# 信道容量

2020年7月28日 11:21

## 1. 概述

首先信道不像信源是不可控的，所以我们一般是根据信道来设计信源，目的是让信源编码通过信道后可以被最大程度还原。通常情况，我们知道的Y关于X的越多，我们的译码结果就越准，那么对其统计平均那就是互信息的概念。

$$I(X;Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(y|x)}{p(y)}$$

$$=H(Y)-H(X|Y)$$

根据互信息的凸性：当 $p(y|x)$ 固定，关于 $p(x)$ 上凸。
当 $p(x)$ 固定，关于 $p(y|x)$ 下凸。

这里我们对应的第一种情形，所以我们可以求得 $I(X;Y)$ 的最大值，也就是关于这个信道的信道容量 $C$

对于一个信道，我们通常考虑的都是一连串的编码，更还有连续的信道。对于离散的信道，为了归一化，我们讨论的信道容量为

$$C=\lim_{N\to\infty}\frac{1}{N}\max_{p(x)}I(x_1,x_2,\cdots x_N;y_1,\cdots,y_N)$$

这和熵率的概念很接近。我们要求信道容量，这实际上是个极大化的问题。
最普通的办法，我们可以采用KKT条件来分析，在一些极端情况，我们可以采用更简单的分析办法。
所以，我们先分析离散无记忆信道。

## 2. 离散无记忆信道的信道容量分析

首先我们对 $I(x_1,x_2,\cdots,x_N;y_1,\cdots,y_N)$ 进行简化，有以下定理

$$I(x_1,x_2,\cdots,x_N;y_1y_2\cdots y_N)\le\sum_{n=1}^{N}I(x_n;y_n)$$

证：

$$I(x_1,x_2,\cdots,x_N;y_1,\cdots,y_N)=I(\vec{X},\vec{Y})=H(\vec{Y})-H(\vec{Y}|\vec{X})$$

分别讨论 $H(\vec{Y})$ 和 $H(\vec{Y}|\vec{X})$

$$H(\vec{Y})=H(y_1)+H(y_2|y_1)+H(y_3|y_1,y_2)+\cdots+H(y_N|y_1\cdots y_{N-1})$$

由条件减少熵

$$H(\vec{Y})\le H(y_1)+H(y_2)+\cdots+H(y_N)=\sum_{i=1}^{N}H(y_i)$$，等号当且仅当 $y_i$ 互相独立时取到。

$$H(\vec{Y}|\vec{X})=-\sum_{x\in X}\sum_{y\in Y}p(x_1,\cdots,x_n,y_1,\cdots,y_n)\log p(y_1,y_2,\cdots y_n|x_1x_2,\cdots,x_n)$$

由于信道是离散无记忆信道，则

$$p(y_1y_2\cdots,y_n|x_1x_2,\cdots,x_n)=\prod_{n=1}^{N}q(y_n|x_n)$$

$$\Rightarrow H(\vec{Y}|\vec{X})=-\sum_{x\in X}\sum_{y\in Y}p(x_1,\cdots,x_n;y_1,\cdots,y_n)\log\prod_{n=1}^{N}q(y_n|x_n)$$

$$=-\sum_{n=1}^{N}\sum_{x\in X}\sum_{y\in Y}p(\vec{x},\vec{y})\log q(y_n|x_n)$$

由于 $\log q(y_n|x_n)$ 只与n有关，则

$$H(\vec{Y}|\vec{X})=-\sum_{n=1}^{N}\sum_{x_n\in X}\sum_{y_n\in Y}p(x_n,y_n)\log q(y_n|x_n)=\sum_{n=1}^{N}H(y_n|x_n)$$

则 $I(\vec{X};\vec{Y})\le\sum_{n=1}^{N}I(x_n;y_n)$，等号仅当 $y_i$ 相互独立时取到。

$$p(\vec{y})=\sum_{\vec{x}}p(\vec{x})q(\vec{y}|\vec{x})=\sum_{\vec{x}}p(\vec{x})\prod_{n=1}^{N}q(y_n|x_n)=\prod_{n=1}^{N}p(y_n)=\prod_{n=1}^{N}\sum_{x_n\in X}q(y_n|x_n)\cdot p(x_n)$$

$\Rightarrow$

$$p(\vec{x})=\prod_{n=1}^{N}p(x_n)$$

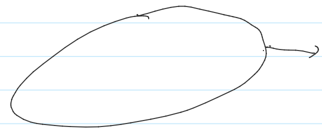
也就是说，对于离散无记忆信道，$y_n$ 相互独立等价于输入 $x_n$ 相互独立。等号在输入相互独立时取到。
当信源平稳时，$I(X_n;Y_n)=I(X;Y)$
记输入字母概率分布 $P=(p(a_1),p(a_2),\cdots,p(a_k))$，前向概率转移矩阵 $Q$，
$I(\vec{X},\vec{Y})=I(P,Q)$，

$$C=\max_{P}I(P\cdot Q)$$

### 2.1 KKT条件

## 2.1 KKT条件

KKT条件是优化问题中在机器学习中广泛使用的一种情形



这里

$$\sum_{k=1}^{N} P(a_k)=1,\quad 0\le P(a_k)\le 1$$

由图像法分析，若最值 $I(p)$ 在分界上取到，$\nabla\cdot I(p)=u\nabla\cdot g(p)$（这里$g$是约束）
若最值在内部，那么 $\nabla\cdot I(p)=0$ 就行，这等于$u$取0值相等价。
则有

$$\begin{cases} g(p)\le 0\\ u\ge 0\\ u\cdot g(p)=0\end{cases}$$

我们代入拉格朗日乘子法 $L=$

$$\sum_{i=1}^{N}\sum_{j=1}^{M} P_X(x_i)P_{Y|X}(y_j|x_i)\log\frac{P_{Y|X}(y_j|x_i)}{P_Y(y_j)}+\lambda\left(\sum_{i=1}^{N}P(x_i)-1\right)+\sum_{i=1}^{N}u_i p(x_i)$$

$$\sum_{i=1}^{N}P(x_i)\cdot\sum_{j=1}^{M}P_{Y|X}(y_j|x_i)\log\frac{P_{Y|X}(y_j|x_i)}{P_Y(y_j)}+\lambda\left(\sum_{i=1}^{N}P(x_i)-1\right)+\sum_{i=1}^{N}u_i p(x_i)$$

对$P(x_i)$求导.

$$\frac{\partial L}{\partial P(x_i)}=P(x_i)\left[\sum_{j=1}^{M}P_{Y|X}(y_j|x_i)\log\frac{P_{Y|X}(y_j|x_i)}{P_Y(y_j)}\right]'+\sum_{j=1}^{M}P_{Y|X}(y_j|x_i)\log\frac{P_{Y|X}(y_j|x_i)}{P_Y(y_j)}+\lambda+u$$

$$=P(x_i)\left[\sum_{j=1}^{M}P_{Y|X}(y_j|x_i)\log\frac{P_{Y|X}(y_j|x_i)}{P(y_j|x_i)P(x_i)+\sum_{k\ne i,k=0}^{N}P(x_k)P(y_j|x_k)}\right]'+I(Y;x_i)+\lambda+u$$

$$=P(x_i)\left\{\sum_{j=1}^{M}P_{Y|X}(y_j|x_i)\cdot\log\left[P_{Y|X}(y_j|x_i)-\log\left(P(y_j|x_i)P(x_i)+\sum_{k\ne i,k=0}^{N}P(x_k)P(y_j|x_k)\right)\right]\right\}'$$

$$+I(Y;x_i)+\lambda+u$$

$$=P(x_i)\sum_{j=1}^{M}P_{Y|X}(y_j|x_i)\cdot\frac{-P(y_j|x_i)}{P_Y(y_j)\ln 2}+I(Y;x_i)+\lambda+u_i$$

$$=\sum_{j=1}^{M}-P_{Y|X}(y_j|x_i)\cdot P(x_i|y_j)\Big/\frac{1}{\log_2 e}+I(Y;x_i)+\lambda+u_i$$

$$=I(Y;x_i)-\log_2 e+\lambda+u$$

我们让

$$\frac{\partial L}{\partial p(x_i)}=0\implies I(Y;x_i)=\log_2 e-(\lambda+u_i)=C,$$

虽然我们无法求出$\lambda$和$u$，但我们对$I(X;Y)$分析可以发现：

$I(X;Y)=\sum p(x_i)\cdot I(Y;x_i)$

对于边界 $P(x_i)=0$，$p(x_i)\cdot I(Y;x_i)=0$.
边界 $p(x_i)>0$，$p(x_i)\cdot I(Y;x_i)\ne 0$，$u_i=0$.
要求最大值，就需要让较小的$I(Y;x_i)$对应的$p(x_i)=0$。而$p(x_i)$时 $I(Y;x_i)=\log_2 e-\lambda-u_i=\log_2 e-\lambda$.
是一个常数，设为$C$.

$$I(X;Y)=\sum p(x_i)I(Y;x_i)=C\cdot\sum_{p(x_i)>0}p(x_i)=C.$$

## 2.2 特殊情况下的信道容量计算

对于$I(X;Y)$的最大值优化问题在当输入字母表的个数非常多时求解过程较为复杂。但如果信道具有某些性质时可能会简化问题。设输入概率为$p$，转移概率为$q$

$$I(x_i;Y)=\sum_{j=1}^{M}q(y_j|x_i)\log\frac{q(y_j|x_i)}{p(y_j)}=C.$$

$$I(x_k;Y)=\sum_{j=1}^{M}q(y_j|x_k)\log\frac{q(y_j|x_k)}{p(y_j)}=C.$$

$$I(x_k;Y)=\sum_{j=1}^{M} q(y_j|x_k)\log\frac{q(y_j|x_k)}{p(y_j)} = C.$$

可以发现若 $p(y_j)=\frac{1}{M}$，且 $q(y_j|x_k)$ 满足矩阵上的置换关系，那么两者必相同。这个矩阵我们也称为信道转移矩阵

$$Q=\begin{bmatrix} q(y_1|x_1) & q(y_2|x_1) & \cdots & q(y_m|x_1) \\ q(y_1|x_2) & q(y_2|x_2) & \cdots & q(y_m|x_2) \\ \vdots & \vdots & \ddots & \vdots \\ q(y_1|x_N) & q(y_2|x_N) & \cdots & q(y_m|x_N) \end{bmatrix}$$

$$P(Y)=P(X)\cdot Q(Y|X)$$

若行互为置换，则 $H(Y|x_k)$ 相同。
若列互为置换，则当 X 等概分布时，$P(Y)$ 也是等概分布

现在我们需要验证一下，当 $P(Y)$ 等概时，且转移矩阵为行置换时，$I(x_i;Y)$ 是否取得了最大值。

$$I(x_i;Y) = \sum_{j=1}^{M} q(y_j|x_i)\log q(y_j|x_i) - \sum_{j=1}^{M} q(y_j|x_i)\log p(y_j)$$

$$= -H(Y|x_i) - \sum_{j=1}^{M} q(y_j|x_i)\log p(y_j)$$

第一项 $H(Y|x_i)$ 是一个定值。我们希望让第二项尽可能大。即

$$\min \sum_{j=1}^{M} q(y_j|x_i)\log p(y_j)$$

$p(y_j)$ 等概 $\Rightarrow$ $p(x_i)$ 相同

$$\sum_{j=1}^{M} q(y_j|x_i)\log\frac{1}{M} = -\log|M|$$

$$I(x_i;Y)=\log|M| - H(Y|x_i) = H(Y)-H(Y|x_i)$$ 不确定为最大值。

由此我们解决了某些特殊情况下的信道容量计算问题，并找到了使其达到最大值的概率分布，总结如下。

(1) 强对称信道：
行互为置换，列互为置换，信道容量在输入等概时取到。

(2) 弱对称信道：
行互为置换，信道容量在输出等概分布时取到。

(3) Gallager-symmetrisch (TUD 中有讲)：
Die Zeilen der Kanalmatrix sind Permutationen voneinander und das Ausgangsalphabet lässt sich so in disjunkte Teilmenge zerlegen, dass für jede Teilmenge der resultierende Teilkanal stark-symmetrisch ist.

即行互为置换，可以分解为几个对称的信道。

例如：

$$Q_1=\begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix} \qquad C=\log 2 - H(\varepsilon) = 1-H(\varepsilon)$$

$$Q_2=\begin{bmatrix} 1-\varepsilon & \varepsilon & 0 \\ 0 & \varepsilon & 1-\varepsilon \end{bmatrix} \qquad C=\log 2 - H(\varepsilon)$$

wenn $P(Y)=\frac{1}{3}$, $\varepsilon=\frac{1}{3}$

$$Q_3=\begin{bmatrix}1-\epsilon & \epsilon & 0\\ \frac{1}{2}(1-\epsilon) & \epsilon & \frac{1}{2}(1-\epsilon)\\ 0 & \epsilon & (1-\epsilon)\end{bmatrix}\quad 若\ \epsilon=\frac{1}{3}\ \Rightarrow \begin{bmatrix}\frac{2}{3} & \frac{1}{3} & 0\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ 0 & \frac{1}{3} & \frac{2}{3}\end{bmatrix}\qquad \begin{array}{l} P(X_1)=0\\ \log 2-H(\frac{1}{3})=1-\epsilon\ \text{bit.}\end{array}$$

通常情况，第一行和第三行为行置换，所以若 $P(X_1)=0$，则可以达到最大值，若进一计算

$$I(Y;X_0)=1-\epsilon\log\frac{1-\epsilon}{P(y_1)}+\epsilon\log\frac{\epsilon}{P(y_2)}$$

$$I(Y;X_1)=\frac{1-\epsilon}{2}\log\frac{1-\epsilon/2}{P(y_1)}+\epsilon\log\frac{\epsilon}{P(y_2)}+\frac{1-\epsilon}{2}\log\frac{\frac{1}{2}(1-\epsilon)}{P(y_3)}$$

$$I(Y;X_2)=1-\epsilon\log\frac{1-\epsilon}{P(y_3)}+\epsilon\log\frac{\epsilon}{P(y_2)}$$

若

$$I(Y;X_0)=I(Y;X_1)=I(Y;X_2)$$

$$\Rightarrow P(y_1)=P(y_3)\ 且\ I(Y;X_1)=(1-\epsilon)\log\frac{(1-\epsilon)/2}{P(y_1)}+\epsilon\log\frac{\epsilon}{P(y_2)}\neq I(Y;X_0)\neq I(Y;X_2)$$

$$\Rightarrow P(X_1)=0,$$

$$Q_4=\begin{bmatrix}1-\epsilon & \epsilon & 0\\ 1-\epsilon & 0 & \epsilon\\ \epsilon & 0 & 1-\epsilon\\ 0 & \epsilon & 1-\epsilon\end{bmatrix}$$

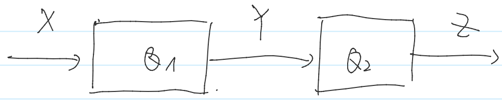
行互为置换，$P(y)=\frac{1}{3}$.
1,3列互为置换，可以推得 $P(x_1)+P(x_2)=P(x_3)+P(x_4)$
$P(x_3)=P(x_2)\Rightarrow P(x_1)=P(x_4)$

$$\Rightarrow 2\epsilon P(x_1)=\frac{1}{3}\qquad P(x_1)=P(x_4)=\frac{1}{6\epsilon}$$

$$\frac{1}{3}=(1-\epsilon)\cdot\frac{1}{6\epsilon}+P(x_2)\cdot(1-\epsilon+\epsilon)\Rightarrow P(x_2)=\frac{2\epsilon-1+\epsilon}{6\epsilon}=\frac{1}{2}-\frac{1}{6\epsilon}=P(x_3)$$

## 2.3 级联信道和并联信道的信道容量

### 2.3.1 级联信道




实质上分析，若Y被给定，X与Z无关，这与马尔可夫链有些类似，那么有

$$P(XZ|Y)=P(X|Y)\cdot P(Z|Y)$$

因此 $I(XZ|Y)=H(X|Y)+H(Z|Y)-H(XZ|Y)=0$
现在我们想讨论一下级联信道的信道容量，即 $I(X;Z)$，根据互信息的可加性

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

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

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

由互信息的非负性

$$I(X;Z)\le I(X;Y)$$

同理

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

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

$$\rightarrow I(X;Z)\le I(Y;Z)$$

这说明信息越处理获得的信息量越少，级联信道的信息量不可能大于其各组成信道的信道容量。

其计算也并不困难，$P(Z)=P(X)\prod_{i=1}^{N}Q_i$，但也不算简单，这涉及到SVD分解，我们会在讨论连续信道的信道容量中再次提到。

下面举一个例子：二元对称信道

其计算也并不困难. $P(Z)=P(X)\prod_{i=1}^{n} Q_i$，但也不算简单. 这涉及到SVD分解. 我们会在讨论连续信道的信道容量中再次提到.

下面举一个例子：二元对称信道.

$$Q_0=\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}.$$

求一个特征值和特征向量.

先求特征值：

$$\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}x=\begin{bmatrix}\lambda & \\ & \lambda\end{bmatrix}\cdot x.$$

$$\begin{bmatrix}1-\varepsilon-\lambda & \varepsilon\\ \varepsilon & 1-\varepsilon-\lambda\end{bmatrix}x=0.$$

$$\det\begin{vmatrix}1-\varepsilon-\lambda & \varepsilon\\ \varepsilon & 1-\varepsilon-\lambda\end{vmatrix}=0 \Rightarrow (1-\varepsilon-\lambda)(1-\varepsilon-\lambda)-\varepsilon^2=0$$

$$(1-\varepsilon-\lambda)(1-\varepsilon-\lambda)=1-\varepsilon-\lambda-\varepsilon+\varepsilon^2+\varepsilon\lambda-\lambda+\lambda\varepsilon+\lambda^2=\varepsilon^2$$

$$\lambda^2+(2\varepsilon-2)\lambda+1-2\varepsilon=0$$

$$\Delta=4(\varepsilon^2-2\varepsilon+1)+8\varepsilon-4=4\varepsilon^2$$

$$\lambda=\frac{-(2\varepsilon-2)\pm 2\varepsilon}{2}=\frac{2-2\varepsilon\pm 2\varepsilon}{2}=1-\varepsilon\pm\varepsilon.$$

再求特征向量

$\lambda=1$ $$\begin{bmatrix}-\varepsilon & \varepsilon\\ \varepsilon & -\varepsilon\end{bmatrix}x=0 \Rightarrow x=\begin{bmatrix}1\\ 1\end{bmatrix}.$$ $Q\cdot x=\lambda I\cdot x$

$\lambda=1-2\varepsilon$ $$\begin{bmatrix}\varepsilon & \varepsilon\\ \varepsilon & \varepsilon\end{bmatrix}x=0 \Rightarrow x=\begin{bmatrix}1\\ -1\end{bmatrix}$$

所以，矩阵可以由正交基变换为

$$\begin{bmatrix}1 & 0\\ 0 & 1-2\varepsilon\end{bmatrix}=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}^{-1}\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

其中 $\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}^{-1}=-\frac{1}{2}\begin{bmatrix}d & -b\\ -c & a\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$

那么，$Q^N$

$$Q=\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1-2\varepsilon\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

$$Q^N=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1^N & \\ & (1-2\varepsilon)^N\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1+(1-2\varepsilon)^N & 1-(1-2\varepsilon)^N\\ 1-(1-2\varepsilon)^N & 1+(1-2\varepsilon)^N\end{bmatrix}.$$

这仍然是一个二元对称信道.

$$\begin{bmatrix}1 & -1\end{bmatrix}\begin{bmatrix} & (1-2\varepsilon)^N\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1-(1-2\varepsilon)^N & 1+(1-2\varepsilon)^N\end{bmatrix}.$$

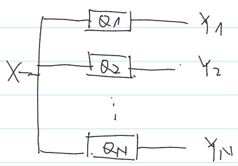
这仍然是一个二元对称信道。

$$\lim_{N\to\infty} Q^N = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix} \longrightarrow \lim_{N\to\infty} C_N = 1-H(\tfrac{1}{2}) = 0$$

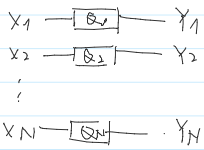
### 2.3.2 并联信道

并联信道和MIMO系统真的有些类似了。这里我们讨论三种。

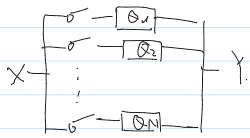
(1) 接收分集




(2) 复用




(3) 选择使用




下面我们逐一讨论

(1). 互信息为 $I(X;Y_1\cdots Y_N) = I(X;Y_1) + I(X;Y_2|Y_1)\cdots + I(X;Y_N|Y_1Y_2\cdots Y_{N-1})$

$$= I(X;Y_2) + I(X;Y_1|Y_2)\cdots + I(X;Y_N|Y_1\cdots Y_{N-1})$$

$$= \cdots$$

$$= I(X;Y_N) + I(X;Y_1|Y_N)\cdots + I(X;Y_{N-1}|Y_1\cdots Y_{N-2}Y_N).$$

由此可知信道容量要比不采用接收分集要大，上界也显而易见，即为$H(X)$。这也与直觉相符。

(2). 信道的转移矩阵，$Q$由$q(y_1\cdots y_N|x_1\cdots x_N)$组成。这与之前的时间复用在模型上是一样的，只不过是使用了空间复用。

$$C = \max_{p(x)} I(\vec{X};\vec{Y}) = \max \sum_{n=1}^{N} I(X_n;Y_n) = \sum_{n=1}^{N} C_n.$$ 当且仅当$\vec{X}$相互独立。

(3) 和信道的转移 $Q = \begin{bmatrix}Q_1 & & & \\ & Q_2 & & \\ & & \ddots & \\ & & & Q_N\end{bmatrix}$

设选择第n个信道的概率为$p_n(c)$，那么从这个矩阵形状上看，每个信道的输入字母$a_{kn}$是不同的，其互信息为

$$I_n(a_{kn};Y) = \sum_{j_n=1}^{J_n} q_n(b_{jn}|a_{kn}) \log \frac{q_n(b_{jn}|a_{kn})}{\sum_{i=1}^{K_n} p_n(a_i)\, q(b_{jn}|a_i)\cdot p_n(c)}.$$

$p_n(c)$是与$X_n$相关的。

那么根据KKT条件

$$I_n(a_{kn};Y) = C = C_n + \log\frac{1}{p_n(c)} \Rightarrow \frac{1}{p_n(c)} = 2^{C-C_n} \Rightarrow p_n(c) = \frac{1}{2^{C-C_n}}$$

为了满足KKT条件，对于$p(x)\neq 0$，有

$$I_i(a_{ki};Y) = I_j(a_{kj};Y)$$

$$\sum p_n(c) = 1. \qquad \sum_n \frac{1}{2^{C-C_n}} = 1 \Rightarrow \sum 2^{C_n} = 2^{C} \Rightarrow C = \log_2 \sum_{n=1}^{N} 2^{C_n}.$$

## 2.4 信道达到充分利用时输入输出字母概率的唯一性

《应用信息论基础》中给出了充分的证明，但这里是以KKT条件为基础推导。KKT条件的解已经充分说明了我们在求解时要求了$H(Y)$尽可能的大，所以输出的概率分布是唯一的，输入概率根据信道的转移矩阵可以不唯一。

## 3. 连续信道

对于连续信道，很自然的，我们要从连续变量的互信息入手。

$$I(X;Y)=\int_x\int_y p(x)q(y|x)\log\frac{q(y|x)}{p(y)}\,dx\,dy$$

但是现实中不存在真正的连续，例如，如果我们要发送X，那么我们永远也发不完，我们想发送的精确一点则会费很多时间，而且费电，那么就有一个费用的问题。通常电路系统中是功率。

若用随机矢量 $\vec{X}=(x_1,\cdots,x_n)$ 表示输入，其概率密度为 $p(x_1x_2,\cdots,x_n)$，该信道的费用为 $b(x)$。

那么对于长度为n的随机矢量，平均费用为

$$E[b(\vec{X})]=\sum_x p(x)b(x)=\sum_{n=1}^{N}E[b(\vec{x_n})]$$

同熵率和信道容量一样，将其归一化取极限，此时信道容量将是费用 $\beta$ 的函数

$$C(\beta)=\lim_{N\to\infty}\frac{1}{N}C_N(\beta)$$

其中

$$C_N(\beta)=\sup_{p(x)}\{I(\vec{X};\vec{Y});E\{b(\vec{X})\}\le N\beta\}$$

同离散无记忆信道一样（因为都是取了一个极限）

$$I(\vec{X};\vec{Y})\le\sum_{n=1}^{N}I(x_n;y_n)=NI(X;Y)=NI(\vec{p},\vec{Q})$$

下面我们着重讨论一些比较有代表性的信道。

### 3.1 无记忆加性噪声信道的信道容量费用函数

故名思义

$$Y=X+Z$$

那么信道转移概率函数只与Z有关

$$q(y|x)=p_Z(Z=y-x)=p_Z(y-x)$$

而 $I(X;Y)=H(Y)-H(Y|X)=H(Y)-H(Z)$

而对于加性噪声，$p(z)$ 与 $p(x)$ 完全无关联，那么我们只需让 $H(Y)$ 最大就可以了，这与之前的讨论相类似。

### 3.2 无记忆加性高斯噪声信道的信道容量函数

之前我们推导过，若随机变量的均值为 $u$，方差为 $\sigma^2$，高斯分布具有最大的微分熵

$$H(X)=\int p(x)\cdot\log\frac{1}{p(x)}dx=-\int\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(x-u)^2}{2\sigma^2}\cdot\log\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(x-u)^2}{2\sigma^2}dx$$

$$=-\int p(x)\log\frac{1}{\sqrt{2\pi\sigma^2}}dx-\int p(x)\log_2 e^{-\frac{(x-u)^2}{2\sigma^2}}dx$$

$$=\log\sqrt{2\pi\sigma^2}-\log_2 e\cdot\int p(x)\cdot-\frac{(x-u)^2}{2\sigma^2}dx$$

$$=\log\sqrt{2\pi\sigma^2}+\frac{1}{2}\log e=\frac{1}{2}\log_2 2\pi e\sigma^2$$

若噪声服从 $Z\sim N(0,\sigma_Z^2)$，$E(Z^2)=Var(Z)+u^2=Var(Z)$ 也即噪声的功率 $P_N$

我们来讨论一下Y的均值和方差

$$E(Y)=E(X)+E(Z)=E(X)\text{，若}E(X)=0\text{，则}E(Y)=0$$

$$Var(Y)=E(Y^2)=E((X+Z)^2)=E(X^2)+E(Z^2)+2E(XZ)=E(X^2)+E(Z^2)=P_S+P_N$$

那么 $\sigma_Y^2=P_S+P_N$

当Y也是高斯分布时，取得最大值，那么对于 $I(X;Y)$

$$C(\beta)=\sup\{H(Y)-H(Z);\beta_X\le P_S,\beta_Z\le P_N\}$$

$$=\frac{1}{2}\log 2\pi e(P_S+P_N)-\frac{1}{2}\log 2\pi eP_N=\frac{1}{2}\log\frac{P_S+P_N}{P_N}=\frac{1}{2}\log(1+SNR)$$

#### 3.2.1. 上、下确界

根据高斯分布之和仍是高斯分布这一结论，可知当Y为高斯分布时，X也为高斯分布，那么当输出为高斯分布时，信道容量可以达到最大。当噪声为高斯分布时，对信道的破坏力达到最大，那么上面求得的 $C(\beta)$ 也就是信道容量的下确界。

$$C(P_S)\ge\frac{1}{2}\log\left(1+\frac{P_S}{P_N}\right)$$

可以达到最大。当噪声为高斯分布时，对信道的破坏力达到最大。那么上面求得的C(P)也就是信道容量的下确界，

$$C(P_S) \ge \frac{1}{2}\log\left(1+\frac{P_S}{P_N}\right)$$

上确界我们就需要让H(Z)尽可能小。换句话说，当H(Y)一定时，尽可能让$P_N$最小。最小$P_N$即为高斯分布时。

$$H(Z)=\frac{1}{2}\log 2\pi e P_Z$$

由此，定义了一种熵功率

$$P(e)=\frac{1}{2\pi e}\exp[2H(Z)]$$，这里取了自然对数。猜想这应该是一种极限的结果。

那么上界为

$$C(P_S)\le \frac{1}{2}\frac{P_S+P_N}{P_e}$$

### 3.2.2. 无记忆加性高斯噪声信道的级联和并联

先讨论级联：对于一个加性噪声的信道，级联只是让噪声线性增加了，信噪比降低，应该避免。
再讨论并联：在离散的情形下我们讨论了接收分集、复用以及和信道，这里接收分集和和信道的意义都不大，所以我们只讨论并用信道。
设有N个无记忆的加性高斯噪声信道被并联使用，各信道的输入为$X_n$，输出为$Y_n$，输入的总功率为$P_S$。
那么约束条件为

$$\sum_{i=1}^{N} P_{Si}\le P_S,\quad P_{Si}\ge 0$$

对于每个子信道，在功率一定的情况下，信道容量的最大值为$\frac{1}{2}\log\left(1+\frac{P_{Si}}{P_{Ni}}\right)$，于是这就变成了KKT条件的练习题。

$$L(\vec{P_S})=\sum_{i=1}^{N}\frac{1}{2}\log\left(1+\frac{P_{Si}}{P_{Ni}}\right)-\lambda\left(P_S-\sum_{i=1}^{N}P_{Si}\right)-\mu_i P_{Si}=0$$

求导：

$$\frac{\partial L}{\partial P_{Si}}=\frac{1}{2}\frac{\frac{1}{P_{Ni}}}{1+\frac{P_{Si}}{P_{Ni}}}\cdot\log e+\lambda-\mu_i=0$$

$$=\frac{1}{2}\frac{1}{P_{Si}+P_{Ni}}\log e+\lambda-\mu_i=0$$

$$\Rightarrow P_{Si}+P_{Ni}=\frac{\log e}{2(\mu_i-\lambda)}$$

$$\sum P_{Si}\le P_S \Rightarrow \sum\frac{\log e}{2(\mu_i-\lambda)}\le P_S+P_N$$

(1) 若不在边界上，$\mu=0$，

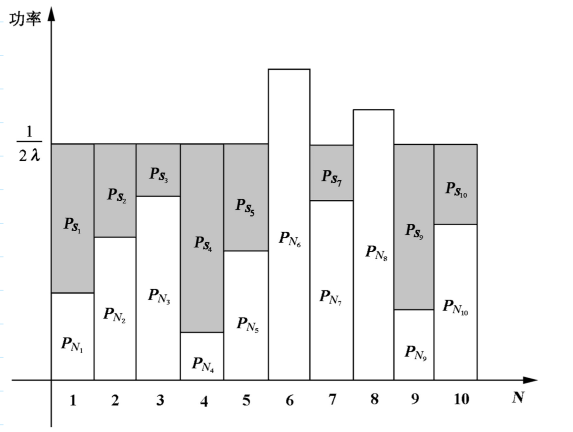
$$\sum\frac{\log e}{-2\lambda}\le P_S+P_N \qquad P_{Si}+P_{Ni}=\frac{\log e}{-2\lambda}$$

$$\Rightarrow P_{Si}+P_{Ni}=\frac{P_S+P_N}{N} \Rightarrow P_{Si}=\frac{P_S+P_N}{N}-P_{Ni}$$

(2) 若在边界上，

$$\frac{\partial L}{\partial x_i}<0 \qquad P_{Si}+P_{Ni}\ge\frac{\log e}{2(\mu_i-\lambda)} \Rightarrow 当P_{Ni}>P_S/N时，P_{Si}=0$$

综上，就有著名的注水定理，如图所示




## 4. 模拟信道的信道容量

### 4.1 K-L展开

模拟信道则就不是离散了，时间轴也是连续的。我们知道电路一般处理的数字信号，所以我们可以使用采样定理使信号变成离散的，这就可以使用上面的连续信道的分析方法来分析信道容量。

但是从模拟信道转化为连续信道有一个从连续到离散的过程，这时就有数据损失。下面我们就来讨论一下如何

模拟信道的输入是时间连续的，我们知道带宽一般都是有限的，所以我们可以利用采样定理使信号变成连续的。这样就可以使用上面的连续信道的分析方法来分析信道容量。

但是从模拟信道转化为连续信道有一个从连续到离散的过程，这时就有数据损失。下面我们就来讨论一下如何尽量减少损失或减小损失。

信息论中常用的方法是 K-L 展开（Karhunen-Loeve Expansion），基本思路是在希尔伯特空间中找到一组正交变换基，将信号展开（例如常见的傅里叶级数和抽样定理），

设信号 $x$，正交函数族为 $u_i$，则

$$\hat{x} = \sum_{i=1}^{\infty} c_i u_i$$

而我们仅需要有限个 $c_i$，那么为了保证信号损失最小，则有

$$\varepsilon = \min_{u_i} E\|x-\hat{x}\|^2 \qquad \hat{x} = \arg\min_{u_i} E\|x-\hat{x}\|^2$$

可以发现这就是使方差最小，具体计算

$$\varepsilon = E\left[(x-\hat{x})\cdot(x-\hat{x})\right] = E\left[\left(\sum_{j=d+1}^{\infty} c_j u_j\right)\left(\sum_{i=d+1}^{\infty} c_i u_i\right)\right]$$

由于正交函数系

$$u_i u_j = 0 \quad \text{when } i \neq j \qquad u_i \cdot u_j = 1 \quad \text{when } i = j$$

$$\Rightarrow \varepsilon = E\left[\sum_{j=d+1}^{\infty} c_j^2\right] = \sum_{j=d+1}^{\infty} u_j^T \cdot x \cdot x^T \cdot u_j = \sum_{j=d+1}^{\infty} u_j^T E[xx^T] u_j$$

或者写成，这里 $e(t)$ 为完备正交函数族

$$E(x_i, x_j^*) = \iint_T e_1^*(t_1) e_2(t_2)\, E(x(t_1) x^*(t_2))\, dt_1 dt_2$$

$$= \int_T\int_T R_x(t_1, t_2)\, e_i^*(t_1) e_2(t_2)\, dt_1 dt_2$$

其中 $R_x(t_1,t_2)$ 是随机信号的相关函数，那么我们可以找到特征值和特征向量

$$\int_T R_x(t_1,t_2) e_j(t_2)\, dt = \lambda_j e_j(t_1)$$

那么 $E[x_i^* x_j]$ 的解则为 $\int_T \lambda_j e_j(t_1) e_j^*(t_1)\, dt_1 = \lambda \delta_{ij}$

我们只需要选取特征值大的就行。注意这里 $x$ 是零均值的，对于非零均值，可以化为 $x - E(x)$，从而变成零均值的随机信号。

还要说明的是，当我们对随机过程应用 K-L 展开时，由于我们并没有定义随机过程的傅里叶变换，所以我们不可能使用随机过程的采样定理，而这也可以看做是对采样函数族进行正交展开。

对于平稳高斯随机过程进行 K-L 展开时，高斯随机过程的功率谱为 $\frac{N_0}{2}$，自相关函数为 $\frac{N_0}{2}\delta(\tau)$，所以，对于零均值的平稳高斯随机变量，它与任何 $L^2$ 函数（再生核希尔伯特空间的函数）的内积都是一个零均值、有限方差的高斯随机变量。而且，K-L 展开由于取的是特征值最大的函数，所以展开的系数是线性独立的，对高斯随机变量而言，则是统计独立。

所以对于所有系数的联合概率密度可以直接写成各一维概率密度的乘积。

$$p(x_1 x_2 \cdots x_N) = \prod_{i=1}^{N} p(x_i)$$

到此，我们就可以将一个模拟信道转化为一个连续信道。

### 4.2 模拟信道下的信道容量费用函数

设输入信号 $x(t)$ 和输出信号 $y(t)$，可以用 $L^2$ 函数展开，那么则有

$$I(x(t); y(t)) = \lim_{N\to\infty} I(\vec{x}; \vec{y})$$

$$= \lim_{N\to\infty} \int_x\int_y p(\vec{x})\, q(\vec{y}|\vec{x}) \log \frac{q(\vec{y}|\vec{x})}{p(\vec{y})}\, d\vec{x}\, d\vec{y}$$

于是就可以定义费用函数为

$$C_T(\beta) = \lim_{N\to\infty} \sup\{ I(\vec{x}; \vec{y}) : E[b(\vec{x})] \le \beta \}$$

因此模拟信道费用函数

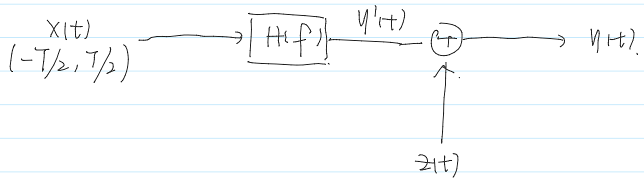
$$C(\beta) = \lim_{T\to\infty} \frac{1}{T} C_T(\beta)$$

### 4.3 一般高斯信道的容量费用函数

一般信道的高斯信道容量费用函数的模型可用如下来表示

$$x(t) \longrightarrow \boxed{H(f)} \xrightarrow{\ y'(t)\ } (+) \longrightarrow y(t)$$

一般信道的[illegible]




我们可以将噪声 $z(t)$ 看做一个单边功率度密度为 $N_0$ 的白噪声经过传输函数 $N(f)$ 滤波器后产生的。
由上面分析可知

$$y'(n) = x(n)\cdot H\left(\frac{n}{T}\right)$$

$z(n)$ 是由 $\int_T g(t-\tau)\,\theta_n(t)\,dt$ 为正交基底的互相独立的高斯随机变量，且功率谱密度为 $\frac{N_0}{2}\cdot N(f)$

那么方向则有

$$y_n = x_n\cdot H\left(\frac{n}{T}\right)+z_n \Rightarrow \frac{y_n}{H(\frac{n}{T})} = x_n+\frac{z_n}{H(\frac{n}{T})}$$

这时我们再次将其转化为了一个连续的并联信道，加上

$$E\sum_{n=-\infty}^{+\infty} x_n^2 \le P_S\cdot T$$

于是我们有了 KKT 条件

$$L(\vec{x};\vec{Y}) = \sum_{i=1}^{N} p(x_n)\,q(y_n|x_n)\log\frac{q(y_n|x_n)}{\sum_{x_k\in X} p(x_k)\cdot q(y_n|x_k)} + u\left(\sum_{n=-\infty}^{+\infty} x_n^2 - P_S T\right)$$

而单个信道的信道容量最大值为

$$C_n = \frac{1}{2}\log\left(1+\frac{P_{Sn}}{P_N}\right)$$

上式则又可写成

$$L(\vec{x};\vec{Y}) = \sum_{i=1}^{N}\frac{1}{2}\log\left(1+\frac{P_{Si}}{P_{ni}}\right) + u\left(\sum_{i=1}^{N}P_{Si} - P_S\right)$$

这本质分析并联信道时的情形一样。这里 $P_{ni} = \frac{N(\frac{n}{T})}{H^2(\frac{n}{T})}$

使用注水定理分配功率即可

## 4.4. 限带，加性白色高斯噪声信道的信道容量

这种情形就是上一种情形的特例，我们只需要带宽计算即可
也可以看做是上述过程的练习题。

这里 $N(f)=1$.
限带的 AWGN 信道，的噪声值为 $\frac{N_0}{2}$.
采样点为 $2WT$.

$$\Rightarrow P_N = \sum P_{Ni} = 2WT\cdot\frac{N_0}{2} = N_0 WT$$

$$\Rightarrow P_S T = P_S \quad , \quad P_{Si} = E(x_n) = \frac{P_S\cdot T}{2WT} = \frac{P_S T}{2W}$$

$$\Rightarrow C_T(\beta) = \sum_{i=1}^{2WT}\frac{1}{2}\log\left(1+\frac{P_S T}{2W}\Big/\frac{N_0}{2}\right) = WT\cdot\log\left(1+\frac{P_S T}{2W}\cdot\frac{2}{N_0}\right) = WT\log\left(1+\frac{P_S}{WN_0}\right)$$

$$\Rightarrow C(\beta) = \lim_{T\to\infty}\frac{1}{T}C_T(\beta) = W\cdot\log\left(1+\frac{P_S}{WN_0}\right)$$

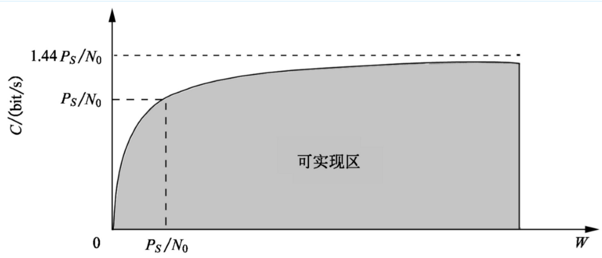
这就是著名的香农公式。
下面我们讨论下他的极限性能
(1) 增大信道容量的各种极限。
从公式中我们看出增大带宽和增加平均功率可以提升信道容量。

$$\lim_{W\to\infty} C(P_S) = \lim_{W\to\infty} W\log\left(1+\frac{P_S}{W\cdot N_0}\right)$$

$$= \lim_{W\to\infty}\log\left(1+\frac{P_S}{WN_0}\right)^{W}$$

$$= \lim_{W\to\infty}\frac{P_S}{N_0}\cdot\log\left(1+\frac{P_S}{WN_0}\right)^{\frac{WN_0}{P_S}}$$

$$= \lim_{W\to\infty}\log_2 e\cdot\frac{P_S}{N_0} \approx 1.44\,P_S/N_0$$



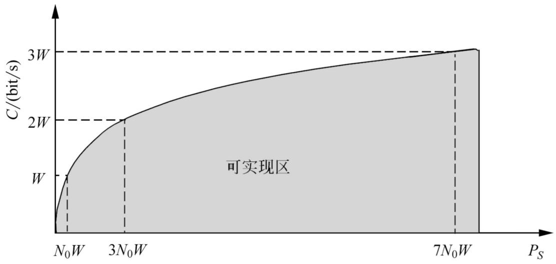
$$= \lim_{W\to\infty} \frac{P_s}{N_0} \cdot \log\left(1+\frac{P_s}{N_0W}\right)^{\frac{N_0W}{P_s}}$$

$$= \lim_{W\to\infty} \log_2 e \cdot \frac{P_s}{N_0} \approx 1.44\, P_s/N_0$$


对于平均功率

$$\lim_{P_s\to\infty} C(P_s) = \infty.$$

可以看到功率可以无限提升信道容量，但 $\lim_{P_s\to\infty}\frac{dC(P_s)}{dP_s} = 0$.
增长速度会不断减少.




(2) 有效利用功率
这里首先要讨论到一个性价比的关系了.

$$E_b = P_s/C.$$

那么直接代入公式

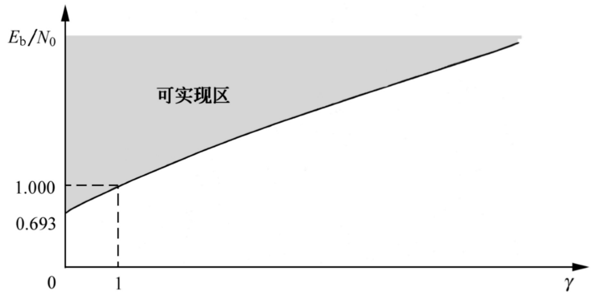
$$E_b = \frac{P_s}{C} = \frac{P_s}{W\log(1+\frac{P_s}{N_0W})}$$

看我们讨论

$$\frac{E_b}{N_0} = \frac{P_s}{N_0 W}\Big/\log\left(1+\frac{P_s}{N_0 W}\right)$$

那么这个性价比就是信噪比的函数. 设信噪比为r. 则为

$$\frac{E_b}{N_0} = \frac{r}{\log(1+r)}$$




$E_b/N_0$ 越小代表单位比特所需的功率越小. 那么最小值为

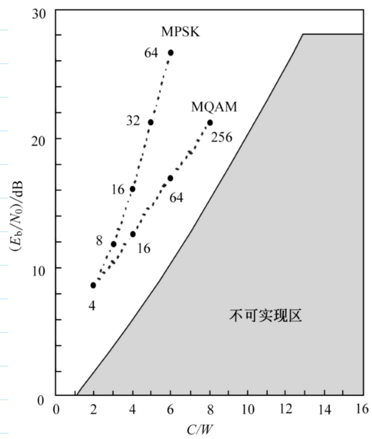
$$\min\frac{E_b}{N_0} = \lim_{r\to 0}\frac{r}{\log_2(1+r)} = \frac{1}{\log_2 e} = 0.693 \approx -1.6\text{dB}.$$

而这也对应于带宽趋于无限. 这一最小值也被称为香农限, 为0.693dB.
物理学中, 热噪声的功率谱密度为 $N_0 = kT$. 那么 $E_{min} = 0.693kT$.

(3). 功率受限下有效利用信道带宽的极限
上面讨论的是功率的性价比. 这里来讨论带宽的.

$$\frac{C}{W} = \log_2\left(1+\frac{P_s}{N_0W}\right) = \log_2\left(1+\frac{E_b}{N_0}\cdot\frac{C}{W}\right)$$

$$\Rightarrow \frac{E_b}{N_0} = \frac{2^{\frac{C}{W}}-1}{C/W}.$$




这个公式说明功率和带宽性价比之间的联系. 在理想情况下, 实际系统的实际曲线与这个越接近越好.

其他的情况不在过多描述. 可参阅应用信息论基础

## 4.5. MIMO系统的信道容量.

这个不是大纲要求, 但是对数学功底的提升有很大帮助.
设信道是限带的AWGN信道. 那么有接收r和信道增益h, 输入s, 噪声n的关系.

$$r = h\cdot s + n.$$

信道容量C为

$$C = B\log\left(1+\frac{\sigma_s^2}{\sigma_n^2}\right) = B\log\frac{\sigma_r^2}{\sigma_n^2}.$$

对于MIMO系统.

$$r_j = \sum_{i=1}^{n} h_{ij}\cdot x_i + n_i$$

所以可以写成矩阵的形式

$$\vec{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = H\vec{s} + \vec{n} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & \ddots & & h_{2n} \\ \vdots & & \ddots & \vdots \\ h_{m1} & \cdots & \cdots & h_{mn} \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_n \end{bmatrix}$$

对于这样一个信道, 自然也是输出为高斯分布时信道容量有最大值.
下面我们则需要求解输出信号的功率. 即方差. 对于向量来说应该是协方差矩阵

对于这样一个信道，当然也是输出为高斯分布时信道容量有最大值
下面我们则需要求解输出信号的功率，即方差，对于向量来说应该是协方差矩阵

$$E[rr^H] = E[(Hs+n)(Hs+n)^H] = E[(Hs+n)(s^HH^H+n^H)]$$

$$= E(Hs\cdot s^HH^H + Hsn^H + ns^HH^H + nn^H)$$

由于s和n相互独立，所以

$$E(Hsn^H + ns^HH^H) = 0$$

$$\Rightarrow E[rr^H] = H R_{ss} H^H + R_{nn}.$$

记协方差矩阵为Q
同理，噪声的协方差矩阵为$R_{nn}$，由循环对称复高斯分布的表达式

$$f_{u,Q}(x) = \det(\pi Q)^{-1}\exp(-(x-u)^HQ^{-1}(x-u))$$

可以推出，其信息熵为(零均值)

$$H(f_Q) = \log\det(\pi e Q)$$

而信道容量为

$$C = H(f_Q) - H(f_n)$$

$$= B\log_2\det(I_n + HR_{ss}H^HR_{nn}^{-1})$$

这是一般情形的表达式
若各个信号分量相互独立，各个噪声分量也相互独立，有

$$R_{ss} = \sigma_s^2\cdot I_n \qquad R_{nn} = \sigma_n^2\cdot I_n$$

信道容量则可以写成

$$C = B\log_2\det(I_n + \frac{\sigma_s^2}{\sigma_n^2}HH^H)$$

若信道矩阵也是一个正交阵，则可以写成

$$HH^H = |h|^2 I_n$$

上式可进一步改写为

$$C = B\log_2\det(I_n + \frac{\sigma_s^2}{\sigma_n^2}I_n\cdot|h|^2)$$

$$= B\log_2\det((1+\frac{\sigma_s^2}{\sigma_n^2}|h|^2)\cdot I_n)$$

$$= B\log_2(1+\frac{\sigma_s^2}{\sigma_n^2})^n\det(I_n)$$

$$= nB\log_2 1+\frac{\sigma_s^2}{\sigma_n^2}$$

可见正交和独立是非常好的性质，可以将MIMO系统等效为并行信道，从而可以简单的提高天线数来提升信道容量。