信息论第十一讲作业解答

中国科学技术大学《信息论 A》006125.01 班助教组

2025 年 5 月 23 日

第 1 题

Consider a channel with additive exponential noise, $Y = X + Z$, where the noise Z obeys an exponential distribution with mean λ, independent of X, and the input X has support $[0, \infty)$ and a mean constraint $\mathbf{E}X \le \mu$. Calculate the information capacity-cost function $C_I(\mu)$ of this channel.

解: 对每个 x 有

$$h(Y|X=x) = h(Y-x|X=x) = h(Y-X|X=x) = h(Z|X=x) = h(Z) = \log_2(e\lambda),$$

所以 $h(Y|X) = \log_2(e\lambda)$. 因为 Y 取值于 $[0,\infty)$, $\mathbf{E}[Y] = \mathbf{E}[X] + \mathbf{E}[Z] \le \mu + \lambda$, 所以 $h(Y) \le \log_2(e(\mu+\lambda))$ 这样

$$I(X;Y) = h(Y) - h(Y|X) \le \log_2\left(1 + \frac{\mu}{\lambda}\right).$$

设随机变量 X 以概率 $\lambda/(\mu+\lambda)$ 取 0, 以概率 $\mu/(\mu+\lambda)$ 服从均值为 $\mu+\lambda$ 的指数分布. 这样 0 有邻域 A 使对每个 $t \in A$ 有

$$\mathbf{E}[e^{tX}] = \frac{\lambda}{\mu+\lambda} + \frac{\mu}{\mu+\lambda}\frac{1}{1-(\mu+\lambda)t}, \tag{1}$$

$$\mathbf{E}[e^{tZ}] = \frac{1}{1-\lambda t}. \tag{2}$$

用矩母函数的性质可得对每个 $t \in A$ 有 $\mathbf{E}[e^{tY}] = \mathbf{E}[e^{tX}]\mathbf{E}[e^{tZ}]$ 即

$$\mathbf{E}[e^{tY}] = \frac{1}{1-(\mu+\lambda)t}. \tag{3}$$

所以 Y 服从均值为 $\mu+\lambda$ 的指数分布, $h(Y) = \log_2(e(\mu+\lambda))$, $I(X;Y) = \log_2(1+\mu/\lambda)$. 同时 $P[X \ge 0] = 1$, $\mathbf{E}[X] = (\mu/(\mu+\lambda))(\mu+\lambda) = \mu$.

作为 $P[X \ge 0] = 1$ 和 $\mathbf{E}[X] \le \mu$ 的条件下 $I(X;Y)$ 的最大值, $C(\mu) = \log_2(1+\mu/\lambda)$. □

寻找 X 的分布的过程和上面的推导过程是相反的. 我们希望 Y 满足 $h(Y)=\log_2(e(\mu+\lambda))$, 就先假设 Y 服从均值为 $\mu+\lambda$ 的指数分布. 这样 Y 有 (3) 式所示的矩母函数. 用 (2) 式所示的 Z 的矩母函数除 Y 的矩母函数可以得到 X 具有 (1) 式所示的矩母函数. 所以 X 应该以概率 $\lambda/(\mu+\lambda)$ 取 0, 以概率 $\mu/(\mu+\lambda)$ 服从均值为 $\mu+\lambda$ 的指数分布.

第 2 题

Show that for a Gaussian signal observed via an additive noise channel, when the noise is also Gaussian, the estimation quality is the worst. Let the signal be $X\sim\mathcal{N}(0,P)$, and the noise Z be independent of X with mean zero and variance N. Prove the following inequality:

$$\mathbf{E}\left[(X-\mathbf{E}\left[X|X+Z\right])^2\right]\le\frac{PN}{P+N},$$

where the equality holds when $Z\sim\mathcal{N}(0,N)$.

证明: 设 $Y=X+Z$，我们有如下引理 (证明见注 1)

引理 1. $\min_g\mathbf{E}\left[(X-g(Y))^2\right]=\mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])^2\right]$, *其中 $g(Y)$ 表示所有根据 Y 对 X 的估计器.*

此时若假设 $g(Y)=aY$，即为线性估计器，根据引理 1 显然有 $\mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])^2\right]\le\mathbf{E}\left[(X-aY)^2\right]$. 接下来考察 $\min_a\mathbf{E}\left[(X-aY)^2\right]$, 可得

$$\begin{aligned}\frac{d}{da}(\mathbf{E}\left[(X-aY)^2\right])&=\frac{d}{da}\left(\mathbf{E}\left[X^2\right]-2a\mathbf{E}\left[XY\right]+a^2\mathbf{E}\left[Y^2\right]\right)\\&=-2\mathbf{E}\left[XY\right]+2a\mathbf{E}\left[Y^2\right],\end{aligned}$$

令上式为零可得

$$a=\frac{\mathbf{E}\left[XY\right]}{\mathbf{E}\left[Y^2\right]}=\frac{\mathbf{E}\left[X(X+Z)\right]}{\mathbf{E}\left[(X+Z)^2\right]}=\frac{\mathbf{E}\left[X^2\right]+\mathbf{E}\left[XZ\right]}{\mathbf{E}\left[X^2\right]+2\mathbf{E}\left[XZ\right]+\mathbf{E}\left[Z^2\right]}=\frac{P}{P+N}.$$

因此

$$\begin{aligned}\mathbf{E}\left[(X-\mathbf{E}\left[X|X+Z\right])^2\right]\le\mathbf{E}\left[(X-aY)^2\right]&=\mathbf{E}\left[X^2\right]-2a\mathbf{E}\left[XY\right]+a^2\mathbf{E}\left[Y^2\right]\\&=P-2\frac{P^2}{P+N}+\left(\frac{P}{P+N}\right)^2\cdot(P+N)\\&=\frac{PN}{P+N}.\end{aligned}\tag{4}$$

若 $Z\sim\mathcal{N}(0,N)$, 则 $Y\sim\mathcal{N}(0,N+P)$, 利用贝叶斯公式可得

$$f_{X|Y}(x|y)=\frac{f_X(x)f_{Y|X}(y|x)}{f_Y(y)}$$

$$
\begin{aligned}
&= \frac{\frac{1}{\sqrt{2\pi P}} e^{-\frac{x^2}{2P}} \cdot \frac{1}{\sqrt{2\pi N}} e^{-\frac{(y-x)^2}{2N}}}{\frac{1}{\sqrt{2\pi(N+P)}} e^{-\frac{y^2}{2(N+P)}}} \\
&\propto e^{-\frac{1}{2}\left(\frac{x^2}{P}+\frac{(y-x)^2}{N}-\frac{y^2}{N+P}\right)} \\
&\propto e^{-\frac{1}{2}\left(\frac{(x-\frac{P}{P+N}y)^2}{\frac{PN}{P+N}}\right)}
\end{aligned}
$$

根据高斯分布的统计性质，此时可得 $\mathbf{E}[X|Y] = \frac{P}{P+N}Y, \mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])^2\right] = \frac{PN}{P+N}$. 所以此时 (4) 等式成立, 即完成证明. □

注 1 (引理 1 的证明)**.** 关于引理 *1*, 需证明 $\mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])^2\right] \le \mathbf{E}\left[(X-g(Y))^2\right]$ 对于任意估计器 $g(Y)$ 都成立.

下面先证明 $\mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])^2|Y\right] \le \mathbf{E}\left[(X-g(Y))^2|Y\right]$ 对任意 g 成立, 再对两边取关于变量 Y 的期望即可.

$$
\begin{aligned}
\mathbf{E}\left[(X-g(Y))^2|Y\right] &= \mathbf{E}\left[\left(X-\mathbf{E}\left[X|Y\right]+\mathbf{E}\left[X|Y\right]-g(Y)\right)^2|Y\right] \\
&= \mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])^2|Y\right] + \mathbf{E}\left[(\mathbf{E}\left[X|Y\right]-g(Y))^2|Y\right] \qquad (5) \\
&\ge \mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])^2|Y\right] \qquad (6)
\end{aligned}
$$

(5) 基于

$$
\begin{aligned}
\mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])(\mathbf{E}\left[X|Y\right]-g(Y))|Y\right] &= (\mathbf{E}\left[X|Y\right]-g(Y))\,\mathbf{E}\left[(X-\mathbf{E}\left[X|Y\right])|Y\right] \\
&= (\mathbf{E}\left[X|Y\right]-g(Y))(\mathbf{E}\left[X|Y\right]-\mathbf{E}\left[X|Y\right]) \\
&= 0
\end{aligned}
$$

所以对 (6) 两边取关于变量 Y 的期望，随即完成证明.

第 3 题

For a continuous random variable S with mean zero and variance σ^2, consider its information rate-distortion function under the squared error distortion measure, $d(s,\hat{s}) = (s-\hat{s})^2$.
a) Show that $R_I(D) = 0$ when $D \ge \sigma^2$.
b) Show that $R_I(D) \ge h(S) - \frac{1}{2}\log(2\pi eD)$ when $D < \sigma^2$.
c) Show that $R_I(D) \le \frac{1}{2}\log\frac{\sigma^2}{D}$ when $D < \sigma^2$.

证明: a): 对于 $D \ge \sigma^2$, 我们可以将 S 只编码为 $\hat{S} = \mathbf{E}[S] = 0$, 便有 $\mathbf{E}[X-\mathbf{E}[X]]^2 = \sigma^2 \le D$. 即此时我们可以达到比 D 更小的失真且码率为 0.

b): 类似于高斯分布下率失真的推导, 我们保留 $h(S)$ 并使用最大熵原理便得到：

$$\begin{aligned} I(S;\hat{S}) &= h(S) - h(S|\hat{S}) \\ &\geq h(S) - h(S-\hat{S}) \\ &\geq h(S) - \frac{1}{2}\log(2\pi e D) \end{aligned}$$

c): 为证明上界, 我们试图构造一个确切的信道, 能够达到这个界, 则我们最优的率失真函数一定不差于他, 令：

$$\hat{S} = \frac{\sigma^2 - D}{\sigma^2}S + Z\,, \quad Z \sim \mathcal{N}(0, \frac{D(\sigma^2-D)}{\sigma^2})\perp S$$

此时：

$$\begin{aligned} \mathbf{E}[d(S,\hat{S})] = \mathbf{E}[(S-\hat{S})^2] = \mathbf{E}[(\frac{D}{\sigma^2}S - Z)^2] = \frac{D^2}{\sigma^2} + \frac{D(\sigma^2-D)}{\sigma^2} = D \\ h(\hat{S}) - h(\hat{S}|S) = h(\frac{\sigma^2-D}{\sigma^2}S + Z) - h(Z) \\ = \frac{1}{2}\log(2\pi e(\sigma^2-D)) - \frac{1}{2}\log(2\pi e\frac{D(\sigma^2-D)}{\sigma^2}) = \frac{1}{2}\log\frac{\sigma^2}{D} \end{aligned}$$

由于我们所求的是 $I(S;\hat{S})$ 极小值, 我们便有 $R(D) \leq \frac{1}{2}\log\frac{\sigma^2}{D}$.

注 2. *c) 中的构造受第 4 题 b) 的启发.*

□

第 4 题

We have seen that a useful trick for calculating rate-distortion functions is to construct suitable test channels from $\hat{S}$ to S. But in the optimization problem for solving rate-distortion functions, we need to characterize the forward channel from S to $\hat{S}$.

a) What is the forward channel $P_{\hat{S}|S}$ for a Bernoulli source under Hamming distortion?
b) What is the forward channel $f_{\hat{S}|S}$ for a Gaussian source under squared error distortion?
c) Calculate the information rate-distortion function for a Laplace source under absoluate error distortion; i.e., $f_S(s) = \frac{1}{2b}e^{-|s|/b}$, and $d(s,\hat{s}) = |s-\hat{s}|$.

a) 解: 设信源符号服从 Bernoulli(δ), $\delta \leq 1/2$. 根据第 3 讲讲义, 如果 $D \geq \delta$ 则 $\hat{S}$ 以概率 1 取 0.

再设 $0 \leq D < \delta$. 此时 $\hat{S}$ 服从 Bernoulli($(\delta-D)/(1-2D)$), 存在独立于 $\hat{S}$ 的 Bernoulli(D) 随机变量 Z 使 $S = \hat{S} \oplus Z$ 以概率 1 成立. 对所有 $s, \hat{s} \in \{0,1\}$ 有

$$P_{\hat{S}|S}(\hat{s}|s) = \frac{P_{\hat{S}}(\hat{s})P_{S|\hat{S}}(s|\hat{s})}{P_S(s)}.$$

用这个公式可以求出

$$P_{\hat{S}|S}(0|0) = \frac{(1-\delta-D)(1-D)}{(1-\delta)(1-2D)}, P_{\hat{S}|S}(1|0) = \frac{D(\delta-D)}{(1-\delta)(1-2D)},$$
$$P_{\hat{S}|S}(0|1) = \frac{D(1-\delta-D)}{\delta(1-2D)}, P_{\hat{S}|S}(1|1) = \frac{(\delta-D)(1-D)}{\delta(1-2D)}. \qquad \square$$

b) 解: 设信源符号服从 $\mathcal{N}(0,\sigma^2)$. 根据第 10 讲讲义, 如果 $D \geq \sigma^2$ 则 $\hat{S}$ 以概率 1 取 0.

再设 $0 < D < \sigma^2$. 此时 $\hat{S}$ 服从 $\mathcal{N}(0,\sigma^2 - D)$, 存在独立于 $\hat{S}$ 的 $\mathcal{N}(0,D)$ 随机变量 Z 使 $S = \hat{S} + Z$ 以概率 1 成立. 对所有实数 s 和 $\hat{s}$ 有

$$\begin{aligned} f_{\hat{S}|S}(\hat{s}|s) &= \frac{f_{\hat{S}}(\hat{s}) f_{S|\hat{S}}(s|\hat{s})}{f_S(s)} = \frac{f_{\hat{S}}(\hat{s}) f_Z(s-\hat{s})}{f_S(s)} \\ &= \sqrt{\frac{\sigma^2}{2\pi D(\sigma^2 - D)}} \exp\left(-\frac{\sigma^2}{2D(\sigma^2-D)}\left(\hat{s} - \frac{\sigma^2 - D}{\sigma^2}s\right)^2\right). \end{aligned} \qquad \square$$

在解 c 问之前我们先推导一点 Laplace 分布的性质.

引理 2. 设 $b > 0$, 对所有实数 x 有

$$f(x) = \frac{1}{2b} e^{-|x|/b},$$

f 是随机变量 X 的概率密度函数. 这样 $\mathbf{E}[|X|] = b$, $h(X) = \log_2(2eb)$, 对所有 $t \in (-1/b, 1/b)$ 有

$$\mathbf{E}[e^{tX}] = \frac{1}{1 - b^2 t^2}.$$

如果 Y 是随机变量且 $\mathbf{E}[|Y|] = b$ 则 $h(Y) \leq \log_2(2eb)$.

证明:

$$\begin{aligned} \mathbf{E}[|X|] &= \int_{-\infty}^{\infty} |x| f(x) dx = 2\int_0^{\infty} x \frac{1}{2b} e^{-x/b} dx = b, \\ h(X) &= -\mathbf{E}[\log_2(f(X))] = -\mathbf{E}\left[-\log_2(2b) - \frac{|X|}{b}\log_2(e)\right] \\ &= \log_2(2b) + \frac{\mathbf{E}[|X|]}{b}\log_2(e) = \log_2(2eb). \end{aligned}$$

对每个 $t \in (-1/b, 1/b)$,

$$\mathbf{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{0} e^{tx} \frac{1}{2b} e^{x/b} dx + \int_0^{\infty} e^{tx} \frac{1}{2b} e^{-x/b} dx = \frac{1}{1 - b^2 t^2},$$

如果 Y 是随机变量且 $\mathbf{E}[|Y|] = b$ 则

$$h(Y) = -\mathbf{E}[\log_2(f_Y(Y))]$$

$$
\begin{aligned}
&= -\mathbf{E}\left[\log_2\left(\frac{f_Y(Y)}{f(Y)}\right)\right] - \mathbf{E}[\log_2(f(Y))] \\
&= -D(f_Y\|f) - \mathbf{E}[\log_2(f(Y))] \\
&\leq -\mathbf{E}[\log_2(f(Y))] \\
&= -\mathbf{E}\left[-\log_2(2b) - \frac{|Y|}{b}\log_2(e)\right] \\
&= \log_2(2eb).
\end{aligned}
$$

□

c) 解: 用 R 表示这个率失真函数. 对每个 $D \in [b, \infty)$, 因为 $\mathbf{E}[d(S,0)] = \mathbf{E}[|S|] = b \leq D$, $I(S;0) = 0$, 所以 $\mathbf{E}[d(S,\hat{S})] \leq D$ 的条件下 $I(S;\hat{S})$ 的最小值等于 0, 即 $R(D) = 0$.

设 $0 < D < b$, $\mathbf{E}[d(S,\hat{S})] \leq D$. 对每个 $\hat{s}$ 有 $h(S|\hat{S}=\hat{s}) = h(S-\hat{s}|\hat{S}=\hat{s}) = h(S-\hat{S}|\hat{S}=\hat{s})$, 所以

$$
h(S|\hat{S}) = h(S-\hat{S}|\hat{S}) \leq h(S-\hat{S}). \tag{7}
$$

我们已经假设了 $\mathbf{E}[|S-\hat{S}|] \leq D$. 再用引理 2 得

$$
h(S-\hat{S}) \leq \log_2(2e\mathbf{E}[|S-\hat{S}|]) \leq \log_2(2eD). \tag{8}
$$

这样 $I(S;\hat{S}) = h(S) - h(S|\hat{S}) \geq \log_2(2eb) - \log_2(2eD) = \log_2(b/D)$.

设 $0 < D < b$, 随机变量 $\hat{S}$ 以概率 D^2/b^2 取 0, 以概率 $(b^2-D^2)/b^2$ 服从参数为 b 的 Laplace 分布, 随机变量 Z 服从参数为 D 的 Laplace 分布, $\hat{S}$ 和 Z 独立. 定义 $S = \hat{S} + Z$. 这样 $\mathbf{E}[d(S,\hat{S})] = \mathbf{E}[|Z|] = D$. 对每个 $t \in (-1/b, 1/b)$ 有

$$
\mathbf{E}[e^{tS}] = \mathbf{E}[e^{t\hat{S}}]\mathbf{E}[e^{tZ}] = \left(\frac{D^2}{b^2} + \frac{b^2-D^2}{b^2}\frac{1}{1-b^2t^2}\right)\frac{1}{1-D^2t^2} = \frac{1}{1-b^2t^2},
$$

所以 S 服从参数为 b 的 Laplace 分布, 即信源符号的分布. 可以看出 (7) 和 (8) 式成立等号, 所以 $I(S;\hat{S}) = \log_2(b/D)$.

对每个 $D \in (0,b)$, 作为 $\mathbf{E}[d(S,\hat{S})] \leq D$ 的条件下 $I(S;\hat{S})$ 的最小值, $R(D) = \log_2(b/D)$.

综上所述,

$$
R(D) = \begin{cases} \log_2\left(\dfrac{b}{D}\right), & 0 < D < b \\ 0, & D \geq b \end{cases}.
$$

□

第 5 题

Consider a memoryless additive noise channel $Y = X + Z$ *where* X *has sopprt* $[-1/2, 1/2]$, *and the noise* Z *is uniform over* $[-1,1]$, *independent of* X. *Calculate the information capacity of the channel,* $C_I = \max_{f_X} I(X;Y)$.

解:

$$I(X;Y) = I(X;X+Z) = h(X+Z) - h(Z) = h(X+Z) - \log_2 2.$$

我们注意到 $Y = X + Z$ 取值于 $[-\frac{3}{2}, \frac{3}{2}]$, 并且:

$$\begin{aligned} f_Y(y) = f_{X+Z}(X+Z=y) = (f_X * f_Z)(y) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} f_X(x) f_Z(y-x)\,\mathrm{d}x \\ &= \begin{cases} \frac{1}{2}\int_{-\frac{1}{2}}^{y+1} f_X(x)\,\mathrm{d}x, & -\frac{3}{2} \le y < -\frac{1}{2} \\ \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} f_X(x)\,\mathrm{d}x = \frac{1}{2}, & -\frac{1}{2} \le y \le \frac{1}{2} \\ \frac{1}{2}\int_{y-1}^{\frac{1}{2}} f_X(x)\,\mathrm{d}x, & \frac{1}{2} < y \le \frac{3}{2} \end{cases}, \end{aligned} \tag{9}$$

即 Y 在 $[-\frac{1}{2}, \frac{1}{2}]$ 上的概率密度函数为 $\frac{1}{2}$, 而在 $[-\frac{3}{2}, -\frac{1}{2})$ 和 $(\frac{1}{2}, \frac{3}{2}]$ 上的分布依赖于 $f_X(x)$.

我们又由最大熵原理, 在固定支撑集上的最大熵分布为均匀分布, 那么 f_Y 在支撑集 $[-\frac{3}{2}, -\frac{1}{2})$ 和 $(\frac{1}{2}, \frac{3}{2}]$ 上的概率密度函数应均为 $\frac{1}{4}$, 根据 (9), 对 $\forall y \in [-\frac{3}{2}, -\frac{1}{2})$ 有:

$$\frac{1}{2}\int_{-\frac{1}{2}}^{y+1} f_X(x)\,\mathrm{d}x = \frac{1}{4},$$

此时则有

$$\lim_{y \to -\frac{3}{2}} \int_{-\frac{1}{2}}^{y+1} f_X(x)\,\mathrm{d}x = \frac{1}{2}.$$

对于 $(\frac{1}{2}, \frac{3}{2}]$ 的情况分析同理, 即在 $\frac{1}{2}$ 和 $-\frac{1}{2}$ 的邻域上累积分布函数分别达到 $\frac{1}{2}$. 那么在连续意义上, 我们可以用 $\frac{1}{2}\left(\delta_n(x-\frac{1}{2}) + \delta_n(x+\frac{1}{2})\right)$ 来描述之, $\delta_n(x)$ 为狄拉克函数. 若没有连续要求, 我们也可以令 $P_X(x=-\frac{1}{2}) = P_X(x=\frac{1}{2}) = \frac{1}{2}$ 来描述, 此时我们便很容易算出 Y 的分布, 即

$$f_Y(y) = \begin{cases} \frac{1}{4}, & y \in [-\frac{3}{2}, -\frac{1}{2}) \\ \frac{1}{2}, & y \in [-\frac{1}{2}, \frac{1}{2}] \\ \frac{1}{4}, & y \in (\frac{1}{2}, \frac{3}{2}] \end{cases},$$

可知该分布满足最大熵. 此时

$$h(X+Z) = \int_{-\frac{3}{2}}^{-\frac{1}{2}} \frac{1}{4}\log_2 4\,\mathrm{d}x + \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{2}\log_2 2\,\mathrm{d}x + \int_{\frac{1}{2}}^{\frac{3}{2}} \frac{1}{4}\log_2 4\,\mathrm{d}x = \frac{3}{2},$$

$$C = \max_{f(X)} I(X;Y) = \frac{3}{2} - 1 = \frac{1}{2}.$$

□

第 6 题

Consider a memoryless Gaussian channel $Y = X + Z$, where X has an average power constraint P, and the noise is $Z \sim \mathcal{N}(0, \sigma_Z^2)$. Suppose that, besides Y, the decoder also observes a noisy version of Z, $V = Z + W$ where $W \sim \mathcal{N}(0, \sigma_W^2)$ is independent of Z and X. What is the information capacity-cost function of this channel model?

解: 对每个 x, 依次用 Y 的定义, 第 10 讲 Proposition 10.3 和 X, (Z, V) 的独立性得 $h(Y, V|X = x) = h(x + Z, V|X = x) = h(Z, V|X = x) = h(Z, V)$. 所以

$$h(Y, V|X) = h(Z, V). \tag{10}$$

Z 和 W 独立且都是高斯的, 所以是联合高斯的. 作为 Z 和 W 的线性组合, (Z, V) 是高斯的. 所以

$$h(Z, V) = \frac{1}{2}\log_2((2\pi e)^2 \det(\mathbf{K}_{Z,V})), \tag{11}$$

其中 $\mathbf{K}_{Z,V}$ 表示 Z 和 V 的协方差矩阵. 因为 $\mathrm{Cov}[Z, V] = \mathbf{E}[Z(Z + W)] = \mathbf{E}[Z^2] + \mathbf{E}[Z]\mathbf{E}[W] = \sigma_Z^2$, $\mathrm{var[V]} = \sigma_Z^2 + \sigma_W^2$, 所以

$$\det(\mathbf{K}_{Z,V}) = \begin{vmatrix} \sigma_Z^2 & \sigma_Z^2 \\ \sigma_Z^2 & \sigma_Z^2 + \sigma_W^2 \end{vmatrix} = \sigma_Z^2\sigma_W^2. \tag{12}$$

根据最大熵定理,

$$h(Y, V) \le \frac{1}{2}\log_2((2\pi e)^2 \det(\mathbf{K}_{Y,V})), \tag{13}$$

其中 $\mathbf{K}_{Y,V}$ 表示 Y 和 V 的协方差矩阵. 因为 $\mathrm{var[Y]} = \mathrm{var[X]} + \sigma_Z^2$,

$$\begin{aligned} \mathrm{Cov}[Y, V] &= \mathbf{E}[YV] - \mathbf{E}[Y]\mathbf{E}[V] \\ &= \mathbf{E}[XZ + XW + Z^2 + ZW] - \mathbf{E}[Y]\mathbf{E}[V] \\ &= \mathbf{E}[X]\mathbf{E}[Z] + \mathbf{E}[X]\mathbf{E}[W] + \mathbf{E}[Z^2] + \mathbf{E}[Z]\mathbf{E}[W] - \mathbf{E}[Y]\mathbf{E}[V] \\ &= \sigma_Z^2, \end{aligned}$$

$\mathrm{var[V]} = \sigma_Z^2 + \sigma_W^2$, 所以

$$\det(\mathbf{K}_{Y,V}) = \begin{vmatrix} \mathrm{var[X]} + \sigma_Z^2 & \sigma_Z^2 \\ \sigma_Z^2 & \sigma_Z^2 + \sigma_W^2 \end{vmatrix} = (\sigma_Z^2 + \sigma_W^2)\mathrm{var[X]} + \sigma_Z^2\sigma_W^2 \le (\sigma_Z^2 + \sigma_W^2)P + \sigma_Z^2\sigma_W^2. \tag{14}$$

综合 $I(X; Y, V) = h(Y, V) - h(Y, V|X)$ 和 (10), (11), (12), (13), (14) 式得

$$I(X; Y, V) \le \frac{1}{2}\log_2\left(1 + \frac{\sigma_Z^2 + \sigma_W^2}{\sigma_Z^2\sigma_W^2}P\right). \tag{15}$$

设 $X \sim \mathcal{N}(0,P)$. X, Z 和 W 独立且都是高斯的, 所以是联合高斯的. 作为 X, Z 和 W 的线性组合, (Y,V) 是高斯的. 所以 (13) 式成立等号. 因为 $\mathrm{var}[\mathrm{X}] = P$, 所以 (14) 式成立等号. 这样 (15) 式成立等号.

作为 $\mathbf{E}[X^2] \le P$ 的条件下 $I(X;Y,V)$ 的最大值, 信道容量

$$C(P) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_Z^2 + \sigma_W^2}{\sigma_Z^2\sigma_W^2}P\right). \qquad \square$$

第 7 题

Consider independent Gaussian random variables $X \sim \mathcal{N}(0,\sigma_X^2)$ and $Z \sim \mathcal{N}(0,\sigma_Z^2)$. If there is another random variable V which is only uncorrelated with X, satisfying $\mathbf{E}[XV]=0$ and $\mathbf{E}[V^2] \le \sigma_Z^2$, prove that $I(X;X+V) \ge I(X;X+Z)$, and discuss the condition under which equality holds.

证明: 由于 $I(X;X+V) = h(X) - h(X|X+V)$, 欲证明 $I(X;X+V) \ge I(X;X+Z)$, 只需证:

$$h(X|X+V) \le h(X|X+Z).$$

我们有下面的不等式串:

$$\begin{aligned}
h(X|X+V) &= h(X - \frac{\sigma_X^2}{\sigma_X^2 + \mathbf{E}[V^2]}(X+V)|X+V) && (16)\\
&= h\left(\frac{\sigma_X^2 V - \mathbf{E}[V^2]X}{\sigma_X^2 + \mathbf{E}[V^2]}|X+V\right)\\
&\le h\left(\frac{\sigma_X^2 V - \mathbf{E}[V^2]X}{\sigma_X^2 + \mathbf{E}[V^2]}\right) && (17)\\
&\le \frac{1}{2}\log\left(2\pi e\mathbf{Var}\left[\frac{\sigma_X^2 V - \mathbf{E}[V^2]X}{\sigma_X^2 + \mathbf{E}[V^2]}\right]\right) && (18)\\
&\le \frac{1}{2}\log\left(2\pi e\mathbf{E}[\frac{\sigma_X^2 V - \mathbf{E}[V^2]X}{\sigma_X^2 + \mathbf{E}[V^2]}]^2\right) && (19)\\
&= \frac{1}{2}\log\left(2\pi e\frac{\sigma_X^2\mathbf{E}[V^2]}{\sigma_X^2 + \mathbf{E}[V^2]}\right) && (20)\\
&\le \frac{1}{2}\log\left(2\pi e\frac{\sigma_X^2\sigma_Z^2}{\sigma_X^2 + \sigma_Z^2}\right) && (21)\\
&= h(X|X+Z) && (22)
\end{aligned}$$

(16) 基于平移条件不影响条件熵, 系数 k 的选取是当 V 是与 X 独立的高斯变量时（这是我们希望的取等条件）, 保证 $X - k(X+V)$ 和 $X+V$ 独立, 那么此处的必要条件是:

$$\mathbf{E}[((1-k)X - kV)(X+V)] = (1-k)\sigma_X^2 - k\mathbf{E}[V^2] = 0 \Rightarrow k = \frac{\sigma_X^2}{\sigma_X^2 + \mathbf{E}[V^2]}$$

(17) 基于条件减少熵.
(18) 基于方差约束下的最大熵原理.
(19) 基于 $\mathbf{E}[X^2] = \mathbf{E}[X]^2 + \mathbf{Var}[X] \geq \mathbf{Var}[X]$.
(20) 基于展开后使用 $\mathbf{E}[XV] = 0$ 并化简.
(21) 基于函数 $\frac{\sigma_X^2 \mathbf{E}[V^2]}{\sigma_X^2 + \mathbf{E}[V^2]}$ 关于 $\mathbf{E}[V^2]$ 单调递增.
(22) 基于以下事实:

$$\begin{aligned}
h(X|X+Z) &= h(X+Z|X) + h(X) - h(X+Z) \\
&= \frac{1}{2}\log(2\pi e\sigma_Z^2) + \frac{1}{2}\log(2\pi e\sigma_X^2) - \frac{1}{2}\log(2\pi e(\sigma_X^2 + \sigma_Z^2)) \\
&= \frac{1}{2}\log\left(2\pi e\frac{\sigma_X^2\sigma_Z^2}{\sigma_X^2 + \sigma_Z^2}\right)
\end{aligned}$$

设 $I(X;X+V) = I(X;X+Z)$. 这样 $h(X|X+V) = h(X|X+Z)$, 所以 (17), (18), (19) 和 (21) 成立等号. 因为 (17), (18) 和 (19) 成立等号, 所以

$$\frac{\sigma_X^2 V - \mathbf{E}[V^2]X}{\sigma_X^2 + \mathbf{E}[V^2]}$$

独立于 $X+V$, 服从正态分布, 且有均值 0. 因为 (21) 成立等号, 所以 $\mathbf{E}[V^2] = \sigma_Z^2$.

反过来, 如果

$$\frac{\sigma_X^2 V - \mathbf{E}[V^2]X}{\sigma_X^2 + \mathbf{E}[V^2]}$$

独立于 $X+V$, 服从正态分布, 有均值 0, 且 $\mathbf{E}[V^2] = \sigma_Z^2$, 则可以验证 $I(X;X+V) = I(X;X+Z)$.

□

第 8 题

Consider the parallel Gaussian channel model in Section 11.4.

a) Derive the water-filling optimal solution.

b) Show that as $P \to \infty$, the rate loss due to using uniform power allocation $P_i = P/k$, $i = 1, \ldots, k$, instead of the water-filling optimal solution, asymptotically vanishes.

a) 证明: 用 A 表示满足 $\sum_{i=1}^k P_i \leq P$ 的所有 $(P_1, P_2, \cdots, P_k) \in [0,\infty)^k$ 组成的集合.

$$\begin{aligned}
I(\underline{X};\underline{Y}) &= h(\underline{Y}) - h(\underline{Y}|\underline{X}) \\
&= h(\underline{Y}) - h(Z_1, Z_2, \cdots, Z_k) \\
&= h(\underline{Y}) - \sum_{i=1}^k \frac{1}{2}\log_2(2\pi e\sigma_i^2)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{k} h(Y_i) - \sum_{i=1}^{k} \frac{1}{2}\log_2(2\pi e \sigma_i^2) \\
&\leq \sum_{i=1}^{k} \frac{1}{2}\log_2(2\pi e(\mathbf{E}[X_i^2]+\sigma_i^2)) - \sum_{i=1}^{k} \frac{1}{2}\log_2(2\pi e \sigma_i^2) \\
&= \sum_{i=1}^{k} \frac{1}{2}\log_2\left(1+\frac{\mathbf{E}[X_i^2]}{\sigma_i^2}\right) \\
&\leq \max_{(P_1,P_2,\cdots,P_k)\in A} \sum_{i=1}^{k} \frac{1}{2}\log_2\left(1+\frac{P_i}{\sigma_i^2}\right),
\end{aligned} \tag{23}
$$

其中第二行的等号用了第 10 讲 Proposition 10.3, 第五行的小于等于号是因为对所有正整数 $i \leq k$ 有 $\mathrm{var[Y_i]} = \mathrm{var[X_i]} + \mathrm{var[Z_i]} \leq \mathbf{E}[X_i^2] + \sigma_i^2$, 最后一行的小于等于号是因为 $\sum_{i=1}^{k} \mathbf{E}[X_i^2] \leq P$.

设 $(P_1^*, P_2^*, \cdots, P_k^*) \in A$,

$$
\sum_{i=1}^{k} \frac{1}{2}\log_2\left(1+\frac{P_i^*}{\sigma_i^2}\right) = \max_{(P_1,P_2,\cdots,P_k)\in A} \sum_{i=1}^{k} \frac{1}{2}\log_2\left(1+\frac{P_i}{\sigma_i^2}\right), \tag{24}
$$

X_1, X_2, $\cdots$, X_k 分别服从 $\mathcal{N}(0,P_1^*)$, $\mathcal{N}(0,P_2^*)$, $\cdots$, $\mathcal{N}(0,P_k^*)$ 且独立. 这样 Y_1, Y_2, $\cdots$, Y_k 独立, 对每个正整数 $i \leq k$, Y_i 服从 $\mathcal{N}(0, \mathbf{E}[X_i^2]+\sigma_i^2)$. 所以 (23) 式第四和第五行的小于等于号现在是等号. (23) 式最后一行的小于等于号现在当然也是等号.

作为 $I(\underline{X};\underline{Y})$ 的最大值, $C(P)$ 等于 (23) 式的最后一行. 设 $(P_1^*, P_2^*, \cdots, P_k^*) \in A$ 且 (24) 式成立. 我们用 KKT 条件 [1, 5.49 式] 来求 $(P_1^*, P_2^*, \cdots, P_k^*)$. 可以看出

$$
-\sum_{i=1}^{k} \ln\left(1+\frac{P_i^*}{\sigma_i^2}\right) = \min_{(P_1,P_2,\cdots,P_k)\in A}\left(-\sum_{i=1}^{k} \ln\left(1+\frac{P_i}{\sigma_i^2}\right)\right).
$$

如果 μ, ν_1, ν_2, $\cdots$, $\nu_k \geq 0$ 则对所有 P_1, P_2, $\cdots$, $P_k \geq 0$ 和正整数 $i \leq k$ 有

$$
\frac{\partial}{\partial P_i}\left(-\sum_{j=1}^{k} \ln\left(1+\frac{P_j}{\sigma_j^2}\right) + \mu\left(\sum_{j=1}^{k} P_j - P\right) + \sum_{j=1}^{k} \nu_j(-P_j)\right) = -\frac{1}{\sigma_i^2+P_i} + \mu - \nu_i.
$$

所以存在非负数 μ, ν_1, ν_2, $\cdots$, ν_k 满足

$$
\begin{aligned}
\sum_{i=1}^{k} P_i^* - P \leq 0, \\
-P_i^* \leq 0, \forall i \in \{1,2,\cdots,k\}, \\
\mu\left(\sum_{i=1}^{k} P_i^* - P\right) = 0,
\end{aligned} \tag{25}
$$

$$\nu_i(-P_i^*) = 0, \forall i \in \{1, 2, \cdots, k\}, \tag{26}$$

$$-\frac{1}{\sigma_i^2 + P_i^*} + \mu - \nu_i = 0, \forall i \in \{1, 2, \cdots, k\}, \tag{27}$$

即 KKT 条件. 由 (27) 式得 $\mu = 1/(\sigma_1^2 + P_1^*) + \nu_1 > 0$, 对每个正整数 $i \le k$ 有

$$P_i^* = \frac{1}{\mu - \nu_i} - \sigma_i^2.$$

如果正整数 $i \le k$, $\nu_i = 0$, 则 $P_i^* = 1/\mu - \sigma_i^2$, $1/\mu - \sigma_i^2 \ge 0$. 如果正整数 $i \le k$, $\nu_i > 0$, 则由 (26) 式得 $P_i^* = 0$, 所以 $1/(\mu - \nu_i) = \sigma_i^2 > 0$, $1/\mu - \sigma_i^2 < 1/(\mu - \nu_i) - \sigma_i^2 = 0$. 总之对每个正整数 $i \le k$ 有

$$P_i^* = \max\left(\frac{1}{\mu} - \sigma_i^2, 0\right). \tag{28}$$

因为 $\mu > 0$ 且 (25) 式成立, 所以 $\sum_{i=1}^k P_i^* = P$,

$$\sum_{i=1}^{k} \max\left(\frac{1}{\mu} - \sigma_i^2, 0\right) = P. \tag{29}$$

因此只要 $(P_1^*, P_2^*, \cdots, P_k^*) \in A$ 且 (24) 式成立就存在 $\mu > 0$ 满足 (29) 式且使 (28) 式对所有正整数 $i \le k$ 成立. 把 μ 换成 -2λ 我们就得到了讲义的 39 和 40 式. 我们在前面构造的 $(X_1, X_2, \cdots, X_k)$ 服从均值为 0_k, 协方差矩阵为

$$\begin{bmatrix} P_1^* & & & \\ & P_2^* & & \\ & & \ddots & \\ & & & P_k^* \end{bmatrix}$$

的高斯分布. □

b) 证明: 根据 (a) 问的结论, 对每个 $P \ge 0$ 存在 $L(P) \ge 0$ (即 (a) 问中的 $1/\mu$) 满足

$$\sum_{i=1}^{k} \max(L(P) - \sigma_i^2, 0) = P,$$

$$C(P) = \sum_{i=1}^{k} \frac{1}{2} \log_2\left(1 + \frac{\max(L(P) - \sigma_i^2, 0)}{\sigma_i^2}\right),$$

其中 C 表示容量-代价函数. 设正整数 $m \le k$, $\sigma_m = \max(\sigma_1, \sigma_2, \cdots, \sigma_k)$, $N = \sum_{i=1}^k \sigma_i^2$. 如果 $P \ge 0$ 且 $L(P) \le \sigma_m^2$ 则

$$P \le \sum_{i=1}^{k} \max(\sigma_m^2 - \sigma_i^2, 0) = k\sigma_m^2 - N,$$

所以对每个 $P > k\sigma_m^2 - N$ 有 $L(P) > \sigma_m^2$, 也就有

$$P = \sum_{i=1}^{k}(L(P) - \sigma_i^2) = kL(P) - N,$$

$$C(P) = \sum_{i=1}^{k} \frac{1}{2}\log_2\left(1 + \frac{L(P) - \sigma_i^2}{\sigma_i^2}\right) = \sum_{i=1}^{k} \frac{1}{2}\log_2\left(\frac{P+N}{k\sigma_i^2}\right).$$

对每个 $P > k\sigma_m^2 - N$, 均匀分配功率达到的码率与容量的差为

$$C(P) - \sum_{i=1}^{k} \frac{1}{2}\log_2\left(1 + \frac{P}{k\sigma_i^2}\right) = \sum_{i=1}^{k} \frac{1}{2}\log_2\left(\frac{P+N}{P+k\sigma_i^2}\right).$$

$P \to \infty$ 时这个差趋于 0. □

第 9 题

Consider a channel with two inputs (X_1, X_2) and two outputs (Y_1, Y_2) obeying the following channel law:

$$Y_1 = X_1 + Z_1,$$
$$Y_2 = h(X_1) + X_2 + Z_2,$$

where $h(\cdot)$ is a given function, and (Z_1, Z_2) are independent noises. A decoding scheme is as follows: first decode X_1 from Y_1, and then decode X_2 from $Y_2 - h(X_1)$. Use a reasoning based on mutual information analysis to argue that this decoding scheme is suboptimal in general.

证明: 根据题中解码方案，即需证 $I(X_1, X_2; Y_1, Y_2) \geq I(X_1; Y_1) + I(X_2; Y_2 - h(X_1))$. 利用互信息的链式法则和微分熵的性质可得

$$\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= I(X_1; Y_1, Y_2) + I(X_2; Y_1, Y_2|X_1) \\
&= I(X_1; Y_1, Y_2) + h(Y_1, Y_2|X_1) - h(Y_1, Y_2|X_1, X_2) \\
&= I(X_1; Y_1, Y_2) + h(Y_1|X_1) + h(Y_2|X_1, Y_1) - h(Y_1, Y_2|X_1, X_2) \\
&= I(X_1; Y_1, Y_2) + h(Z_1) + h(X_2 + Z_2) - (h(Y_1|X_1, X_2) + h(Y_2|X_1, X_2, Y_1)) \quad (30) \\
&= I(X_1; Y_1, Y_2) + h(Z_1) + h(X_2 + Z_2) - h(Z_1) - h(Z_2) \quad (31) \\
&= I(X_1; Y_1, Y_2) + h(X_2 + Z_2) - h(X_2 + Z_2|X_2) \\
&= I(X_1; Y_1) + I(X_1; Y_2|Y_1) + I(X_2; X_2 + Z_2) \\
&\geq I(X_1; Y_1) + I(X_2; Y_2 - h(X_1)). \quad (32)
\end{aligned}$$

(30) 基于

$$h(Y_1|X_1) = h(X_1 + Z_1|X_1) = h(Z_1|X_1) = h(Z_1),$$
$$h(Y_2|X_1, Y_1) = h(h(X_1) + X_2 + Z_2|X_1, Z_1) = h(X_2 + Z_2|X_1, Z_1) = h(X_2 + Z_2).$$

(31) 基于

$$h(Y_1|X_1, X_2) = h(X_1 + Z_1|X_1, X_2) = h(Z_1|X_1, X_2) = h(Z_1),$$
$$h(Y_2|X_1, X_2, Y_1) = h(h(X_1) + X_2 + Z_2|X_1, X_2, Z_1) = h(Z_2|X_1, X_2, Z_1) = h(Z_2).$$

由此可得题中解码方案是次优的, 当且仅当存在 $X_1 \leftrightarrow Y_1 \leftrightarrow Y_2$ 时, 该解码方案才是最优的. □

第 10 题

Consider a joint source channel coding setup where the source is a memoryless Gaussian source with mean zero and variance Q, and the channel is a memoryless Gaussian channel whose Gaussian noise has mean zero and variance σ^2. Let the conversion ratio between source and channel be $r = 1$. Consider an average squared error distortion D for source reproduction and an average input power constraint P for channel transmission.

a) *Identify the fundamental performance limit between D and P.*
b) *Show that it is possible to design simple symbol-level mappings to achieve the fundamental performance limit.*
c) *Verify that the "double matching" conditions in Section 6.5 hold for the designed symbol-level mappings.*

a) 解: 高斯信源的率失真函数为：

$$R(D) = \frac{1}{2}\log\frac{Q}{D}$$

高斯信道的信道容量为：

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

根据理论性能界 $R(D) \leq C(P)$，得

$$D \geq \frac{Q\sigma^2}{\sigma^2 + P}$$

□

b) 证明: 考虑采用线性映射：$X = f(S) = \alpha S$, $\hat{S} = g(Y) = \beta Y$. 依题有 $Y = X + Z$, 信道噪声 $Z \sim \mathcal{N}(0, \sigma^2)$, 此时分析该 JSCC 方案的平均失真:

$$\begin{aligned} D &= \mathbf{E}_{(S,\hat{S})}[(S-\hat{S})^2] \\ &= \mathbf{E}_{(S,Z)}[(S-\beta(\alpha S+Z))^2] \\ &= \mathbf{E}_S[(1-\alpha\beta)^2 S^2] + \mathbf{E}_Z[\beta^2 Z^2] \\ &= (1-\alpha\beta)^2 Q + \beta^2\sigma^2 \\ &\geq \frac{Q\sigma^2}{\sigma^2 + Q\alpha^2} \qquad (33) \\ &\geq \frac{Q\sigma^2}{\sigma^2 + P} \qquad (34) \end{aligned}$$

(33) 基于 $h(\beta) = (1-\alpha\beta)^2 Q + \beta^2\sigma^2 = (\alpha^2 Q + \sigma^2)\beta^2 - 2\alpha Q\beta + Q \geq \frac{Q\sigma^2}{\sigma^2+Q\alpha^2}$, 当且仅当 $\beta = \frac{\alpha Q}{\alpha^2 Q + \sigma^2}$ 时取等.
(34) 基于信道输入 X 要满足功率约束 $\mathbf{E}[X^2] = \alpha^2 Q \leq P$, 当 X 功率恰好为 P, 即 $\alpha = \sqrt{\frac{P}{Q}}$ 时取等.

综上, 当 $(\alpha, \beta) = (\sqrt{\frac{P}{Q}}, \frac{\sqrt{PQ}}{\sigma^2+P})$ 时, $D = \frac{Q\sigma^2}{\sigma^2+P}$, 此时该方案达到理论性能界. □

c) 证明:

对于 b) 中我们设计的 $X = f(S) = \sqrt{\frac{P}{Q}}S$, $\hat{S} = g(Y) = \frac{\sqrt{PQ}}{\sigma^2+P}Y, D = \frac{Q\sigma^2}{\sigma^2+P}$,

- 可得 $Y = \sqrt{\frac{P}{Q}}S + Z$, 则 $Y \sim \mathcal{N}(0, P+\sigma^2)$, 有
$$I(S;g(Y)) = I(S;Y) = h(Y) - h(Y|S) = h(\sqrt{\tfrac{P}{Q}}S + Z) - h(Z) = \frac{1}{2}\log(1+\frac{P}{\sigma^2}) = R(D);$$
- 同理有
$$I(f(S);Y) = I(S;Y) = \frac{1}{2}\log(1+\frac{P}{\sigma^2}) = C(P);$$
- 以及
$$I(S;g(Y)) = I(S;Y) = I(f(S);Y).$$

因此，“双重匹配”条件成立。 □

参考文献

[1] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004. [Online]. Available: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf