EE 4TM4: Digital Communications II
Scalar Gaussian Channel

I. Differential Entropy

Let X be a continuous random variable with probability density function (pdf) $f(x)$ (in short $X \sim f(x)$). The differential entropy of X is defined as

$$h(X) = -\int f(x) \log f(x) dx.$$

For example, if $X \sim \text{Unif}[a, b]$, then

$$h(X) = \log(b - a).$$

As another example, if $X \sim N(\mu, \sigma^2)$, then

$$h(X) = \frac{1}{2} \log(2\pi e \sigma^2).$$

The differential entropy $h(X)$ is a concave function of $f(x)$. However, unlike entropy it is not always nonnegative, and hence should not be interpreted directly as a measure of information. The differential entropy is invariant under translation but not under scaling.

- Translation: For any constant a, $h(X + a) = h(X)$.
- Scaling: For any nonzero constant a, $h(aX) = h(X) + \log |a|$.

The maximum differential entropy of a continuous random variable $X \sim f(x)$ under the average power constraint $E(X^2) \leq P$ is

$$\max_{f(x):E(X^2)\leq P} h(X) = \frac{1}{2} \log(2\pi e P),$$

and is attained when X is Gaussian with zero mean and power P, i.e., $X \sim N(0, P)$. Thus, for any $X \sim f(x)$,

$$h(X) = h(X - E(X)) \leq \frac{1}{2} \log(2\pi e \text{Var}(X)).$$

Proof: Let g be any density with variance P and ϕ be a zero-mean Gaussian density with variance P. We have

$$\begin{aligned}0 &\le D(g\|\phi)\\ &= \int g\log(g/\phi)\\ &= -h(g) - \int g\log\phi\\ &= -h(g) - \int \phi\log\phi\\ &= -h(g) + h(\phi),\end{aligned}$$

where the substitution $\int g\log\phi = \int \phi\log\phi$ follows from the fact that g and ϕ yield the same variance (note that $\log\phi = ax^2 + b$). ∎

Let $X \sim F(x)$ be an arbitrary random variable and $Y|\{X = x\} \sim f(y|x)$ be continuous for every x. The conditional differential entropy $h(Y|X)$ of Y given X is defined as

$$h(Y|X) = -E_{X,Y}(\log f(Y|X)).$$

Note that $I(X;Y) = h(Y) - h(Y|X)$. Since $I(X;Y) \ge 0$, it follows that

$$h(Y|X) \le h(Y).$$

II. Additive White Gaussian Noise Channel

Channel model: $Y(t) = X(t) + N(t)$, where $\{N(t)\}_t$ is an i.i.d. Gaussian process with $N(t) \sim \mathcal{N}(0, \sigma_N^2)$ for all t.

Theorem 1: The capacity of additive white Gaussian noise channel with average power constraint P is given by

$$C_{AWGN} = \frac{1}{2}\log\frac{P+\sigma_N^2}{\sigma_N^2} = \frac{1}{2}\log(1+\mathrm{SNR}).$$

The achievability part follows by discretizing the channel input and output and by a weak convergence argument.

Proof of the converse. First note that the proof of the converse for the DMC with input cost constraint applies to arbitrary (not necessarily discrete) memoryless channels. Therefore, we have

$$C_{AWGN} \le \sup_{F(x):E[X^2]\le P} I(X;Y).$$

Now for any $X \sim F(x)$ with $E[X^2] \le P$,

$$\begin{aligned}I(X;Y) &= h(Y) - h(Y|X)\\ &= h(Y) - h(N|X)\\ &= h(Y) - h(N)\\ &\le \frac{1}{2}\log(2\pi e(P+\sigma_N^2)) - \frac{1}{2}\log(2\pi e\sigma_N^2)\\ &= \frac{1}{2}\log\frac{P+\sigma_N^2}{\sigma_N^2},\end{aligned}$$

where the inequality becomes equality if $X \sim N(0, P)$.

III. Scalar Fading Channels: Ergodic Capacity

Channel model: $Y(t) = H(t)X(t) + N(t)$, where $\{H(t)\}_t$ is a stationary and ergodic process with stationary distribution $p(h)$, $\{N(t)\}_t$ is an i.i.d. Gaussian process with $N(t) \sim \mathcal{N}(0, \sigma_N^2)$ for all t.

1) $\{H(t)\}_t$ is known at the receiver: (coherent channel capacity)

$$\begin{aligned} C &= \max_{E[X^2]\le P} I(X;Y|H) \\ &= \max_{E[X^2]\le P} E_H I(X;Y|H) \\ &= \frac{1}{2} E_H \log\left(1 + \frac{H^2 P}{\sigma_N^2}\right). \end{aligned}$$

Define the average SNR at the receiver as

$$\mathrm{SNR} = \frac{E_H[H^2]P}{\sigma_N^2}.$$

By Jensen's inequality,

$$\frac{1}{2} E_H \log\left(1 + \frac{H^2 P}{\sigma_N^2}\right) \le \frac{1}{2}\log\left(1 + \frac{E_H[H^2]P}{\sigma_N^2}\right) = \frac{1}{2}\log\left(1 + \mathrm{SNR}\right).$$

Bad news: the capacity of fading channel is smaller than that of the AWGN channel with the same receiver SNR. In the low SNR regime, we have

$$\frac{1}{2} E_H \log\left(1 + \frac{H^2 P}{\sigma_N^2}\right) \approx \frac{1}{2} E_H\left(\frac{H^2 P}{\sigma_N^2}\right)\log_2 e = \frac{1}{2}\mathrm{SNR}\log_2 e \approx C_{\text{AWGN}}.$$

At high SNR,

$$\begin{aligned} \frac{1}{2} E_H \log\left(1 + \frac{H^2 P}{\sigma_N^2}\right) &\approx \frac{1}{2} E_H \log\left(\frac{H^2 P}{\sigma_N^2}\right) \\ &= \frac{1}{2}\log \mathrm{SNR} + \frac{1}{2} E_H \log\left(\frac{H^2}{E_H[H^2]}\right) \\ &\approx C_{\text{AWGN}} + \frac{1}{2} E_H \log\left(\frac{H^2}{E_H[H^2]}\right). \end{aligned}$$

2) $\{H(t)\}_t$ is known at the transmitter and the receiver: (dynamic power control and opportunistic communication)

$$\begin{aligned} C &= \max_{E[X^2(H)]\le P} I(X;Y|H) \\ &= \max_{E[X^2(H)]\le P} E_H \max_{X(h)} I(X;Y|H=h) \end{aligned}$$

Define $E[X^2(h)] = P(h)$. We have

$$\max_{E[X^2(h)]\le P} E_H \max_{X(h)} I(X;Y|H=h) = \max_{E_H P(H)\le P} \frac{1}{2} E_H \log\left(1 + \frac{H^2 P(H)}{\sigma_N^2}\right).$$

The maximizer $P^*(h)$ is

$$P^*(h) = \max\left(\lambda - \frac{\sigma_N^2}{h^2}, 0\right)$$

with λ given by

$$E_H[P^*(H)] = P.$$

Therefore,

$$C = \frac{1}{2}\int_{|h|=\sigma_N\lambda^{-1/2}}^{\infty} p(h)\log\left(\frac{h^2\lambda}{\sigma_N^2}\right)\mathrm{d}h$$

In the high SNR regime, λ is very large and $P^*(h) \approx \lambda$. Therefore the capacity converges to that of the case with no state information at transmitter.

Suppose the channel gain $\mathbf{h}^2$ has a peak value G_{max}. At low SNR, we have

$$\mathbb{P}(h^2 \approx G_{\text{max}})\left(\lambda - \frac{\sigma_N^2}{G_{\text{max}}}\right) = P$$

and thus

$$\begin{aligned}
\frac{1}{2}E_H \log\left(1 + \frac{H^2P^*(H)}{\sigma_N^2}\right) &\approx \frac{1}{2}\mathbb{P}(h^2 \approx G_{\text{max}})\log\left(1 + \frac{G_{\text{max}}P}{\sigma_N^2\mathbb{P}(h^2 \approx G_{\text{max}})}\right)\\
&\approx \frac{G_{\text{max}}P}{2\sigma_N^2}\log_2 e\\
&= \frac{1}{2}\mathrm{SNR}\frac{G_{\text{max}}}{E_H[H^2]}\log_2 e\\
&\approx \frac{G_{\text{max}}}{E_H[H^2]}C_{\text{AWGN}}.
\end{aligned}$$

Good news: the capacity gain over the AWGN channel depends on the peakness of the fading process, which can be unbounded.

Since the capacity scales linearly with the power at low SNR ($\frac{1}{2}\log(1+\mathrm{SNR}) \approx \frac{1}{2}\mathrm{SNR}\log_2 e$) and logrithmically at high SNR ($\frac{1}{2}\log(1+\mathrm{SNR}) \approx \frac{1}{2}\log(\mathrm{SNR})$), power control is more effective in the low SNR regime. We will see that at high SNR, a more efficient method is to increase the degree of freedom, say, using multiple antennas.

When proving the capacity formula for the case where the state process is know at both the transmitter and receiver, we adopted the multiplexing technique, i.e., using different coding schemes for different channel states. But such a scheme may cause long decoding delay. Furthermore, for fading processes with continuous amplitude, there are essentially infinite states, which make the multiplexing method infeasible. Fortunately, for fading channels with additive Gaussian noise, the multiplexing method can be replaced by dynamic power control. That is because we if assume the fading process is $\{\sqrt{P^*(H(t))}H(t)\}_t$ and the power constraint is $E[X^2] \le 1$, then the capacity with the state information at the receiver is exactly

$$\frac{1}{2}E_H\log\left(1 + \frac{P^*(H)H^2}{\sigma_N^2}\right).$$

Since $P^*(H(t))$ is a deterministic function of $H(t)$, the receiver can compute $\{\sqrt{P^*(H(t))}H(t)\}_t$ from the fading process $\{H(t)\}_t$. Out assumption is thus valid. So in order to achieve the capacity, we only need use the power control policy $P^*(h)$ and a coding scheme with power constraint $E[X^2] \le 1$ designed for the effective fading process $\{\sqrt{P^*(H(t))}H(t)\}_t$. Note: the resulting transmission power is $E_H[P^*(H)] = P$.

REFERENCES

[1] T. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Edition. Hoboken, NJ: Wiley, 2006.

[2] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2011.

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge, U.K.: Cambridge University Press, 2005.