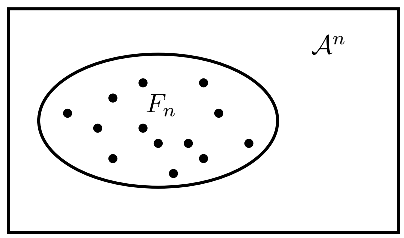
17.1 Channel coding with input constraints

Motivations: Let us look at the additive Gaussian noise. Then the Shannon capacity is infinite, since $\sup_{P_X} I(X; X+Z) = \infty$ achieved by $X \sim \mathcal{N}(0, P)$ and $P \to \infty$. But this is at the price of infinite second moment. In reality, limitation of transmission power $\Rightarrow$ constraints on the encoding operations $\Rightarrow$ constraints on input distribution.

Definition 17.1. An (n, M, ϵ)-code satisfies the input constraint $F_n \subset \mathcal{A}^n$ if the encoder is $f : [M] \to F_n$. (Without constraint, the encoder maps into $\mathcal{A}^n$).



Codewords all land in a subset of $\mathcal{A}^n$

Definition 17.2 (Separable cost constraint)**.** A channel with separable cost constraint is specified as follows:

1. $\mathcal{A}, \mathcal{B}$: input/output spaces
2. $P_{Y^n|X^n} : \mathcal{A}^n \to \mathcal{B}^n$, $n = 1, 2, \ldots$
3. Cost $\mathsf{c} : \mathcal{A} \to \bar{\mathbb{R}}$

Input constraint: average per-letter cost of a codeword x^n (with slight abuse of notation)

$$\mathsf{c}(x^n) = \frac{1}{n} \sum_{k=1}^{n} \mathsf{c}(x_k) \le P$$

Example: $\mathcal{A} = \mathcal{B} = \mathbb{R}$

- Average power constraint (separable):

$$\frac{1}{n} \sum_{i=1}^{n} |x_i|^2 \le P \quad \Leftrightarrow \quad \|x^n\|_2 \le \sqrt{nP}$$

- Peak power constraint (non-separable):

$$\max_{1 \le i \le n} |x_i| \le A \quad \Leftrightarrow \quad \|x^n\|_\infty \le A$$

Definition 17.3. Some basic definitions in parallel with the channel capacity without input constraint.

- A code is an (n, M, ϵ, P)-code if it is an (n, M, ϵ)-code satisfying input constraint $F_n \triangleq \{x^n : \frac{1}{n}\sum \mathsf{c}(x_k) \le P\}$
- Finite-n fundamental limits:

$$
\begin{aligned}
M^*(n, \epsilon, P) &= \max\{M : \exists (n, M, \epsilon, P)\text{-code}\} \\
M^*_{max}(n, \epsilon, P) &= \max\{M : \exists (n, M, \epsilon, P)_{max}\text{-code}\}
\end{aligned}
$$

- ϵ-capacity and Shannon capacity

$$
\begin{aligned}
C_\epsilon(P) &= \liminf_{n\to\infty} \frac{1}{n} \log M^*(n, \epsilon, P) \\
C(P) &= \lim_{\epsilon\downarrow 0} C_\epsilon(P)
\end{aligned}
$$

- Information capacity

$$
C_i(P) = \liminf_{n\to\infty} \frac{1}{n} \sup_{P_{X^n} : \mathbb{E}[\sum_{k=1}^n \mathsf{c}(X_k)] \le nP} I(X^n; Y^n)
$$

- Information stability: Channel is information stable if for all (admissible) P, there exists a sequence of channel input distributions P_{X^n} such that the following two properties hold:

$$
\frac{1}{n} i_{P_{X^n,Y^n}}(X^n; Y^n) \xrightarrow{i.P.} C_i(P) \tag{17.1}
$$

$$
\mathbb{P}[c(X^n) > P + \delta] \to 0 \qquad \forall \delta > 0\,. \tag{17.2}
$$

Note: These are the usual definitions, except that in $C_i(P)$, we are permitted to maximize $I(X^n; Y^n)$ using input distributions from the constraint set $\{P_{X^n} : \mathbb{E}[\sum_{k=1}^n \mathsf{c}(X_k)] \le nP\}$ instead of the distributions supported on F_n.

Definition 17.4 (Admissible constraint)**.** P is an admissible constraint if $\exists x_0 \in \mathcal{A}$ s.t. $\mathsf{c}(x_0) \le P \Leftrightarrow \exists P_X : \mathbb{E}[\mathsf{c}(X)] \le P$. The set of admissible P's is denoted by $\mathcal{D}_{\mathsf{c}}$, and can be either in the form (P_0, ∞) or $[P_0, \infty)$, where $P_0 \triangleq \inf_{x\in\mathcal{A}} \mathsf{c}(x)$.

Clearly, if $P \notin \mathcal{D}_{\mathsf{c}}$, then there is no code (even a useless one, with 1 codeword) satisfying the input constraint. So in the remaining we always assume $P \in \mathcal{D}_{\mathsf{c}}$.

Proposition 17.1. *Define* $f(P) = \sup_{P_X:\mathbb{E}[\mathsf{c}(X)]\le P} I(X; Y)$*. Then*

1. *f is concave and non-decreasing. The domain of f,* $\operatorname{dom} f \triangleq \{x : f(x) > -\infty\} = \mathcal{D}_{\mathsf{c}}$.
2. *One of the following is true: $f(P)$ is continuous and finite on (P_0, ∞), or $f = \infty$ on (P_0, ∞).*

Furthermore, both properties hold for the function $P \mapsto C_i(P)$.

Proof. In (1) all statements are obvious, except for concavity, which follows from the concavity of $P_X \mapsto I(X;Y)$. For any P_{X_i} such that $\mathbb{E}[\mathsf{c}(X_i)] \le P_i, i = 0, 1$, let $X \sim \bar{\lambda} P_{X_0} + \lambda P_{X_1}$. Then $\mathbb{E}[\mathsf{c}(X)] \le \bar{\lambda} P_0 + \lambda P_1$ and $I(X;Y) \ge \bar{\lambda} I(X_0;Y_0) + \lambda I(X_1;Y_1)$. Hence $f(\bar{\lambda} P_0 + \lambda P_1) \ge \bar{\lambda} f(P_0) + \lambda f(P_1)$. The second claim follows from concavity of $f(\cdot)$.

To extend these results to $C_i(P)$ observe that for every n

$$P \mapsto \frac{1}{n} \sup_{P_{X^n}:\mathbb{E}[\mathsf{c}(X^n)]\le P} I(X^n;Y^n)$$

is concave. Then taking $\liminf_{n\to\infty}$ the same holds for $C_i(P)$. □

An immediate consequence is that memoryless input is optimal for memoryless channel with separable cost, which gives us the single-letter formula of the information capacity:

Corollary 17.1 (Single-letterization). *Information capacity of stationary memoryless channel with separable cost:*

$$C_i(P) = f(P) = \sup_{\mathbb{E}[\mathsf{c}(X)]\le P} I(X;Y).$$

Proof. $C_i(P) \ge f(P)$ is obvious by using $P_{X^n} = (P_X)^n$. For "≤", use the concavity of $f(\cdot)$, we have that for any P_{X^n},

$$I(X^n;Y^n) \le \sum_{j=1}^n I(X_j;Y_j) \le \sum_{j=1}^n f(\mathbb{E}[\mathsf{c}(X_j)]) \le n f\Big(\frac{1}{n}\sum_{j=1}^n \mathbb{E}[\mathsf{c}(X_j)]\Big) \le n f(P).$$

□

17.2 Capacity under input constraint $C(P) \stackrel{?}{=} C_i(P)$

Theorem 17.1 (General weak converse)**.**

$$C_\epsilon(P) \le \frac{C_i(P)}{1-\epsilon}$$

Proof. The argument is the same as before: Take any (n, M, ϵ, P)-code, $W \to X^n \to Y^n \to \hat{W}$. Apply Fano's inequality, we have

$$-h(\epsilon) + (1-\epsilon)\log M \le I(W;\hat{W}) \le I(X^n;Y^n) \le \sup_{P_{X^n}:\mathbb{E}[\mathsf{c}(X^n)]\le P} I(X^n;Y^n) \le n f(P)$$

□

Theorem 17.2 (Extended Feinstein's Lemma)**.** *Fix a random transformation $P_{Y|X}$. $\forall P_X, \forall F \subset \mathcal{X}, \forall \gamma > 0, \forall M$, there exists an $(M,\epsilon)_{\max}$-code with:*

- *Encoder satisfies the input constraint: $f : [M] \to F \subset \mathcal{X}$;*
- *Probability of error bound:*

$$\epsilon P_X(F) \le \mathbb{P}[i(X;Y) < \log\gamma] + \frac{M}{\gamma}$$

Note: when $F = \mathcal{X}$, it reduces to the original Feinstein's Lemma.

Proof. Similar to the proof of the original Feinstein's Lemma, define the preliminary decoding regions $E_c = \{y : i(c; y) \geq \log\gamma\}$ for all $c \in \mathcal{X}$. Sequentially pick codewords $\{c_1, \dots, c_M\}$ **from the set** F and the final decoding region $\{D_1, \dots, D_M\}$ where $D_j \triangleq E_{c_j} \backslash \cup_{k=1}^{j-1} D_k$. The stopping criterion is that M is maximal, i.e.,

$$
\begin{aligned}
&\forall x_0 \in F, P_Y[E_{x_0} \backslash \cup_{j=1}^M D_j \big| X = x_0] < 1 - \epsilon \\
\Leftrightarrow\ &\forall x_0 \in \mathcal{X}, P_Y[E_{x_0} \backslash \cup_{j=1}^M D_j \big| X = x_0] < (1-\epsilon)\mathbf{1}[x_0 \in F] + \mathbf{1}[x_0 \in F^c] \\
\Rightarrow\ &\text{average over } x_0 \sim P_X,\ \mathbb{P}[\{i(X;Y) \geq \log\gamma\} \backslash \cup_{j=1}^M D_j] \leq (1-\epsilon)P_X(F) + P_X(F^c) = 1 - \epsilon P_X(F)
\end{aligned}
$$

From here, we can complete the proof by following the same steps as in the proof of Feinstein's lemma (Theorem 15.3). □

Theorem 17.3 (Achievability)**.** *For any information stable channel with input constraints and $P > P_0$ we have*

$$C(P) \geq C_i(P) \tag{17.3}$$

Proof. Let us consider a special case of the stationary memoryless channel (the proof for general information stable channel follows similarly). So we assume $P_{Y^n|X^n} = (P_{Y|X})^n$.

Fix $n \geq 1$. Since the channel is stationary memoryless, we have $P_{Y^n|X^n} = (P_{Y|X})^n$. Choose a P_X such that $\mathbb{E}[\mathsf{c}(X)] < P$, Pick $\log M = n(I(X;Y) - 2\delta)$ and $\log\gamma = n(I(X;Y) - \delta)$.

With the input constraint set $F_n = \{x^n : \frac{1}{n}\sum \mathsf{c}(x_k) \leq P\}$, and iid input distribution $P_{X^n} = P_X^n$, we apply the extended Feinstein's Lemma, there exists an $(n, M, \epsilon_n, P)_{\max}$-code with the encoder satisfying input constraint F and the error probability

$$\epsilon_n \underbrace{P_X(F)}_{\to 1} \leq \underbrace{P(i(X^n;Y^n) \leq n(I(X;Y) - \delta))}_{\to 0 \text{ as } n\to\infty \text{ by WLLN and stationary memoryless assumption}} + \underbrace{\exp(-n\delta)}_{\to 0}$$

Also, since $\mathbb{E}[\mathsf{c}(X)] < P$, by WLLN, we have $P_{X^n}(F_n) = P(\frac{1}{n}\sum \mathsf{c}(x_k) \leq P) \to 1$.

$$
\begin{aligned}
&\epsilon_n(1 + o(1)) \leq o(1) \\
\Rightarrow\ &\epsilon_n \to 0 \text{ as } n \to \infty \\
\Rightarrow\ &\forall\epsilon, \exists n_0, \text{ s.t. } \forall n \geq n_0, \exists(n, M, \epsilon_n, P)_{\max}\text{-code, with } \epsilon_n \leq \epsilon
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&C_\epsilon(P) \geq \frac{1}{n}\log M = I(X;Y) - 2\delta, \quad \forall\delta > 0, \forall P_X \text{ s.t. } \mathbb{E}[\mathsf{c}(X)] < P \\
\Rightarrow\ &C_\epsilon(P) \geq \sup_{P_X:\mathbb{E}[\mathsf{c}(X)]<P} \lim_{\delta\to 0}(I(X;Y) - 2\delta) \\
\Rightarrow\ &C_\epsilon(P) \geq \sup_{P_X:\mathbb{E}[\mathsf{c}(X)]<P} I(X;Y) = C_i(P-) = C_i(P)
\end{aligned}
$$

where the last equality is from the continuity of C_i on (P_0, ∞) by Proposition 17.1. Notice that for general information stable channel, we just need to use the definition to show that $P(i(X^n;Y^n) \leq n(C_i - \delta)) \to 0$, and all the rest follows. □

Theorem 17.4 (Shannon capacity)**.** *For an information stable channel with cost constraint and for any admissible constraint P we have*

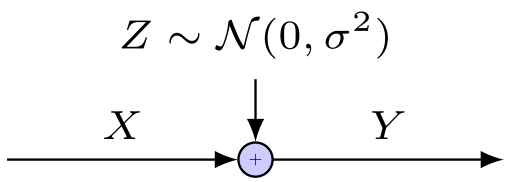
$$C(P) = C_i(P).$$

Proof. The case of $P = P_0$ is treated in the homework. So assume $P > P_0$. Theorem 17.1 shows $C_\epsilon(P) \le \frac{C_i(P)}{1-\epsilon}$, thus $C(P) \le C_i(P)$. On the other hand, from Theorem 17.3 we have $C(P) \ge C_i(P)$. $\square$

Note: In homework, you will show that $C(P_0) = C_i(P_0)$ also holds, even though $C_i(P)$ may be discontinuous at P_0.

17.3 Applications

17.3.1 Stationary AWGN channel



Definition 17.5 (AWGN)**.** The additive Gaussian noise (AWGN) channel is a stationary memoryless additive-noise channel with separable cost constraint: $\mathcal{A} = \mathcal{B} = \mathbb{R}$, $\mathsf{c}(x) = x^2$, $P_{Y|X}$ is given by $Y = X + Z$, where $Z \sim \mathcal{N}(0, \sigma^2) \perp\!\!\!\perp X$, and average power constraint $\mathbb{E}X^2 \le P$.

In other words, $Y^n = X^n + Z^n$, where $Z^n \sim \mathcal{N}(0, I_n)$.

Note: Here "white" = uncorrelated $\stackrel{\text{Gaussian}}{=}$ independent.
Note: Complex AWGN channel is similarly defined: $\mathcal{A} = \mathcal{B} = \mathbb{C}$, $\mathsf{c}(x) = |x|^2$, and $Z^n \sim \mathbb{C}\mathcal{N}(0, I_n)$

Theorem 17.5. *For stationary ($\mathbb{C}$)-AWGN channel, the channel capacity is equal to information capacity, and is given by:*

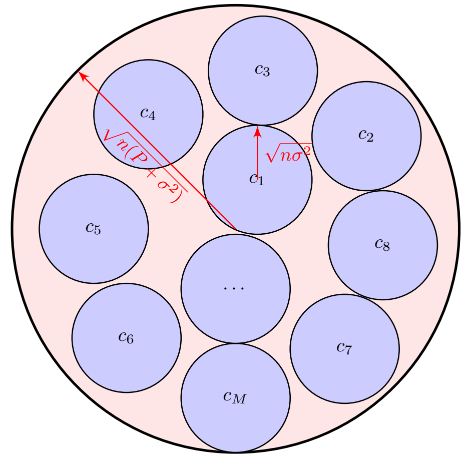
$$C(P) = C_i(P) = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right) \quad \textit{for AWGN}$$
$$C(P) = C_i(P) = \log\left(1 + \frac{P}{\sigma^2}\right) \quad \textit{for } \mathbb{C}\textit{-AWGN}$$

Proof. By Corollary 17.1,

$$C_i = \sup_{P_X : \mathbb{E}X^2 \le P} I(X; X + Z)$$

Then use Theorem 4.6 (Gaussian saddlepoint) to conclude $X \sim \mathcal{N}(0, P)$ (or $\mathbb{C}\mathcal{N}(0, P)$) is the unique caid. $\square$

Note: Since $Z^n \sim \mathcal{N}(0, \sigma^2)$, then with high probability, $\|Z^n\|_2$ concentrates around $\sqrt{n\sigma^2}$. Similarly, due the power constraint and the fact that $Z^n \perp\!\!\!\perp X^n$, the received vector Y^n lies in an ℓ_2-ball of radius $\sqrt{n(P+\sigma^2)}$. Since the noise can at most perturb the codeword by $\sqrt{n\sigma^2}$ in Euclidean distance, if we can pack M balls of radius $\sqrt{n\sigma^2}$ into the ℓ_2-ball of radius $\sqrt{n(P+\sigma^2)}$ centered at the origin, then this gives a good codebook and decision regions. The packing number is related to the volume ratio. Note that the volume of an ℓ_2-ball of radius r in $\mathbb{R}^n$ is given by $c_n r^n$ for some constant c_n. Then $\frac{c_n(n(P+\sigma^2))^{n/2}}{c_n(n\sigma^2)^{n/2}} = \left(1+\frac{P}{\sigma^2}\right)^{n/2}$. Take the log and divide by n, we get $\frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right)$.



Theorem 17.5 applies to Gaussian noise. What if the noise is non-Gaussian and how sensitive is the capacity formula $\frac{1}{2}\log(1+\text{SNR}$ to the Gaussian assumption? Recall the Gaussian saddlepoint result we have studied in Lecture 4 where we showed that for the same variance, Gaussian noise is the worst which shows that the capacity of any non-Gaussian noise is at least $\frac{1}{2}\log(1+\text{SNR})$. Conversely, it turns out the increase of the capacity can be controlled by how non-Gaussian the noise is (in terms of KL divergence). The following result is due to Ihara.

Theorem 17.6 (Additive Non-Gaussian noise)**.** *Let Z be a real-valued random variable independent of X and $\mathbb{E}Z^2 < \infty$. Let $\sigma^2 = \operatorname{Var} Z$. Then*

$$\frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right) \le \sup_{P_X:\mathbb{E}X^2\le P} I(X;X+Z) \le \frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right) + D(P_Z\|\mathcal{N}(\mathbb{E}Z,\sigma^2)).$$

Proof. Homework. □

Note: The quantity $D(P_Z\|\mathcal{N}(\mathbb{E}Z,\sigma^2))$ is sometimes called the *non-Gaussianness* of Z, where $\mathcal{N}(\mathbb{E}Z,\sigma^2)$ is a Gaussian with the same mean and variance as Z. So if Z has a non-Gaussian density, say, Z is uniform on $[0,1]$, then the capacity can only differ by a constant compared to AWGN, which still scales as $\frac{1}{2}\log\text{SNR}$ in the high-SNR regime. On the other hand, if Z is discrete, then $D(P_Z\|\mathcal{N}(\mathbb{E}Z,\sigma^2)) = \infty$ and indeed in this case one can show that the capacity is infinite because the noise is "too weak".

17.3.2 Parallel AWGN channel

Definition 17.6 (Parallel AWGN)**.** A parallel AWGN channel with L branches is defined as follows: $\mathcal{A} = \mathcal{B} = \mathbb{R}^L$; $\mathsf{c}(x) = \sum_{k=1}^L |x_k|^2$; $P_{Y^L|X^L} : Y_k = X_k + Z_k$, for $k = 1,\ldots,L$, and $Z_k \sim \mathcal{N}(0,\sigma_k^2)$ are independent for each branch.

Theorem 17.7 (Waterfilling)**.** *The capacity of L-parallel AWGN channel is given by*

$$C = \frac{1}{2}\sum_{j=1}^L \log^+ \frac{T}{\sigma_j^2}$$

where $\log^+(x) \triangleq \max(\log x, 0)$, and $T \ge 0$ is determined by

$$P = \sum_{j=1}^L |T-\sigma_j^2|^+$$

Proof.

$$\begin{aligned}
C_i(P) &= \sup_{P_{X^L}:\sum \mathbb{E}[X_i^2]\le P} I(X^L;Y^L) \\
&\le \sup_{\sum P_k\le P, P_k\ge 0} \sum_{k=1}^{L} \sup_{\mathbb{E}[X_k^2]\le P_k} I(X_k;Y_k) \\
&= \sup_{\sum P_k\le P, P_k\ge 0} \sum_{k=1}^{L} \frac{1}{2}\log(1+\frac{P_k}{\sigma_k^2})
\end{aligned}$$

with equality if $X_k \sim \mathcal{N}(0,P_k)$ are independent. So the question boils down to the last maximization problem – **power allocation**: Denote the Lagragian multipliers for the constraint $\sum P_k \le P$ by λ and for the constraint $P_k \ge 0$ by μ_k. We want to solve $\max \sum \frac{1}{2}\log(1+\frac{P_k}{\sigma_k^2}) - \mu_k P_k + \lambda(P - \sum P_k)$. First-order condition on P_k gives that

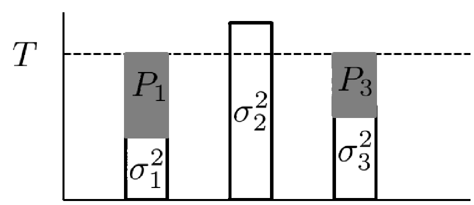
$$\frac{1}{2}\frac{1}{\sigma_k^2+P_k} = \lambda - \mu_k, \quad \mu_k P_k = 0$$

therefore the optimal solution is

$$P_k = |T-\sigma_k^2|^+, \quad T \text{ is chosen such that } P = \sum_{k=1}^{L}|T-\sigma_k^2|^+$$

□

Note: The figure illustrates the power allocation via water-filling. In this particular case, the second branch is too noisy (σ_2 too big) such that it is better be discarded, i.e., the assigned power is zero.



waterfilling across 3 parallel channels

Note: [Significance of the waterfilling theorem] In the high SNR regime, the capacity for 1 AWGN channel is approximately $\frac{1}{2}\log P$, while the capacity for L parallel AWGN channel is approximately $\frac{L}{2}\log(\frac{P}{L}) \approx \frac{L}{2}\log P$ for large P. This L-fold increase in capacity at high SNR regime leads to the powerful technique of spatial multiplexing in MIMO.

Also notice that this gain does not come from multipath diversity. Consider the scheme that a single stream of data is sent through every parallel channel simultaneously, with multipath diversity, the effective noise level is reduced to $\frac{1}{L}$, and the capacity is approximately $\log(LP)$, which is much smaller than $\frac{L}{2}\log(\frac{P}{L})$ for P large.

17.4* Non-stationary AWGN

Definition 17.7 (Non-stationary AWGN)**.** A non-stationary AWGN channel is defined as follows: $\mathcal{A} = \mathcal{B} = \mathbb{R}$, $\mathsf{c}(x) = x^2$, $P_{Y_j|X_j} : Y_j = X_j + Z_j$, where $Z_j \sim \mathcal{N}(0,\sigma_j^2)$.

Theorem 17.8. *Assume that for every T the following limits exist:*

$$\tilde{C}_i(T) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2} \log^+ \frac{T}{\sigma_j^2}$$

$$\tilde{P}(T) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} |T - \sigma_j^2|^+$$

then the capacity of the non-stationary AWGN channel is given by the parameterized form: $C(T) = \tilde{C}_i(T)$ with input power constraint $\tilde{P}(T)$.

Proof. Fix $T > 0$. Then it is clear from the waterfilling solution that

$$\sup I(X^n; Y^n) = \sum_{j=1}^{n} \frac{1}{2} \log^+ \frac{T}{\sigma_j^2}, \tag{17.4}$$

where supremum is over all P_{X^n} such that

$$\mathbb{E}[\mathsf{c}(X^n)] \le \frac{1}{n} \sum_{j=1}^{n} |T - \sigma_j^2|^+ . \tag{17.5}$$

Now, by assumption, the LHS of (17.5) converges to $\tilde{P}(T)$. Thus, we have that for every $\delta > 0$

$$C_i(\tilde{P}(T) - \delta) \le \tilde{C}_i(T) \tag{17.6}$$

$$C_i(\tilde{P}(T) + \delta) \ge \tilde{C}_i(T) \tag{17.7}$$

Taking $\delta \to 0$ and invoking continuity of $P \mapsto C_i(P)$, we get that the information capacity satisfies

$$C_i(\tilde{P}(T)) = \tilde{C}_i(T) .$$

The channel is information stable. Indeed, from (16.17)

$$\mathrm{Var}(i(X_j; Y_j)) = \frac{\log^2 e}{2} \frac{P_j}{P_j + \sigma_j^2} \le \frac{\log^2 e}{2}$$

and thus

$$\sum_{j=1}^{n} \frac{1}{n^2} \mathrm{Var}(i(X_j; Y_j)) < \infty .$$

From here information stability follows via Theorem 16.9. □

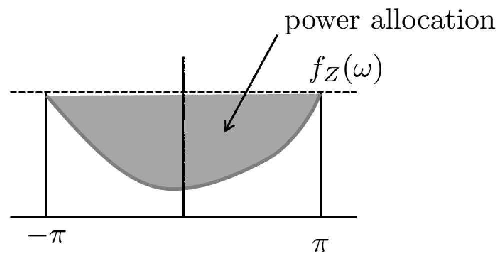
Note: Non-stationary AWGN is primarily interesting due to its relationship to the stationary Additive Colored Gaussian noise channel in the following discussion.

17.5* Stationary Additive Colored Gaussian noise channel

Definition 17.8 (Additive colored Gaussian noise channel). An Additive Colored Gaussian noise channel is defined as follows: $\mathcal{A} = \mathcal{B} = \mathbb{R}$, $\mathsf{c}(x) = x^2$, $P_{Y_j|X_j} : Y_j = X_j + Z_j$, where Z_j is a stationary Gaussian process with spectral density $f_Z(\omega) > 0, \omega \in [-\pi, \pi]$.

Theorem 17.9. *The capacity of stationary ACGN channel is given by the parameterized form:*

$$C(T) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{2}\log^+ \frac{T}{f_Z(\omega)} d\omega$$
$$P(T) = \frac{1}{2\pi}\int_0^{2\pi} \left|T - f_Z(\omega)\right|^+ d\omega$$



waterfilling across spectrum for stationary ACGN channel

Proof. Take $n \geq 1$, consider the diagonalization of the covariance matrix of Z^n:

$$Cov(Z^n) = \Sigma = U^* \widetilde{\Sigma} U, \text{ such that } \widetilde{\Sigma} = diag(\sigma_1, \dots, \sigma_n)$$

Since $Cov(Z^n)$ is positive semi-definite, U is a unitary matrix. Define $\widetilde{X}^n = UX^n$ and $\widetilde{Y}^n = UY^n$, the channel between $\widetilde{X}^n$ and $\widetilde{Y}^n$ is thus

$$\widetilde{Y}^n = \widetilde{X}^n + UZ^n,$$
$$Cov(UZ^n) = UCov(Z^n)U^* = \widetilde{\Sigma}$$

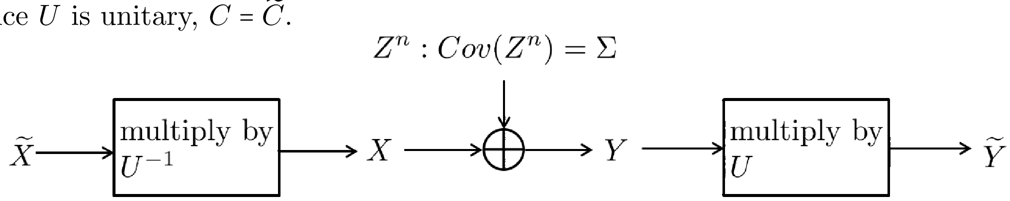
Therefore we have the equivalent channel as follows:

$$\widetilde{Y}^n = \widetilde{X}^n + \widetilde{Z}^n, \ \widetilde{Z}^n_j \sim \mathcal{N}(0, \sigma_j^2) \text{ indep across } j$$

By Theorem 17.8, we have that

$$\widetilde{C} = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \log^+ \frac{T}{\sigma_j^2} = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{2}\log^+ \frac{T}{f_Z(\omega)} d\omega. \text{ (by Szegö, Theorem 5.6)}$$
$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n |T - \sigma_j^2|^+ = P(T)$$

Finally since U is unitary, $C = \widetilde{C}$.



stationary additive Gausian noise channel □

Note: Noise is born white, the colored noise is essentially due to some filtering.

17.6* Additive White Gaussian Noise channel with Intersymbol Interference

Definition 17.9 (AWGN with ISI)**.** An AWGN channel with ISI is defined as follows: $\mathcal{A} = \mathcal{B} = \mathbb{R}$, $\mathsf{c}(x) = x^2$, and the channel law $P_{Y^n|X^n}$ is given by

$$Y_k = \sum_{j=1}^{n} h_{k-j} X_j + Z_k \,, \qquad k = 1, \dots, n$$

where $Z_k \sim \mathcal{N}(0,1)$ is white Gaussian noise, $\{h_k, k = -\infty, \dots, \infty\}$ are coefficients of a discrete-time channel filter.

Theorem 17.10. *Suppose that the sequence $\{h_k\}$ is an inverse Fourier transform of a frequency response $H(\omega)$:*

$$h_k = \frac{1}{2\pi} \int_0^{2\pi} e^{i\omega k} H(\omega) d\omega \,.$$

Assume also that $H(\omega)$ is a continuous function on $[0, 2\pi]$. Then the capacity of the AWGN channel with ISI is given by

$$C(T) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} \log^+ (T |H(\omega)|^2) d\omega$$
$$P(T) = \frac{1}{2\pi} \int_0^{2\pi} \left| T - \frac{1}{|H(\omega)|^2} \right|^+ d\omega$$

Proof. (Sketch) At the decoder apply the inverse filter with frequency response $\omega \mapsto \frac{1}{H(\omega)}$. The equivalent channel then becomes a stationary colored-noise Gaussian channel:

$$\tilde{Y}_j = X_j + \tilde{Z}_j \,,$$

where $\tilde{Z}_j$ is a stationary Gaussian process with spectral density

$$f_{\tilde{Z}}(\omega) = \frac{1}{|H(\omega)|^2} \,.$$

Then apply Theorem 17.9 to the resulting channel.

Remark: to make the above argument rigorous one must simply carefully analyze the non-zero error introduced by truncating the deconvolution filter to finite n. □

17.7* Gaussian channels with amplitude constraints

We have examined some classical results of additive Gaussian noise channels. In the following, we will list some more recent results without proof.

Theorem 17.11 (Amplitude-constrained capacity of AWGN channel)**.** *For an AWGN channel $Y_i = X_i + Z_i$ with amplitude constraint $|X_i| \le A$ and energy constraint $\sum_{i=1}^{n} X_i^2 \le nP$, we denote the capacity by:*

$$C(A, P) = \max_{P_X : |X| \le A, \mathbb{E}|X|^2 \le P} I(X; X + Z).$$

Capacity achieving input distribution P_X^* is discrete, with finitely many atoms on $[-A, A]$. Moreover, the convergence speed of $\lim_{A \to \infty} C(A, P) = \frac{1}{2} \log(1 + P)$ is of the order $e^{-O(A^2)}$.

For details, see [Smi71] and [PW14, Section III].

17.8* Gaussian channels with fading

Fading channels are often used to model the urban signal propagation with multipath or shadowing. The received signal Y_i is modeled to be affected by multiplicative fading coefficient H_i and additive noise Z_i:

$$Y_i = H_i X_i + Z_i, \quad Z_i \sim \mathcal{N}(0,1)$$

In the coherent case (also known as CSIR – for channel state information at the receiver), the receiver has access to the channel state information of H_i, i.e. the channel output is effectively (Y_i, H_i). Whenever H_j is a stationary ergodic process, we have the channel capacity given by:

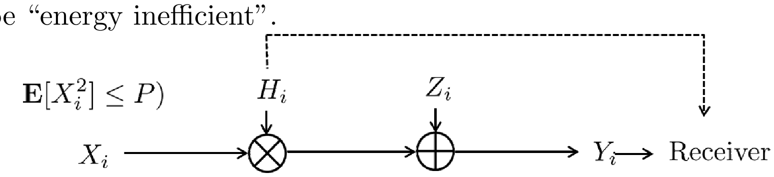
$$C(P) = \mathbb{E}[\frac{1}{2}\log(1 + P|H|^2)]$$

and the capacity achieving input distribution is the usual $P_X = \mathcal{N}(0, P)$. Note that the capacity $C(P)$ is in the order of $\log(P)$ and we call the channel "energy efficient".

In the non-coherent case where the receiver does not have the information of H_i, no simple expression for the channel capacity is known. It is known, however, that the capacity achieving input distribution is discrete, and the capacity

$$C(P) = O(\log\log P), \qquad P \to \infty \tag{17.8}$$

This channel is said to be "energy inefficient".



Fading channel

With introduction of multiple antenna channels, there are endless variations, theoretical open problems and practically unresolved issues in the topic of fading channels. We recommend consulting textbook [TV05] for details.