k} h(Z_i) \\ &\le \frac{1}{2}\sum_{i=1}^{k} \log[2\pi e(EY_i^2)] - \frac{1}{2}\sum_{i=1}^{k} \log(2\pi e N_i) \\ &= \frac{1}{2}\sum_{i=1}^{k} \log(EY_i^2) - \frac{1}{2}\sum_{i=1}^{k} \log N_i \\ &= \frac{1}{2}\sum_{i=1}^{k} \log(EX_i^2 + EZ_i^2) - \frac{1}{2}\sum_{i=1}^{k} \log N_i \\ &= \frac{1}{2}\sum_{i=1}^{k} \log(P_i + N_i) - \frac{1}{2}\sum_{i=1}^{k} \log N_i \\ &= \frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{P_i}{N_i}\right) \end{aligned}$$

Maximization of $\sum_i \log(P_i + N_i)$

- Constraints: $\sum_i P_i \leq P$ and $P_i \geq 0$
- $\sum_i P_i \leq P$ can be replaced by $\sum_i P_i = P$ because $\log(P_i + N_i)$ is increasing in P_i.
- Ignore the constraints $P_i \geq 0$ for the time being. Use Lagrange multiplier to obtain

  $$P_i = \nu - N_i$$

  where the constant ν is chosen such that

  $$\sum_{i=1}^{k} P_i = \sum_{i=1}^{k} (\nu - N_i) = P$$

- This solution, which has a water-filling interpretation, would be a valid solution if $\nu \geq N_i$ so that $P_i \geq 0$ for all i.

Capacity of Parallel Gaussian Channels

By means of Proposition 11.23 (an application of the Karush-Kuhn-Tucker (KKT) condition), we obtain that in the general,

$$C(P) = \frac{1}{2} \sum_{i=1}^{k} \log \left(1 + \frac{P_i^*}{N_i} \right)$$

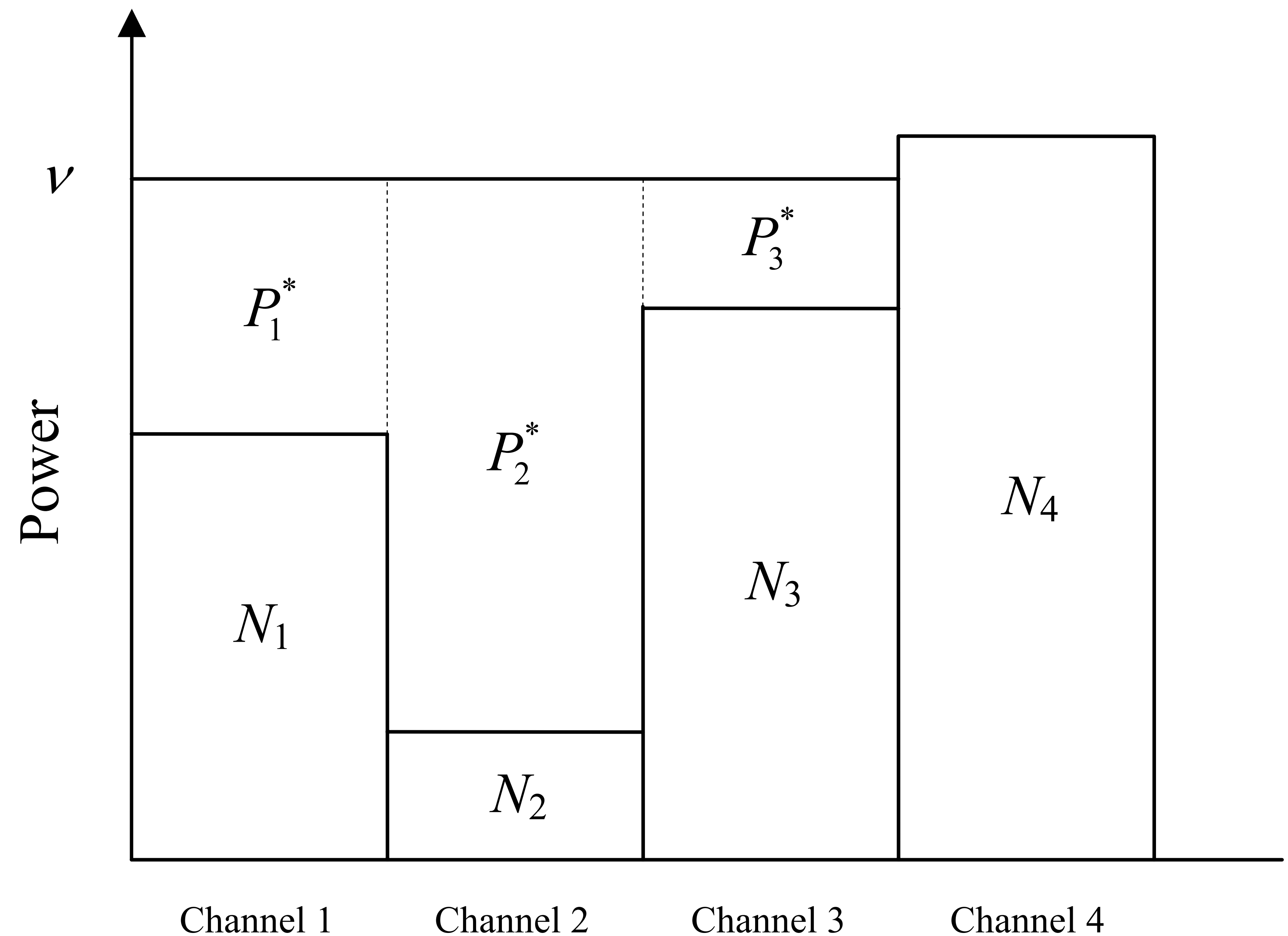
where $\{P_i^*, 1 \leq i \leq k\}$ is the optimal input power allocation among the channels given by

$$P_i^* = (\nu - N_i)^+, \quad 1 \leq i \leq k$$

with ν satisfying

$$\sum_{i=1}^{k} (\nu - N_i)^+ = P$$

Water-Filling



11.6 Correlated Gaussian Channels

- Same model as for parallel Gaussian channels except that $\mathbf{Z} \sim \mathcal{N}(0, K_{\mathbf{Z}})$.
- $EZ_i = 0$ for all i.
- The total input power constraint continues to be $E\sum_{i=1}^{k} X_i^2 \leq P$.
- The problem can be reduced to the problem of parallel Gaussian channels by decorrelating the noise vector.

Decorrelation of the Noise Vector

- Let $K_{\mathbf{Z}}$ be diagonalizable as $Q\Lambda Q^\top$ and consider

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}$$

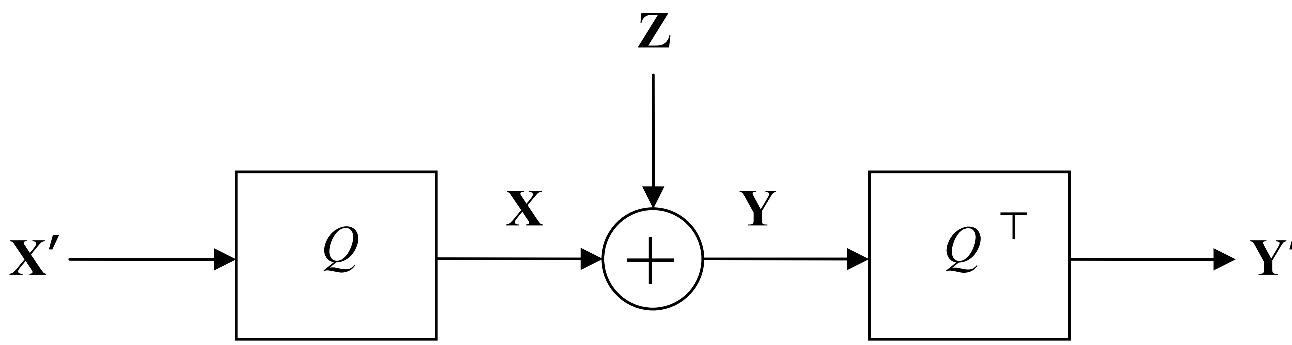
- Then

$$Q^\top \mathbf{Y} = Q^\top \mathbf{X} + Q^\top \mathbf{Z}$$

- Let $\mathbf{X}' = Q^\top \mathbf{X}$, $\mathbf{Y}' = Q^\top \mathbf{Y}$, and $\mathbf{Z}' = Q^\top \mathbf{Z}$ to obtain

$$\mathbf{Y}' = \mathbf{X}' + \mathbf{Z}'$$

- $K_{\mathbf{Z}'} = \Lambda$, $Z_i' \sim \mathcal{N}(0, \lambda_i)$, and Z_i', $1 \leq i \leq k$ are mutually independent.



- $\mathbf{X} = Q\mathbf{X}'$ and $\mathbf{Y}' = Q^\top \mathbf{Y}$ as prescribed.

- $\mathbf{Z}'$ is the equivalent noise vector, making it a system of parallel Gaussian channels.

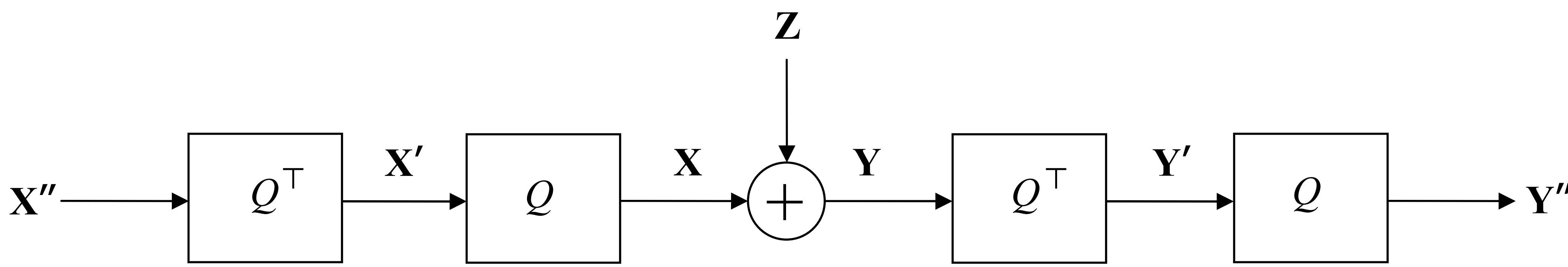
- The only difference between this system and the original system are the linear transformations Q and $Q^\top$ before and after the original system.

- By Proposition 10.9, the total input power constraint for the original system translates to the total input power constraint

$$E \sum_{i=1}^{k} (X_i')^2 \leq P$$

  for this system.

- Let the capacity of this system be C' and the capacity of the original system be C. Obviously, $C \geq C'$.

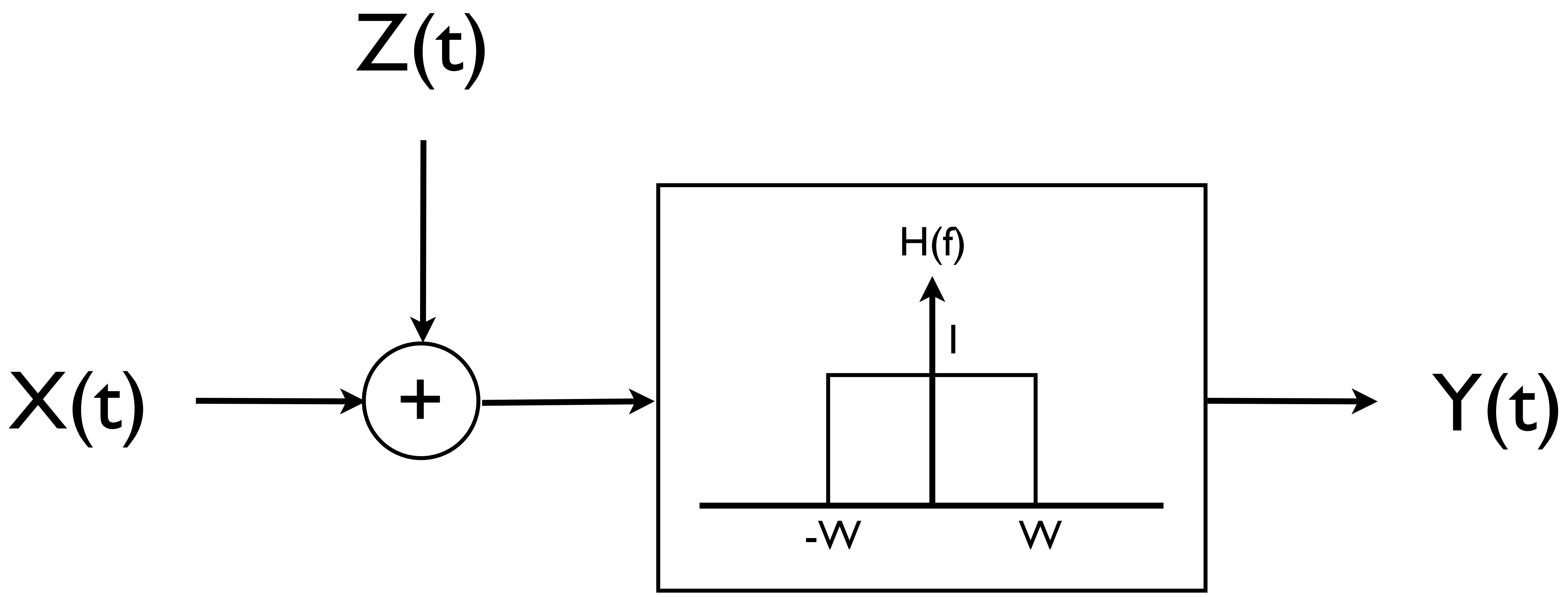


- Let the capacity of the above system be C''.
- Then, $C \geq C' \geq C''$.
- But since the above system is equivalent to the original system, $C'' = C$.
- Therefore, $C' = C$, or the equivalent system of parallel Gaussian channels is the same as the original system of correlated Gaussian channels.
- Hence, the capacity of the original system is given by

$$\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{a_i^*}{\lambda_i}\right)$$

where a_i^* is the optimal power allocated to the ith channel in the equivalent system, and its value can be obtained by water-filling.

11.7 The Bandlimited White Gaussian Channel



- Both input and output are in continuous time.
- $Z(t)$ is a zero-mean white Gaussian noise process with $S_Z(f) = \frac{N_0}{2}$, called an additive white Gaussian noise (AWGN).

Signal Analysis Preliminaries

Definition 11.24 The Fourier transform of a signal $g(t)$ is defined as

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi f t} dt.$$

The signal $g(t)$ can be recovered from $G(f)$ as

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi f t} df,$$

and $g(t)$ is called the inverse Fourier transform of $G(f)$. The functions $g(t)$ and $G(f)$ are said to form a transform pair, denoted by

$$g(t) \rightleftharpoons G(f).$$

The variables t and f are referred to as time and frequency, respectively.

Energy Signal:

- $$\int_{-\infty}^{\infty} |g(t)|^2 dt < \infty$$

- the Fourier transform of an energy signal exists.

Definition 11.25 Let $g_1(t)$ and $g_2(t)$ be a pair of energy signals. The cross-correlation function for $g_1(t)$ and $g_2(t)$ is defined as

$$R_{12}(\tau) = \int_{-\infty}^{\infty} g_1(t) g_2(t-\tau) dt$$

Proposition 11.26 For a pair of energy signals $g_1(t)$ and $g_2(t)$

$$R_{12}(\tau) \rightleftharpoons G_1(f) G_2^*(f),$$

where $G_2^*(f)$ denotes the complex conjugate of $G_2(f)$.

Definition 11.27 For a wide-sense stationary process $\{X(t), -\infty < t < \infty\}$, the autocorrelation function is defined as

$$R_X(\tau) = E[X(t+\tau)X(t)]$$

which does not depend on t, and the power spectral density is defined as

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f \tau} d\tau$$

i.e.,

$$R_X(\tau) \rightleftharpoons S_X(f)$$

Remark A process $X(t)$ is wide-sense stationary if $EX(t)$ does not depend on t and $E[X(t+\tau)X(t)]$ depends only on τ.

Let $\{(X(t), Y(t)), -\infty < t < \infty\}$ be a bivariate wide-sense stationary process. Their cross-correlation functions are defined as

$$R_{XY}(\tau) = E[X(t+\tau)Y(t)]$$

and

$$R_{YX}(\tau) = E[Y(t+\tau)X(t)]$$

which do not depend on t. The cross-spectral densities are defined as

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j2\pi f\tau} d\tau$$

and

$$S_{YX}(f) = \int_{-\infty}^{\infty} R_{YX}(\tau) e^{-j2\pi f\tau} d\tau$$

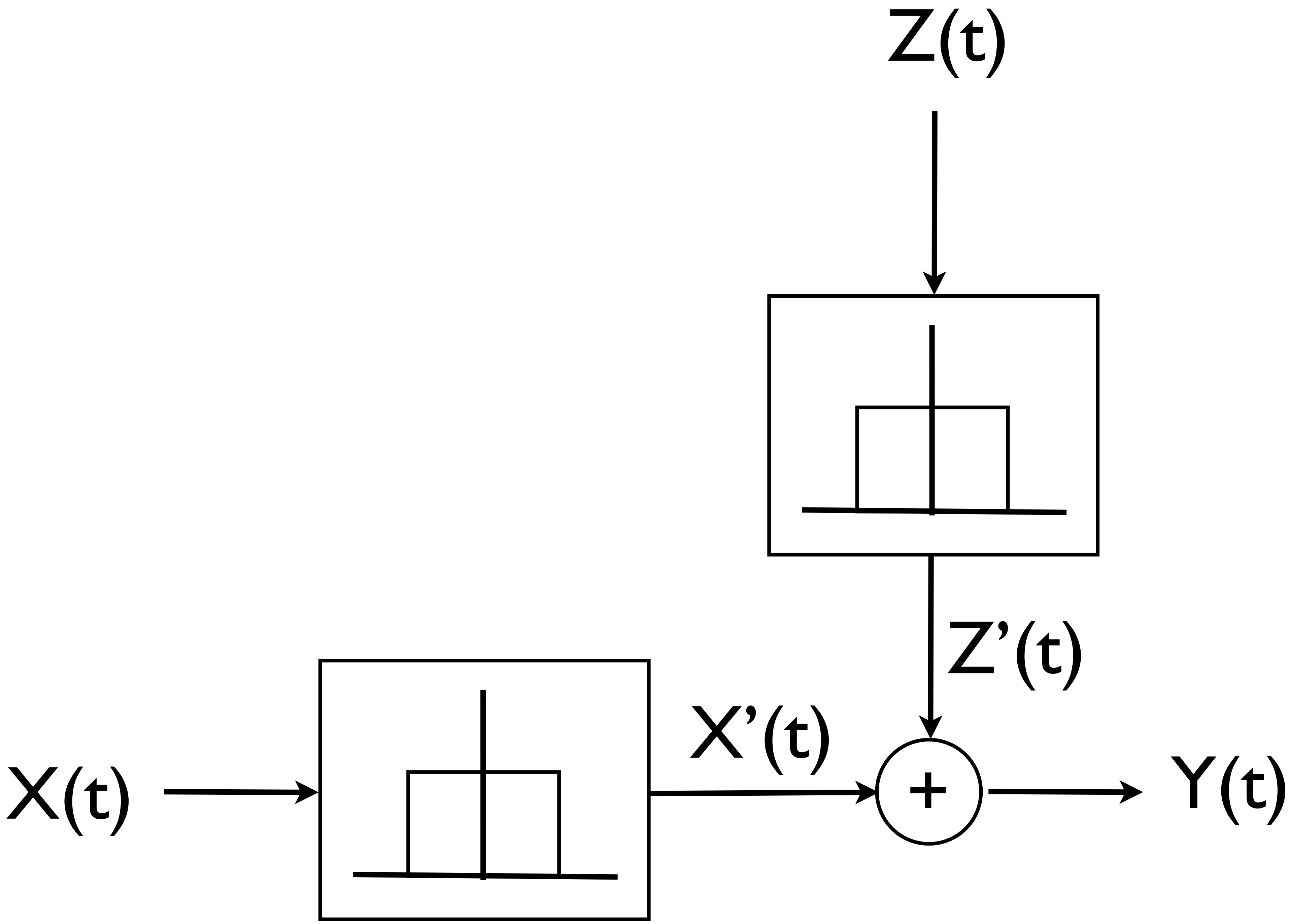
i.e.,

$$R_{XY}(\tau) \rightleftharpoons S_{XY}(f)$$

and

$$R_{YX}(\tau) \rightleftharpoons S_{YX}(f)$$

An Equivalent Model



- $Y'(t) = X'(t) + Z'(t)$
- $X'(t)$ and $Z'(t)$ are filtered versions of $X(t)$ and $Z(t)$, respectively.
- Both $X'(t)$ and $Z'(t)$ are bandlimited to $[0, W]$.
- Regard $X'(t)$ as the channel input and $Z'(t)$ as the additive noise process.
- Impose a power constraint on $X'(t)$.

Theorem 11.29 (Sampling Theorem) Let $g(t)$ be a signal with Fourier transform $G(f)$ that vanishes for $f \notin [-W, W]$. Then

$$g(t) = \sum_{i=-\infty}^{\infty} g\left(\frac{i}{2W}\right) \text{sinc}(2Wt - i)$$

for $-\infty < t < \infty$, where

$$\text{sinc}(t) = \frac{sin(\pi t)}{\pi t}$$

called the sinc function, is defined to be 1 at $t = 0$ by continuity.

Remarks

- $\text{sinc}(t) = 0$ for every integer $i \neq 0$.
- $\text{sinc}(2Wt - i) = \text{sinc}\left(2W\left(t - \frac{i}{2W}\right)\right) = 1$ for $t = \frac{i}{2W}$ and vanishes for $t = \frac{j}{2W}$ for every integer $j \neq i$.

- Let

$$g_i = \frac{1}{\sqrt{2W}} g\left(\frac{i}{2W}\right)$$

  and

$$\psi_i(t) = \sqrt{2W}\operatorname{sinc}(2Wt - i)$$

- Then

$$g(t) = \sum_{i=-\infty}^{\infty} g_i \psi_i(t)$$

Proposition 11.30 $\psi_i(t)$, $-\infty < i < \infty$ form an orthonormal basis for signals which are bandlimited to $[0, W]$.

Heuristic Treatment of the Bandlimited Channel

- Assume the input process $X'(t)$ has a Fourier transform, so that

$$X'(t) = \sum_{i=-\infty}^{\infty} X'_i \psi_i(t)$$

- There is a one-to-one correspondence between $\{X'(t)\}$ and $\{X_i\}$.
- Likewise, assume $Y'(t)$ can be written as

$$Y(t) = \sum_{i=-\infty}^{\infty} Y'_i \psi_i(t)$$

- With these assumptions, the waveform channel can be regarded as a discrete-time channel defined at $t = \frac{i}{2W}$, with the ith input and output of the channel being X'_i and Y_i, respectively.

To complete the model of the discrete-time channel, we need to

1. understand the effect of the noise process $Z'(t)$ on $Y(t)$ at the sampling points
2. relate the power constraint on X'_i to the power constraint on $X'(t)$.

Proposition 11.31 $Z'\left(\frac{i}{2W}\right)$, $-\infty < i < \infty$ are i.i.d. Gaussian random variables with zero mean and variance N_0W.

Proof

- $Z'(t)$ is a filtered version of $Z(t)$, so $Z'(t)$ is also a zero-mean Gaussian process.
- $Z'\left(\frac{i}{2W}\right)$, $-\infty < i < \infty$ are zero-mean Gaussian random variables.
- $$S_{Z'}(f) = \begin{cases} \frac{N_0}{2} & -W \leq f \leq W \\ 0 & \text{otherwise.} \end{cases}$$
- $$S_{Z'}(f) \rightleftharpoons R_{Z'}(\tau) = N_0 W \operatorname{sinc}(2W\tau)$$
- $$R_{Z'}\left(\frac{i}{2W}\right) = \begin{cases} 0 & i \neq 0 \\ N_0 W & i = 0 \end{cases}$$
- $Z'\left(\frac{i}{2W}\right)$, $-\infty < i < \infty$ are uncorrelated and hence independent because they are jointly Gaussian.
- Since $Z'\left(\frac{i}{2W}\right)$ has zero mean, its variance is given by $R_{Z'}(0) = N_0 W$.

- Recall that $Y(t) = X'(t) + Z'(t)$.
- Letting

$$Z_i' = \frac{1}{\sqrt{2W}} Z'\left(\frac{i}{2W}\right)$$

  we have

$$Y_i = X_i' + Z_i'$$

- Since $Z'\left(\frac{i}{2W}\right)$ are i.i.d. $\sim \mathcal{N}(0, N_0 W)$, Z_i' are i.i.d. $\sim \mathcal{N}(0, \frac{N_0}{2})$.
- So the bandlimited white Gaussian channel is equivalent to a memoryless Gaussian channel with noise power equal to $\frac{N_0}{2}$.

Relating the Power Constraints

- Let P' be the average energy (i.e., the second moment) of the X_i's.
- Since $\psi_i(t), -\infty < i < \infty$ are orthonormal, each has unit energy and their energy adds up.
- Therefore, $X'(t)$ accumulates energy from the samples at a rate equal to $2WP'$.
- Consider

$$2WP' \leq P$$

where P is the average power constraint on the input process $X'(t)$, we obtain

$$P' \leq \frac{P}{2W}$$

Capacity of the Bandlimited White Gaussian Channel

- $$\frac{1}{2}\log\left(1+\frac{P/2W}{N_0/2}\right)=\frac{1}{2}\log\left(1+\frac{P}{N_0W}\right) \quad \text{bits per sample}$$
- Since there are $2W$ samples per unit time, the capacity is

$$W\log\left(1+\frac{P}{N_0W}\right) \quad \text{bits per unit time}$$

- For the white Gaussian channel bandlimited to $[f_l, f_h]$, where f_l is a multiple of $W = f_h - f_l$, apply the bandpass version of the sampling theorem to obtain the same capacity formula.

11.8 The Bandlimited Colored Gaussian Channel

- Bandlimited to $[0, W]$ with input power constraint P.
- $Z(t)$ is a zero-mean additive colored Gaussian noise.
- Divide $[0, W]$ into k subintervals, each with width $\Delta_k = \frac{W}{k}$.
- Assume that the noise power over the ith subinterval is a constant $S_{Z,i}$.
- The capacity of the ith sub-channel is

$$\Delta_k \log\left(1 + \frac{P_i}{2S_{Z,i}\Delta_k}\right)$$

- The noise process $Z_i'(t)$ of the ith sub-channel is obtained by passing $Z(t)$ through the corresponding ideal bandpass filter.
- It can be shown (see Problem 9) that the noise processes $Z_i(t)$, $1 \leq i \leq k$ are independent.

- By sampling the channels in time, the k sub-channels can be regarded as a system of parallel Gaussian channels.
- Thus the channel capacity is equal to the sum of the capacities of the individual sub-channels when the power allocation among the k sub-channels is optimal.
- Let P_i^* be the optimal power allocation for the ith sub-channel.
- The channel capacity is equal to

$$\sum_{i=1}^{k} \Delta_k \log\left(1 + \frac{P_i^*}{2S_{Z,i}\Delta_k}\right) = \sum_{i=1}^{k} \Delta_k \log\left(1 + \frac{\frac{P_i^*}{2\Delta_k}}{S_{Z,i}}\right)$$

where By Proposition 11.23,

$$\frac{P_i^*}{2\Delta_k} = (\nu - S_{Z,i})^+$$

with

$$\sum_{i=1}^{k} P_i^* = P$$

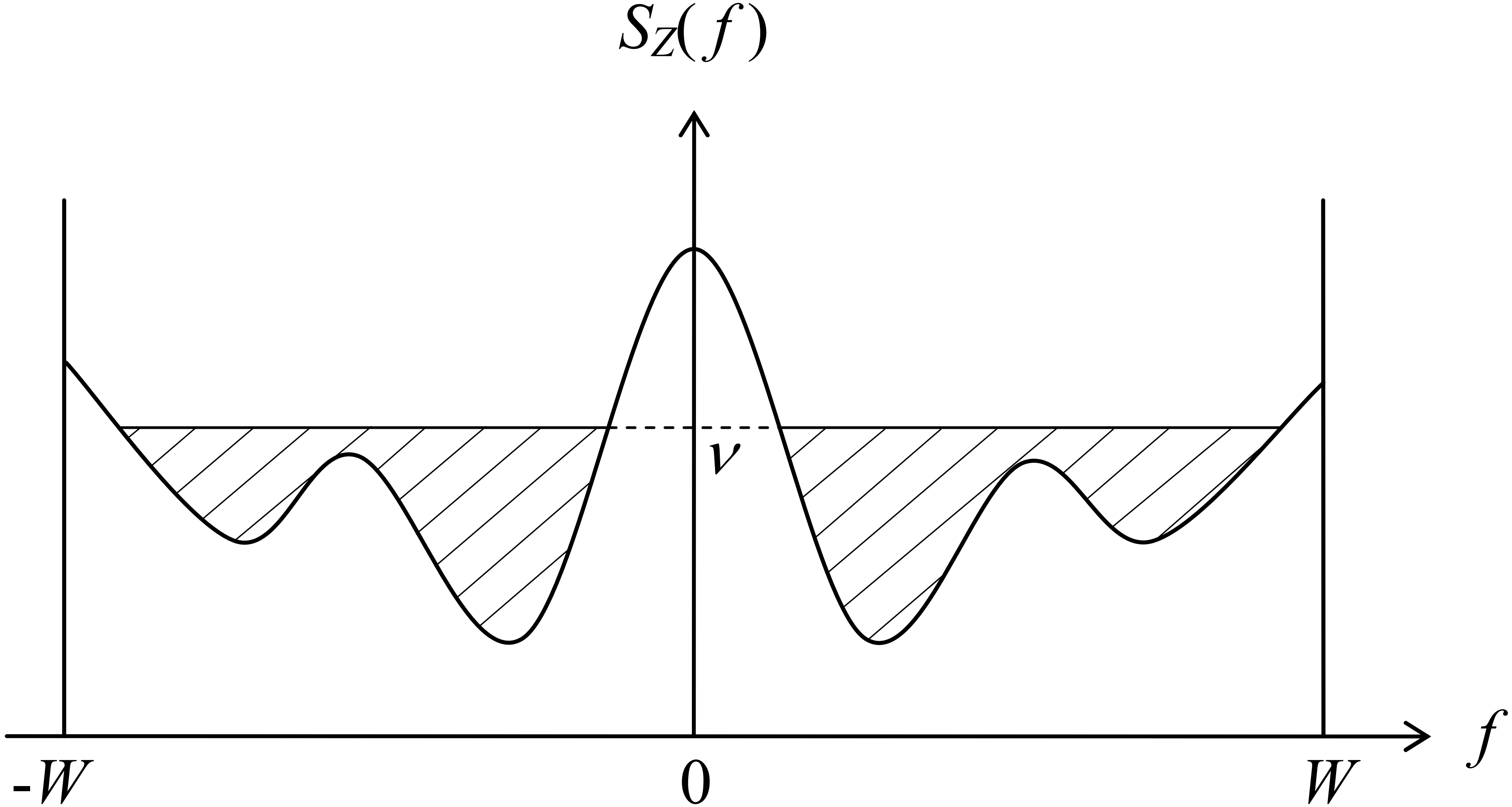
- As $k \to \infty$,

$$\sum_{i=1}^{k} \Delta_k \log\left(1 + \frac{\frac{P_i^*}{2\Delta_k}}{S_{Z,i}}\right)$$
$$\to \quad \frac{1}{2}\int_{-W}^{W} \log\left(1 + \frac{(\nu - S_Z(f))^+}{S_Z(f)}\right) df \quad \text{bits per unit time}$$

and

$$\sum_{i=1}^{k} P_i^* = P \quad \to \quad \int_{-W}^{W} (\nu - S_Z(f))^+ df = P$$

Water-Filling



11.9 Zero-Mean Noise is the Worst Additive Noise

- We will show that in terms of the capacity of the system, the zero-mean Gaussian noise is the worst additive noise given that the noise vector has a fixed correlation matrix.
- The diagonal elements of the correlation matrix specify the power of the individual noise variables.
- The other elements in the matrix give a characterization of the correlation between the noise variables.

Two Lemmas

Lemma 11.33 Let $\mathbf{X}$ be a zero-mean random vector and

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}$$

where $\mathbf{Z}$ is independent of $\mathbf{X}$. Then

$$\tilde{K}_{\mathbf{Y}} = \tilde{K}_{\mathbf{X}} + \tilde{K}_{\mathbf{Z}}$$

Remark The scalar case has been proved in the proof of Theorem 11.21.

Lemma 11.34 Let $\mathbf{Y}^* \sim \mathcal{N}(0, K)$ and $\mathbf{Y}$ be any random vector with correlation matrix K. Then

$$\int f_{\mathbf{Y}^*}(\mathbf{y}) \log f_{\mathbf{Y}^*}(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{S}_{\mathbf{Y}}} f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}^*}(\mathbf{y}) d\mathbf{y}.$$

Remark A similar technique has been used in proving Theorems 2.50 and 10.41 (maximum entropy distributions).

Theorem 11.32 For a fixed zero-mean Gaussian random vector $\mathbf{X}^*$, let

$$\mathbf{Y} = \mathbf{X}^* + \mathbf{Z},$$

where the joint pdf of $\mathbf{Z}$ exists and $\mathbf{Z}$ is independent of $\mathbf{X}^*$. Under the constraint that the correlation matrix of $\mathbf{Z}$ is equal to K, where K is any symmetric positive definite matrix, $I(\mathbf{X}^*;\mathbf{Y})$ is minimized if and only if $\mathbf{Z} \sim \mathcal{N}(0, K)$.

Proof

$$\begin{aligned}
& I(\mathbf{X}^*;\mathbf{Y}^*) - I(\mathbf{X}^*;\mathbf{Y}) \\
& \stackrel{a)}{=} \quad h(\mathbf{Y}^*) - h(\mathbf{Z}^*) - h(\mathbf{Y}) + h(\mathbf{Z}) \\
& = \quad -\int f_{\mathbf{Y}^*}(\mathbf{y}) \log f_{\mathbf{Y}^*}(\mathbf{y}) d\mathbf{y} + \int f_{\mathbf{Z}^*}(\mathbf{z}) \log f_{\mathbf{Z}^*}(\mathbf{z}) d\mathbf{z} \\
& \qquad + \int f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} - \int_{\mathcal{S}_{\mathbf{Z}}} f_{\mathbf{Z}}(\mathbf{z}) \log f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \\
& \stackrel{b)}{=} \quad -\int f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}^*}(\mathbf{y}) d\mathbf{y} + \int_{\mathcal{S}_{\mathbf{Z}}} f_{\mathbf{Z}}(\mathbf{z}) \log f_{\mathbf{Z}^*}(\mathbf{z}) d\mathbf{z} \\
& \qquad + \int f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} - \int_{\mathcal{S}_{\mathbf{Z}}} f_{\mathbf{Z}}(\mathbf{z}) \log f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}
\end{aligned}$$

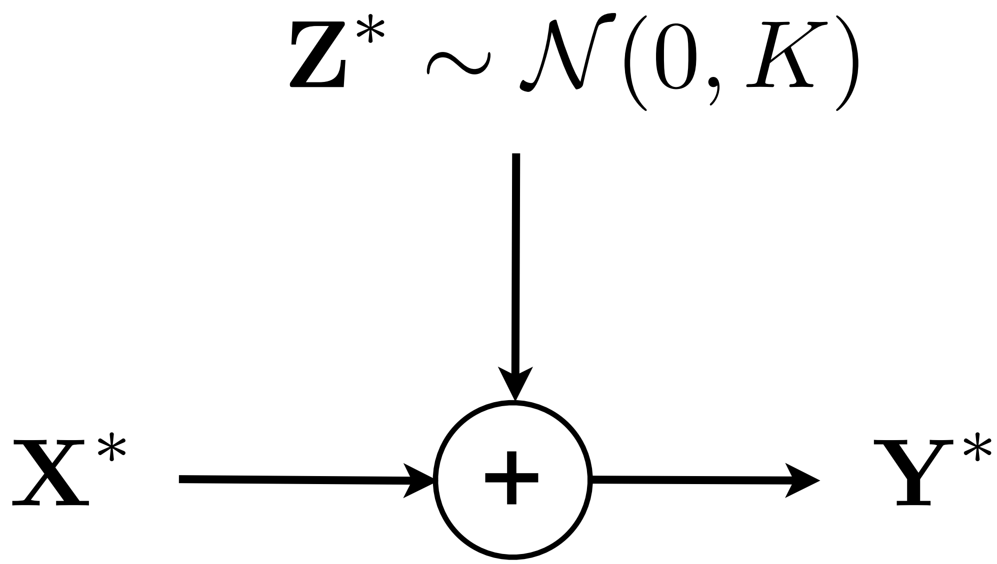
Proof (cont.)

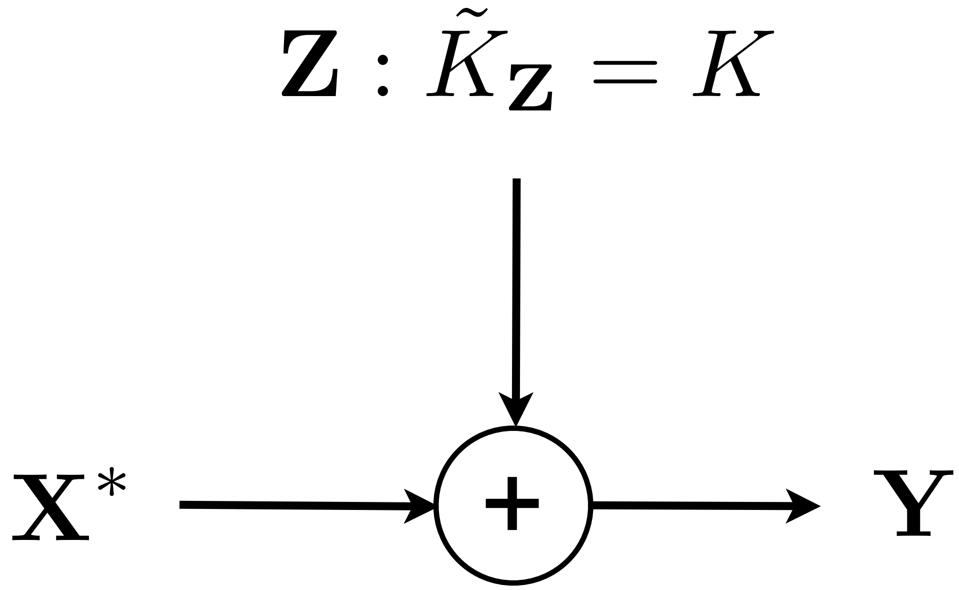
$$
\begin{aligned}
&= \int \log\left(\frac{f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{Y}^*}(\mathbf{y})}\right) f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} + \int_{\mathcal{S}_{\mathbf{Z}}} \log\left(\frac{f_{\mathbf{Z}^*}(\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})}\right) f_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} \\
&\overset{c)}{=} \int_{\mathcal{S}_{\mathbf{Z}}} \int \log\left(\frac{f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Z}^*}(\mathbf{z})}{f_{\mathbf{Y}^*}(\mathbf{y}) f_{\mathbf{Z}}(\mathbf{z})}\right) f_{\mathbf{YZ}}(\mathbf{y},\mathbf{z})d\mathbf{y}d\mathbf{z} \\
&\overset{d)}{\leq} \log\left(\int_{\mathcal{S}_{\mathbf{Z}}} \int \frac{f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{Z}^*}(\mathbf{z})}{f_{\mathbf{Y}^*}(\mathbf{y}) f_{\mathbf{Z}}(\mathbf{z})} f_{\mathbf{YZ}}(\mathbf{y},\mathbf{z})d\mathbf{y}d\mathbf{z}\right) \\
&\overset{e)}{=} \log\left(\int \left[\frac{1}{f_{\mathbf{Y}^*}(\mathbf{y})} \int_{\mathcal{S}_{\mathbf{Z}}} f_{\mathbf{X}^*}(\mathbf{y}-\mathbf{z}) f_{\mathbf{Z}^*}(\mathbf{z})d\mathbf{z}\right] f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y}\right) \\
&\overset{f)}{\leq} \log\left(\int \frac{f_{\mathbf{Y}^*}(\mathbf{y})}{f_{\mathbf{Y}^*}(\mathbf{y})} f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y}\right) \\
&= 0
\end{aligned}
$$

Gaussian is the Worst

- Consider a system of correlated Gaussian channels with noise vector $\mathbf{Z}^* \sim \mathcal{N}(0, K)$, and so $\tilde{K}_{\mathbf{Z}} = K$. Call this the zero-mean Gaussian system and let C^* be its capacity.
- Consider another system with exactly the same specification except that the noise vector $\mathbf{Z}$ may neither be zero-mean nor Gaussian. We require that the joint pdf of $\mathbf{Z}$ exists. Call this system as the alternative system and let C be its capacity.
- Let $\mathbf{X}^*$ be the zero-mean Gaussian input vector that achieves the capacity of the zero-mean Gaussian system.
- Let $\mathbf{Y}^*$ be the output of the zero-mean Gaussian system with $\mathbf{X}^*$ as input.
- Let $\mathbf{Y}$ be the output of the alternative system with $\mathbf{X}^*$ as input.
- Then

$$C \geq I(\mathbf{X}^*; \mathbf{Y}) \geq I(\mathbf{X}^*; \mathbf{Y}^*) = C^*$$


$\mathbf{Z}^* \sim \mathcal{N}(0, K)$
$\mathbf{X}^*$
+
$\mathbf{Y}^*$


$\mathbf{Z} : \tilde{K}_{\mathbf{Z}} = K$
$\mathbf{X}^*$
+
$\mathbf{Y}$