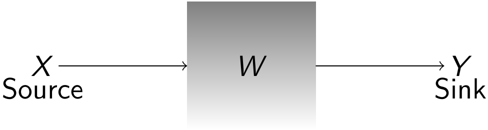
Linear Universal Decoder for Binary Memoryless Channels

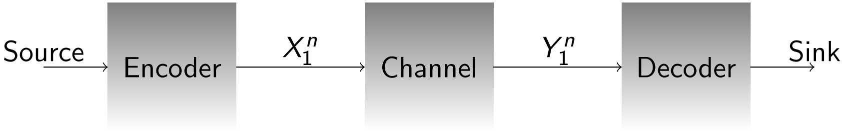
Rethnakaran Pulikkoonattu

July 16, 2009

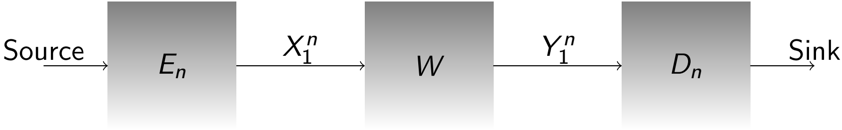
Digital Communication problem



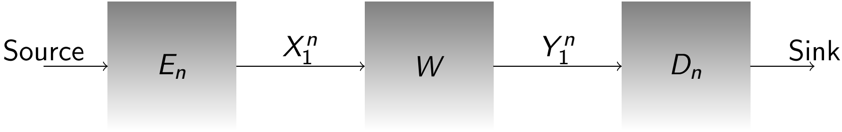
Shannon said...



Shannon said...

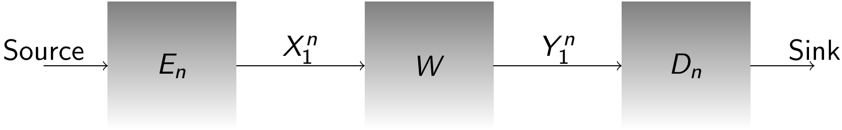


Shannon said...



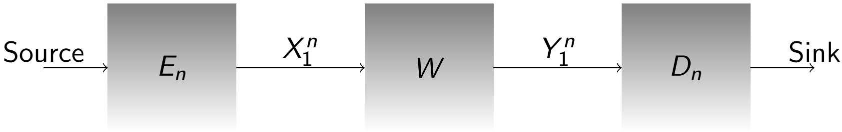
- For $R < C(W)$, there exist E_n and D_n which ensure reliable communication.

Shannon said...



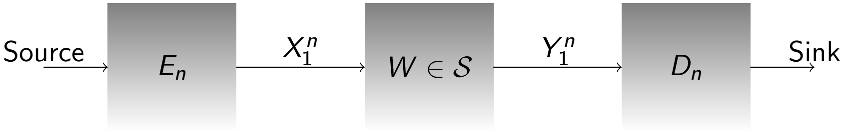
- For $R < C(W)$, there exist E_n and D_n which ensure reliable communication.
- E_n and D_n depends on W

Shannon said...

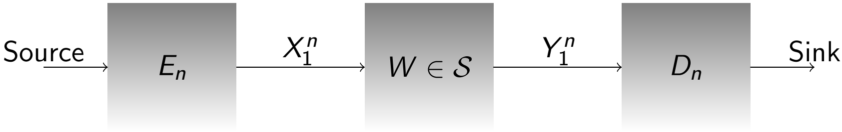


- For $R < C(W)$, there exist E_n and D_n which ensure reliable communication.
- E_n and D_n depends on W

Family of channels

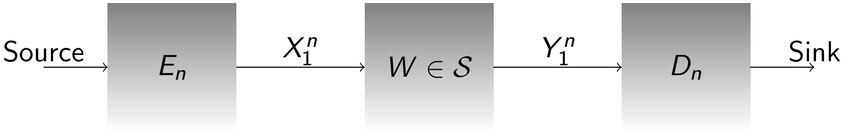


Family of channels



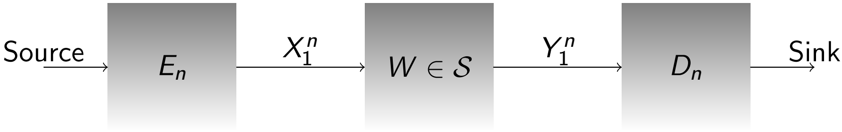
- Channel W is not known

Family of channels



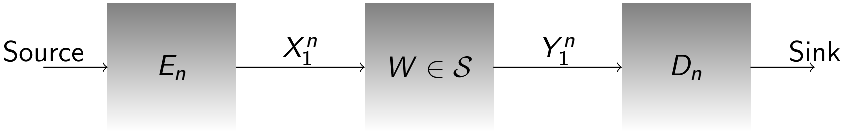
- Channel W is not known
- E_n and D_n cannot depend on W
- No feedback

Family of channels



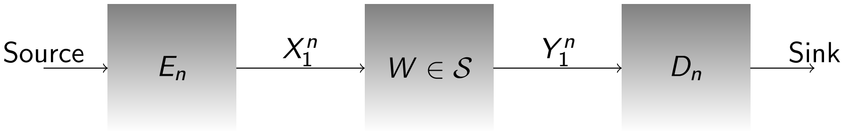
- Channel W is not known
- E_n and D_n cannot depend on W
- No feedback
- Reliable communication still possible? If so, at what rate?

Family of channels

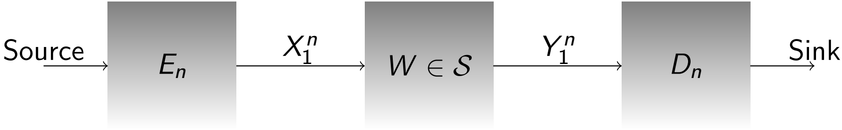


- Channel W is not known
- E_n and D_n cannot depend on W
- No feedback
- Reliable communication still possible? If so, at what rate?
- [Csiszar,Korner 1981],[Goppa 1975],[Narayan,Csiszar 1995],[Feder Lapidoth 1998],[Lapidoth,Telatar 1998],[Merhav,Kaplan,Lapidoth,Shamai 1994] and others

Compound channels



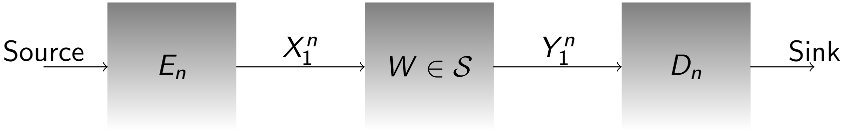
Compound channels



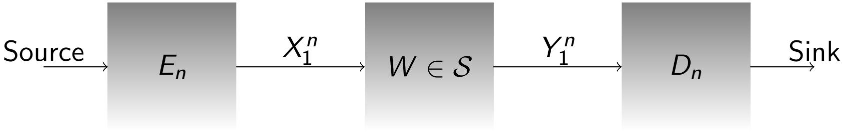
Compound capacity $C(\mathcal{S})$ is the largest rate possible [Blackwell,Breiman,Thomasian, 1959]

$$C(\mathcal{S}) = \inf_{P_X} \max_{W \in \mathcal{S}} I\left(P_X, W\right)$$

Coding for Compound channels

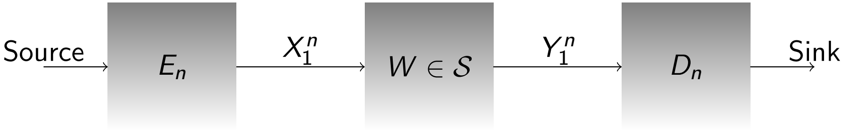


Coding for Compound channels



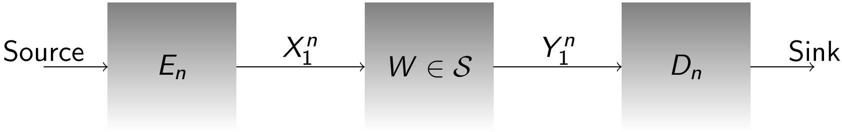
- Notion of Universality

Coding for Compound channels



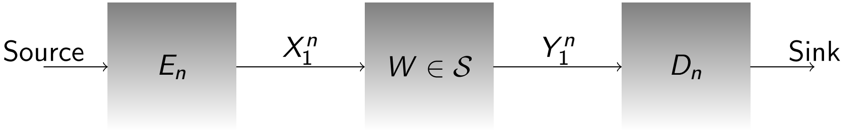
- Notion of Universality
- Universal Encoder:

Coding for Compound channels



- Notion of Universality
- Universal Encoder:Find a fixed P_X and a random codebook

Coding for Compound channels



- Notion of Universality
- Universal Encoder:Find a fixed P_X and a random codebook
- Universal Decoder

Universal Decoder

- Maximum Mutual Information (MMI) decoder [Goppa 1975]

Universal Decoder

- Maximum Mutual Information (MMI) decoder [Goppa 1975]
- MMI is universal
- Computes empirical mutual information (EMI) of a received word y with all elements $x_m \in \mathcal{C}$ of the codebook

Universal Decoder

- Maximum Mutual Information (MMI) decoder [Goppa 1975]
- MMI is universal
- Computes empirical mutual information (EMI) of a received word y with all elements $x_m \in \mathcal{C}$ of the codebook
- Pick the x_m with the maximum EMI

Universal Decoder

- Maximum Mutual Information (MMI) decoder [Goppa 1975]
- MMI is universal
- Computes empirical mutual information (EMI) of a received word y with all elements $x_m \in \mathcal{C}$ of the codebook
- Pick the x_m with the maximum EMI
- Complexity is exponential for random codebook

Universal Decoder

- Maximum Mutual Information (MMI) decoder [Goppa 1975]
- MMI is universal
- Computes empirical mutual information (EMI) of a received word y with all elements $x_m \in \mathcal{C}$ of the codebook
- Pick the x_m with the maximum EMI
- Complexity is exponential for random codebook
- Complexity is exponential also for structured codebook (Say a tree code)

Linear Decoder

A linear decoder induced by a single metric d is a rule given by

$$D_n(y) = \arg max_m d^n\left(x_m, y\right)$$

where

$$d^n\left(x_m, y\right) = \frac{1}{n}\sum_{i=1}^{n} d\left(x_m(i), y(i)\right) = \mathbb{E}_{\hat{P}(x_m,y)}[d]$$

Linear Decoder

A linear decoder induced by a single metric d is a rule given by

$$D_n(y) = \arg max_m d^n(x_m, y)$$

where

$$d^n(x_m, y) = \frac{1}{n}\sum_{i=1}^{n} d(x_m(i), y(i)) = \mathbb{E}_{\hat{P}(x_m, y)}[d]$$

- d^n has an additive structure

Linear Decoder

A linear decoder induced by a single metric d is a rule given by

$$D_n(y) = \arg max_m d^n\left(x_m, y\right)$$

where

$$d^n\left(x_m, y\right) = \frac{1}{n}\sum_{i=1}^{n} d\left(x_m(i), y(i)\right) = \mathbb{E}_{\hat{P}(x_m,y)}[d]$$

- d^n has an additive structure
- $d^{ML}(u, v) = \log W(v|u)$ is a linear decoder for DMC

Linear Decoder

A linear decoder induced by a single metric d is a rule given by

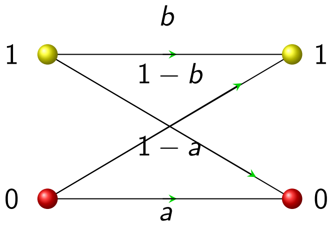
$$D_n(y) = \arg max_m d^n\left(x_m, y\right)$$

where

$$d^n\left(x_m, y\right) = \frac{1}{n}\sum_{i=1}^{n} d\left(x_m(i), y(i)\right) = \mathbb{E}_{\hat{P}(x_m, y)}[d]$$

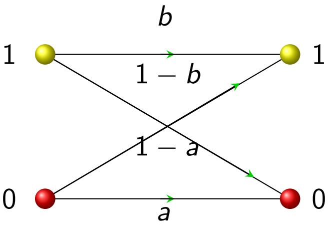
- $d^{ML}(u, v) = \log W(v|u)$ is a linear decoder for DMC
- Why do we care about linear decoders?

Compound Binary Memoryless channels



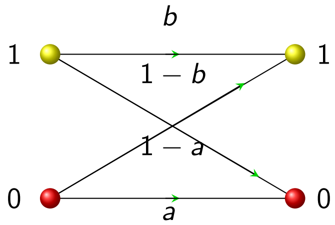
A tale of two games!

Compound Binary Memoryless channels



A tale of two games! Game 1: Find universal input distribution P

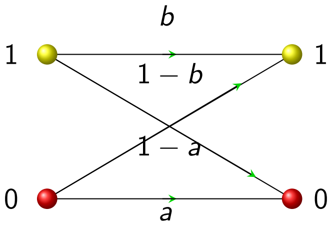
Compound Binary Memoryless channels



A tale of two games! Game 1: Find universal input distribution P

$$P_{opt} = \arg\max_{P} \inf_{W \in \mathcal{S}} \frac{I(P, W)}{C(W)}. \quad (1)$$

Compound Binary Memoryless channels



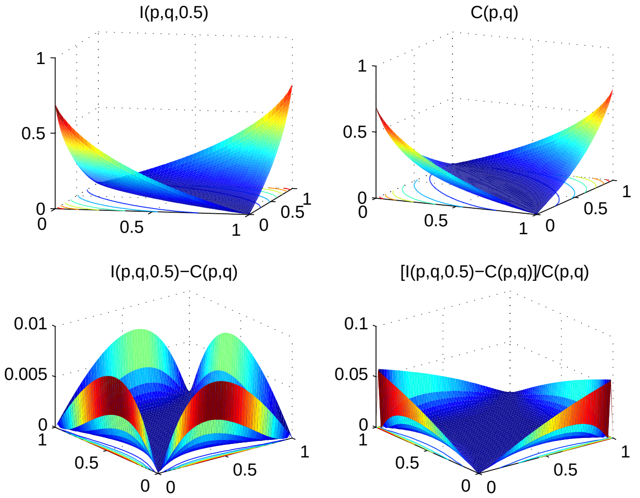
A tale of two games! Game 1: Find universal input distribution P

$$P_{opt} = \arg\max_{P} \inf_{W \in \mathcal{S}} \frac{I(P, W)}{C(W)}. \qquad (1)$$

Game 2: Find linear universal decoder

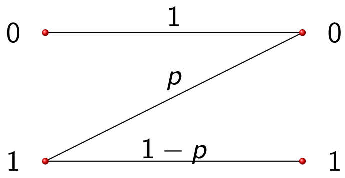
Uniform distribution does well

- Gets over 94.20% of the capacity!



Worst case binary channel

[Majani 1991], [Shulman, Feder 2004]



- Z channel when $p \to 1$
- Most of the channels does get over 98% (Simulation)

Uniform priors

The Z channel capacity is

$$\begin{aligned} C &= \max_u I(X;Y) \\ &= \max_u h(up) - uh(p) \end{aligned}$$

Solving $\frac{\partial}{\partial u}\left[h(up) - uh(p)\right] = 0$, we get the capacity achieving input prior $u^\star$ and it is,

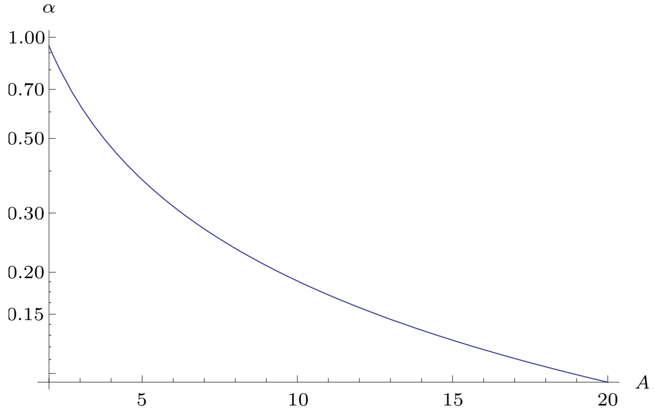
$$u^\star = \frac{p^{\frac{p}{1-p}}}{1 + (1-p)p^{\frac{p}{1-p}}}$$

The worst possible Z channel is in the limit $p \to 1$.

$$\begin{aligned} \lim_{p \to 1} u(p) &= \lim_{p \to 1} \frac{p^{\frac{p}{1-p}}}{1 + (1-p)p^{\frac{p}{1-p}}} \\ &= \frac{1}{e} \end{aligned}$$

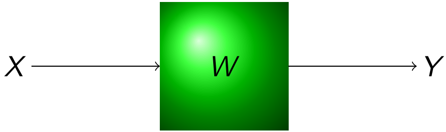
Then,

Uniform prior in general



- Input alphabet size $\mathcal{A}$
- Uniform gets to $\alpha = \frac{e}{\mathcal{A}\log_2(e)}$ times the capacity

Mutual Information



$$
\begin{aligned}
I(P_X, W) &\triangleq I(X;Y) \\
&= H(Y) - H(Y|X) \\
&= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \\
&= \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \\
&= D\left(P_{XY}(x,y) \| P_X(x)P_Y(y)\right)
\end{aligned}
$$

KL Distance

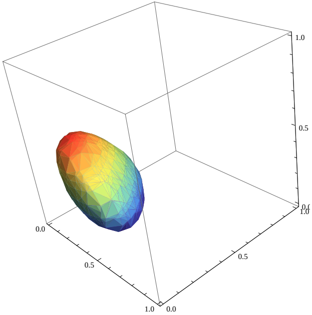
Definition

The Kullback Leibler distance (denoted by $D\left(P \| Q\right)$) between two distributions P and Q where both P and Q are defined over the same probability support space $supp(P)$ is defined as

$$D\left(P \| Q\right) = \sum_{u \in supp(P)} P(u) \log \frac{P(u)}{Q(u)}. \tag{2}$$

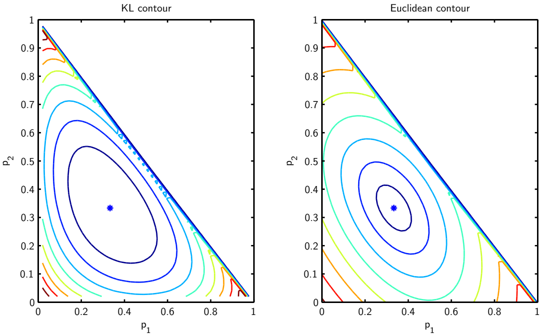
- Think of it as a distance between two distributions
- Has many (not all) properties of a metric (say Euclidean distance).

KL ball



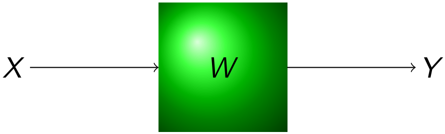
- Not quite the Euclidean ball!

KL versus Euclidean distance



- Not everything is circular!

Game 2: Linear Universal Decoder



Let

$$\begin{aligned} \mu &\doteq P_{X,Y}(x,y) = P_X(x)W(y|x) \\ \mu^p &\doteq P_X(x)P_Y(y) = P_X(x) \circ W(y|x). \end{aligned}$$

mutual information $I(X;Y)$ is,

$$I\left(W\right) \doteq D\left(\mu \| \mu^p\right).$$

Symmetric (Channel) capacity

Channel capacity $C(W)$ is given by

$$C(W) = \inf_{P_X} I\left(P_X, W\right) = \inf_{P_X} D\left(\mu \| \mu^p\right)$$

Symmetric capacity $I(W)$ is mutual information when P_X is uniform.

$$I(W) = \inf_{P_X(x)=\frac{1}{\mathcal{X}}} D\left(\mu \| \mu^p\right)$$

- Recall that for binary channels $I(W) \geq 0.94C(W)$.
- We are happy when we can do 94% of what the happiest lot do.

Game 2: Linear Decoders for Compound BMC

- Compound sets which are convex admit linear decoders [Csiszar, Narayan 1995]

Game 2: Linear Decoders for Compound BMC

- Compound sets which are convex admit linear decoders [Csiszar, Narayan 1995]
- Beyond convex possible! One sided sets admit linear decoders [Abbe,Zheng 2008]

Game 2: Linear Decoders for Compound BMC

- Compound sets which are convex admit linear decoders [Csiszar, Narayan 1995]
- Beyond convex possible! One sided sets admit linear decoders [Abbe,Zheng 2008]
- Union of one sided sets also admit generalized linear decoder![Abbe,Zheng 2008]

One sided sets

A set $\mathcal{S}$ is one sided if the following inequality is satisfied by all channels $\mu_0$in the set $\mathcal{S}$.

$$D\left(\mu_0 \| \mu_{\mathcal{S}}^{p}\right) \geq D\left(\mu_0 \| \mu_{\mathcal{S}}\right) + D\left(\mu_{\mathcal{S}} \| \mu_{\mathcal{S}}^{p}\right)$$

where $W_{\mathcal{S}}$ is the worst channel in the set $\mathcal{S}$, given by,

$$W_{\mathcal{S}} = \arg \min_{W \in cl(S)} I\left(P_X, W\right)$$

One sided sets

A set $\mathcal{S}$ is one sided if the following inequality is satisfied by all channels $\mu_0$in the set $\mathcal{S}$.

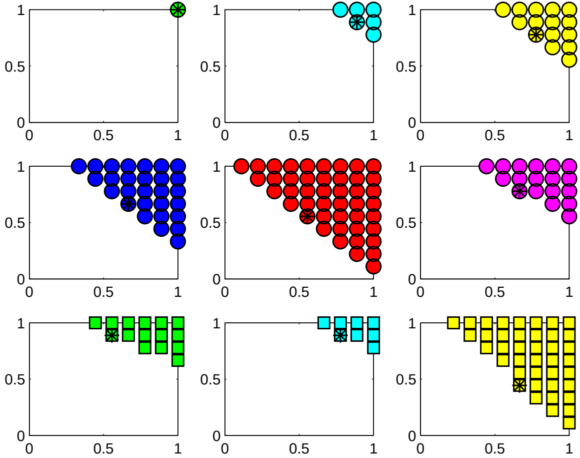
$$D\left(\mu_0 \| \mu_{\mathcal{S}}^p\right) \geq D\left(\mu_0 \| \mu_{\mathcal{S}}\right) + D\left(\mu_{\mathcal{S}} \| \mu_{\mathcal{S}}^p\right)$$

where $W_{\mathcal{S}}$ is the worst channel in the set $\mathcal{S}$, given by,

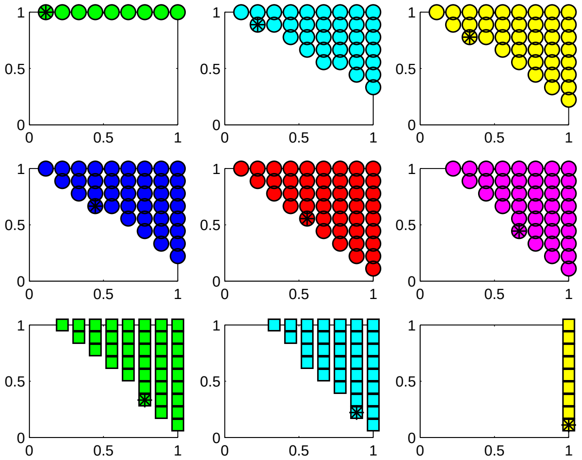
$$W_{\mathcal{S}} = \arg \min_{W \in cl(S)} I\left(P_X, W\right)$$

We can think of geometry now

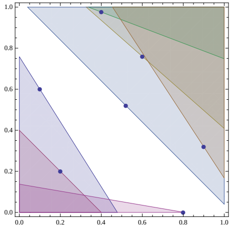
One sided sets



One sided sets



One sided sets:Characteristics

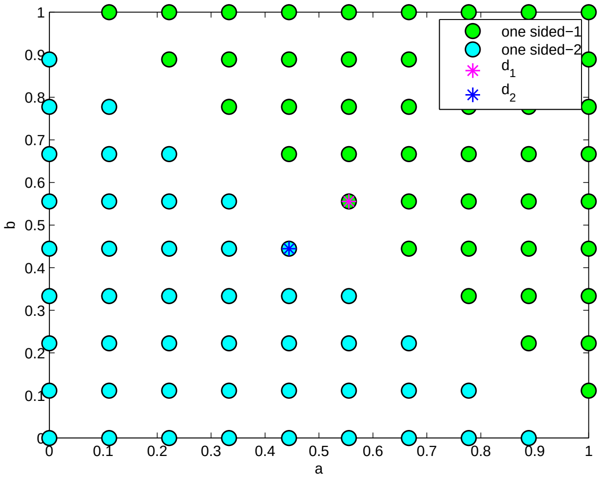


$$\Gamma_{\mathcal{S}}^{\text{onesided}}\left(W_s\right) = \left\{ a, b \middle| \frac{b - b_s}{a - a_s} \geq -\eta\left(W_s\right) \right\}$$

where

$$\eta\left(W_s\right) = \frac{\log\left(\frac{a_s}{1-a_s}\frac{1-a_s+b_s}{1+a_s-b_s}\right)}{\log\left(\frac{b_s}{1-b_s}\frac{1+a_s-b_s}{1-a_s+b_s}\right)}$$

Union of one sided sets



Can we go beyond union of one sided sets?

Y_eS!

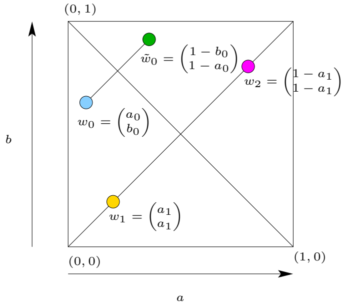
Arbitrary compound set of BMCs

Theorem

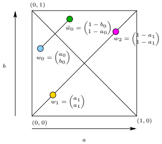
There exists a linear codebook and a generalized linear decoder such that by using this code on any binary symmetric channel W, the symmetric capacity $I(W)$ is achieved.

We get to MMI!Game 2 over.

Arbitrary BMC



Arbitrary BMC



Lemma

With Generalized Maximum Aposteriory (GMAP) rule, induced by two single letter metrics d_1, d_2,

$$
\begin{aligned}
d_1(y,x) &= \log \frac{W_1(y|x)}{W_1(y|0)+W_1(y|1)}, W_1 = BMC(a_1, a_1) \\
d_2(y,x) &= \log \frac{W_2(y|x)}{W_2(y|0)+W_2(y|1)}, W_2 = BMC(1-a_1, 1-a_1),
\end{aligned}
$$

the symmetric capacity $I(W)$ can be achieved for any binary memoryless channel W.

BMC continued

For a BMC W_0, and for any $a_1 \in [0, 1]$, a (mismatched) decoder tuned to one of the channels $W_1 = (a_1, a_1)$ or $W_2 = (1 - a_1, 1 - a_1)$ achieve the symmetric capacity $I(W)$.

BMC continued

Lemma

Let $\mathcal{B}^+$ be the set of BMCs with $a + b \leq 1$. For any BMC $W_0 \in \mathcal{B}^+$, with a mismatched decoder tuned to a BSC W_1 from the same region, i.e., $W_1 \in \mathcal{B}^+$ the following rate $I_{mis}(W_0, W_1)$ can be achieved.

$$I_{mis}(W_0, W_1) = D\left(\mu_0 \| \mu_0^p\right)$$

Proof

$$I_{\text{mis}}\left(W_0, W_1\right) = \inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_{\mu}[d_1]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right)$$

Proof

$$I_{\text{mis}}(W_0, W_1) = \inf_{\substack{\mu : \mu^p = \mu_0^p \\ \mathbb{E}_\mu[d_1] \geq \mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right)$$

$$= \begin{cases} \inf\limits_{\substack{\mu : \frac{a-a_0}{b-b_0} = \frac{1+a_0-b_0}{1+b_0-a_0} \\ a+b \leq a_0+b_0, a_1 \leq \frac{1}{2}}} D\left(\mu \| \mu_0^p\right), & W_1 \in \mathcal{B}^+ \\ \inf\limits_{\substack{\mu : \frac{a-a_0}{b-b_0} = \frac{1+a_0-b_0}{1+b_0-a_0} \\ a+b \geq a_0+b_0, a_1 \geq \frac{1}{2}}} D\left(\mu \| \mu_0^p\right), & W_1 \in \mathcal{B}^- \end{cases}$$

Proof

$$
\begin{aligned}
I_{\text{mis}}(W_0, W_1) &= \inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right) \\
&= \begin{cases} \inf_{\substack{\mu:\frac{a-a_0}{b-b_0}=\frac{1+a_0-b_0}{1+b_0-a_0} \\ a+b\leq a_0+b_0, a_1\leq\frac{1}{2}}} D\left(\mu \| \mu_0^p\right), & W_1 \in \mathcal{B}^+ \\ \inf_{\substack{\mu:\frac{a-a_0}{b-b_0}=\frac{1+a_0-b_0}{1+b_0-a_0} \\ a+b\geq a_0+b_0, a_1\geq\frac{1}{2}}} D\left(\mu \| \mu_0^p\right), & W_1 \in \mathcal{B}^- \end{cases} \\
&= \begin{cases} D\left(\mu_0 \| \mu_0^p\right) & a_1 \leq \frac{1}{2} \\ 0 & a_1 \geq \frac{1}{2} \end{cases} \\
&= D\left(\mu_0 \| \mu_0^p\right)
\end{aligned}
$$

Proof of main theorem

We need to prove,

$$\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu d_1 \vee \mathbb{E}_\mu d_2 \geq \mathbb{E}_{\mu_0} d_1 \vee \mathbb{E}_{\mu_0} d_2}} D\left(\mu \| \mu_0^p\right) = D\left(\mu_0 \| \mu_0^p\right)$$

Proof of main theorem

We need to prove,

$$\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu d_1 \vee \mathbb{E}_\mu d_2 \geq \mathbb{E}_{\mu_0} d_1 \vee \mathbb{E}_{\mu_0} d_2}} D\left(\mu \| \mu_0^p\right) = D\left(\mu_0 \| \mu_0^p\right)$$

w.l.o.g assume $a_1 < \frac{1}{2}$ and $a_0 + b_0 < 1$, then

$$\mathbb{E}_{\mu_0}\left[d_1\right] \vee \mathbb{E}_{\mu_0}\left[d_2\right] = \mathbb{E}_{\mu_0}\left[d_1\right]$$

Proof of main theorem

We need to prove,

$$\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu d_1 \vee \mathbb{E}_\mu d_2 \geq \mathbb{E}_{\mu_0} d_1 \vee \mathbb{E}_{\mu_0} d_2}} D\left(\mu \| \mu_0^p\right) = D\left(\mu_0 \| \mu_0^p\right)$$

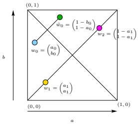
w.l.o.g assume $a_1 < \frac{1}{2}$ and $a_0 + b_0 < 1$, then

$$\mathbb{E}_{\mu_0}[d_1] \vee \mathbb{E}_{\mu_0}[d_2] = \mathbb{E}_{\mu_0}[d_1]$$

$$\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1] \vee \mathbb{E}_\mu[d_2] \geq \mathbb{E}_{\mu_0}[d_1] \vee \mathbb{E}_{\mu_0}[d_2]}} D\left(\mu \| \mu_0^p\right) = \inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1] \vee \mathbb{E}_\mu[d_2] \geq \mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right)$$

Proof of main theorem

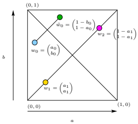
Note that, $\mu_0^p = \tilde{\mu}_0^p$.



Proof of main theorem

Note that, $\mu_0^p = \tilde{\mu}_0^p$.

$$= \min \left(\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu\|\mu_0^p\right), \inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_2]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu\|\mu_0^p\right) \right)$$

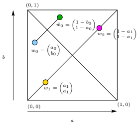


Proof of main theorem

Note that, $\mu_0^p = \tilde{\mu}_0^p$.

$$= \min\left(\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right), \inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_2]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right) \right)$$

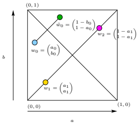
$$= \min\left(\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right), \inf_{\substack{\mu:\mu^p=\tilde{\mu}_0^p \\ \mathbb{E}_\mu[d_2]\geq\mathbb{E}_{\tilde{\mu}_0}[d_2]}} D\left(\mu \| \tilde{\mu}_0^p\right) \right)$$



Proof of main theorem

Note that, $\mu_0^p = \tilde{\mu}_0^p$.

$$
\begin{aligned}
&= \min\left(\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right), \inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_2]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right) \right) \\
&= \min\left(\inf_{\substack{\mu:\mu^p=\mu_0^p \\ \mathbb{E}_\mu[d_1]\geq\mathbb{E}_{\mu_0}[d_1]}} D\left(\mu \| \mu_0^p\right), \inf_{\substack{\mu:\mu^p=\tilde{\mu}_0^p \\ \mathbb{E}_\mu[d_2]\geq\mathbb{E}_{\tilde{\mu}_0}[d_2]}} D\left(\mu \| \tilde{\mu}_0^p\right) \right) \\
&= \min\left(D\left(\mu_0 \| \mu_0^p\right), D\left(\tilde{\mu}_0 \| \tilde{\mu}_0^p\right) \right) \\
&= D\left(\mu_0 \| \mu_0^p\right)
\end{aligned}
$$



Universal Codes

- Polar Codes [Arikan 2008] are provably capacity achieving in certain class of channels.

Universal Codes

- Polar Codes [Arikan 2008] are provably capacity achieving in certain class of channels.
- Polar Codes are symmetric capacity achieving for a given channel (Need the channel knowledge)

Universal Codes

- Polar Codes [Arikan 2008] are provably capacity achieving in certain class of channels.
- Polar Codes are symmetric capacity achieving for a given channel (Need the channel knowledge)
- Are Polar Codes universal?

Universal Codes

- Polar Codes [Arikan 2008] are provably capacity achieving in certain class of channels.
- Polar Codes are symmetric capacity achieving for a given channel (Need the channel knowledge)
- Are Polar Codes universal?
- Under successive cancellation decoding, Not universal (in general).[Hassani,Korada,Urbanke 2009]

Universal Codes

- Polar Codes [Arikan 2008] are provably capacity achieving in certain class of channels.
- Polar Codes are symmetric capacity achieving for a given channel (Need the channel knowledge)
- Are Polar Codes universal?
- Under successive cancellation decoding, Not universal (in general).[Hassani,Korada,Urbanke 2009]
- Under GMAP are they universal?

Polar Codes as universal codes

- For certain special compound sets, Polar codes are universal.

Polar Codes as universal codes

- For certain special compound sets, Polar codes are universal.
- Degraded sets

Polar Codes as universal codes

- For certain special compound sets, Polar codes are universal.
- Degraded sets

$$W_{\mathcal{S}} = \left\{ W | W \times W' = W_0 \right\}$$

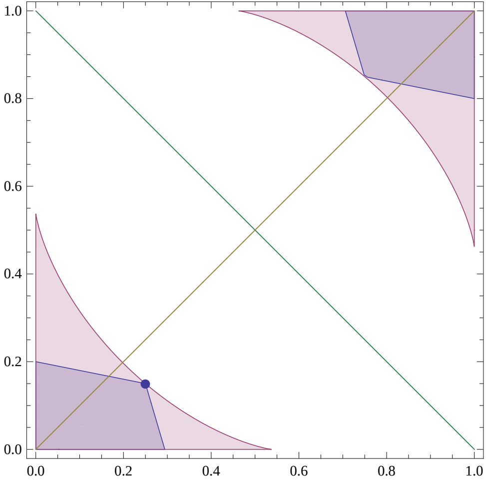
Polar Codes as universal codes

- For certain special compound sets, Polar codes are universal.
- Degraded sets

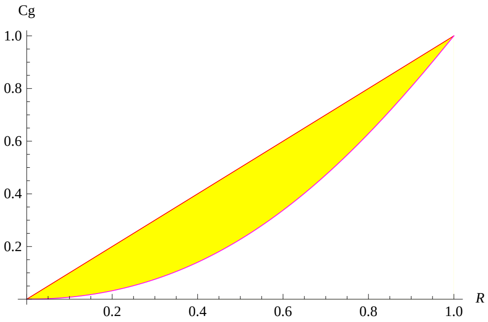
$$W_{\mathcal{S}} = \left\{ W | W \times W' = W_0 \right\}$$

- Polar codes constructed for W_0 is universal for $W_{\mathcal{S}}$ [Hassani,Korada,Urbanke]

Sufficient condition



Gap to compound capacity with Polar codes



Conclusion

- Coding problem for compound channels
- Existence of a universal encoder and decoder for binary memoryless channels
- Linear Decoder exist for compound set of binary memoryless channels
- Polar codes as a universal candidate

Open problems etc

-

Thanks for your patience!
Questions! The story of e follows....

And the story of *e*!

And the story of e!