Broadcast and relay channels

Introductory notes

Paulo J S G Ferreira

SPL — Signal Processing Lab / IEETA
Univ. Aveiro, Portugal

April–May 2010

Contents

Contents

Contents

Contents

Contents

Contents

Broadcast channel

- One sender
- At least two receivers
- Example: TV broadcast
- Best receivers may need extra information to display better picture
- Worst receivers need just the basic information
- What can be achieved?

Simple example

- Orthogonal broadcast channels
- Separate channels to each receiver
- One with capacity C_1, the other with C_2
- Capacity region: rectangle
- Superposition makes things more difficult — and interesting

Speaking two languages

- Consider a speaker who speaks languages A and B
- Listener 1 understands only A
- Listener 2 understands only B
- Imagine that there are 2^{20} words in the language A
- Same for B
- What are the rate options?

Speaking two languages

- Let the speaker pronounce one word / second
- The speaker may communicate 20 bps to listener 1 by speaking language A
- He can communicate 20 bps to listener 2 by speaking language B
- Any rate pair $R_1 + R_2 = 20$ can be achieved by time-sharing
- The surprise: this can be improved

Speaking two languages

- The speaker divides his time equally between languages A and B
- Extra information can be encoded in the order of the words
- 50 words of each language give rise to $\binom{100}{50}$ different orderings
- Selection of one ordering conveys extra information to both listeners
- Since $\binom{n}{k} \approx 2^{nH(k/n)}$, this shows that one can send 10 bps to each listener and at least one extra bit to both

The binomial

- The terms $\binom{n}{k} p^k q^{n-k}$ add up to one and are positive
- Thus, for $p = k/n$

$$
\begin{aligned}
1 &\geq \binom{n}{k} p^k q^{n-k} \\
&= \binom{n}{k} 2^{k \log p + (n-k) \log q} \\
&= \binom{n}{k} 2^{n\left(\frac{k}{n} \log p + \frac{(n-k)}{n} \log q\right)} \\
&= \binom{n}{k} 2^{n(p \log p + q \log q)} = \binom{n}{k} 2^{-nH(k/n)}
\end{aligned}
$$

- More precise results (including lower bounds) are possible

Definition

- A broadcast channel consists of:
  - An input alphabet $\mathcal{X}$
  - Two output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$
  - A probability transition function $p(y_1, y_2|x)$
- The channel is memoryless if

$$p(y_1^n, y_2^n|x^n) = \prod p(y_{1i}, y_{2i}|x_i)$$

Code

- A code for a broadcast channel must account for two rates: R_1 and R_2
- Notation: $(2^{nR_1}, 2^{nR_2}, n)$
- Encoder:

$$X : \{1, 2, \dots, 2^{nR_1}\} \times \{1, 2, \dots, 2^{nR_2}\} \mapsto \mathcal{X}^n$$

- Decoder 1:

$$g_1 : \mathcal{Y}_1^n \mapsto \{1, 2, \dots, 2^{nR_1}\}$$

- Decoder 2:

$$g_2 : \mathcal{Y}_2^n \mapsto \{1, 2, \dots, 2^{nR_2}\}$$

Error probability

- There is an error when the decoded message is not equal to the broadcast message
- The error probability is thus

$$P_e = P(g_1(Y_1^n) \neq W_1 \text{ or } g_2(Y_2^n \neq W_2)$$

- It is assumed that W_1 and W_2 are uniformly distributed

Achievable rates

- A pair of rates (R_1, R_2) is achievable if there exists a sequence of codes $(2^{nR_1}, 2^{nR_2}, n)$ with $P_e \to 0$
- The capacity region is the closure of the set of achievable rates

Capacity region

- The error for receiver 1 depends on $p(x^n, y_1^n)$ but not on the joint distribution $p(x^n, y_1^n, y_2^n)$
- Alternatively, $P_e \to 0$ is equivalent to P_e(each channel) $\to 0$
- Thus, the capacity region of a broadcast channel depends only on the conditional marginal distributions $p(y_1|x)$ and $p(y_2|x)$

Code with common information

- Imagine that the messages have common and independent parts
- In this case an extra rate term, R_0, is required
- Notation: $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n)$
- Again, assume that W_0, W_1, W_2 are uniformly distributed
- The error probability is

$$P_e = P(g_1(Y_1^n) \neq (W_0, W_1) \text{ or } g_2(Y_2^n \neq (W_0, W_2))$$

Physically degraded broadcast

- One receiver can be further downstream than the other
- It would therefore receive a degraded version of the signal
- Physically degraded broadcast channel: when

$$p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1)$$

- Notation: $X \to Y_1 \to Y_2$

Stochastically degraded broadcast

- If there exists $q(y_2|y_1)$ such that

$$p(y_2|x) = \sum_{y_1} p(y_1|x)q(y_2|y_1)$$

- The capacity regions of the physically degraded and stochastically degraded channels are the same
- Why? Because the capacity depends on the conditional marginals
- Assumes physically degraded broadcast from now on

Example

- Alphabets: $\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2 = \{0, 1\}$
- Channels:
$$P_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} q & p \\ p & q \end{pmatrix}$$
that is, a noiseless channel and a BSC(p)
- One can transmit at a rate of $C_2 = 1 - H(p)$ to receiver 2
- The pair $(R_1, R_2) = (C_2, C_2)$ is therefore achievable
- One can transmit at rate 1 to channel 1, with a resulting rate of zero to receiver 2
- Thus, (C_2, C_2) and $(1, 0)$ are achievable
- Timesharing shows that the line joining $(1, 0)$ to (C_2, C_2) is achievable
- Can one do better?

Example

- Design a code for a somewhat noisier channel
- The code can be decoded by the BSP(p)
- The extra redundancy can be used to convey some extra information to the noiseless channel
- Geometrically, the code looks like a set of clouds clustered around a center
- A message is a pair (c, s) where
  - c identifies the cloud
  - s identifies a specific point inside the cloud
- The perfect receiver can decode c and s
- The other receiver can decode c
- This makes it possible to exceed the time-sharing bound

Capacity region

- The capacity region for the degraded broadcast channel $X \to Y_1 \to Y_2$ is the convex hull of the closure of all pairs (R_1, R_2) that satisfy

$$\begin{aligned} R_2 &\le I(U; Y_2) \\ R_1 &\le I(X; Y_1 | U) \end{aligned}$$

  for some joint distribution $p(u)p(x|u)p(y_1, y_2|x)$
- U is an auxiliary random variable, jointly distributed with X
- The cardinality of U cannot exceed $|\mathcal{X}|$, $|\mathcal{Y}_1|$ or $|\mathcal{Y}_2|$

Main idea

- The auxiliary random variable U is associated with a cloud of points
- Each cloud consists of 2^{nR_1} codewords X^n
- The clouds can be distinguished by both receivers
- The better receiver can seen individual points inside each cloud

Proof: codebook

- Select $p(u)$ and $p(x|u)$
- Generate 2^{nR_2} independent codewords $U(w_2)$ of length n according to $\prod p(u_i)$
- For each codeword $U(w_2)$, generate 2^{nR_1} independent codewords $X(w_1, w_2)$ according to $\prod p(x_i|u_i(w_2))$
- The $u(i)$ play the role of cloud centers
- $x(i, j)$ is the jth satellite codeword in the ith cloud
- The cloud center is never sent

Proof: decoding

- Receiver 2 looks for the unique W_2 such that $(U(w_2), Y_2)$ is jointly typical
- If it cannot find it (because it does not exist or because there are more than one) it flags an error
- Receiver 1 looks for the unique (W_1, W_2) such that $(U(W_2), X(W_1, W_2), Y_1)$ is jointly typical
- If it cannot find it, it flags an error

Proof: error analysis

- The condition $R_2 < I(U; Y_2)$ guarantees that W_2 is correct w.h.p. because there are $2^{nI(U;Y_2)}$ distinguishable points as observed by Y_2
- The extra information is viewed as noise by Y_2
- The condition $R_1 \leq I(X; Y_1|U)$ guarantees that Y_1 decodes W_1 correctly w.h.p. given that it has already decoded W_2

Broadcasting common data 1

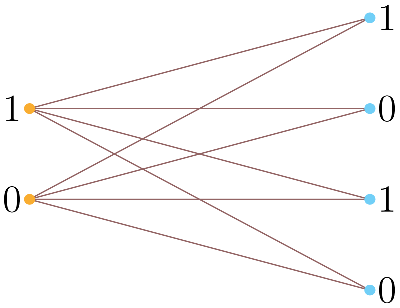
- Let the rate pair (R_1, R_2) be achievable for a broadcast channel with independent information
- Then the rate triple $(R_0, R_1 - R_0, R_2 - R_0)$ is achievable, provided that $R_0 \leq \min(R_1, R_2)$
- The idea is simple: reduce the independent rates by R_0, and have common information transmitted at that rate instead

Broadcasting common data 2

- Let the rate pair (R_1, R_2) be achievable for a degraded broadcast channel
- Then the rate triple $(R_0, R_1, R_2 - R_0)$ is achievable, provided that $R_0 \leq R_2$
- Why? In the degraded channel, the better receiver always decodes all the information sent to the worst one
- Thus the better receiver rate needs not be reduced

A pair of BSC

- Consider two BSC “in parallel”:



A pair of BSC

- This can be converted to a degraded broadcast channel
- How? a BSC(a) in cascade with a BSC(b) has a matrix given by

$$\begin{pmatrix} \bar{a} & a \\ a & \bar{a} \end{pmatrix} \begin{pmatrix} \bar{b} & b \\ b & \bar{b} \end{pmatrix} = \begin{pmatrix} \bar{a}\bar{b} + ab & \bar{a}b + a\bar{b} \\ a\bar{b} + \bar{a}b & ab + \bar{a}\bar{b} \end{pmatrix}$$

- In other words, it is a BSC(c), with $c = a * b \equiv a\bar{b} + \bar{a}b$
- It follows that

$$a = \frac{c - b}{1 - 2b}$$

A pair of BSC

- So a pair BSC(p_1), BSC(p_2) in parallel can be converted to a cascade
- The cascade is formed by BSC(p_1) and BSC(α), where

$$\alpha = \frac{p_2 - p_1}{1 - 2p_1}$$

- The capacity region can now be found from

$$R_2 \le I(U; Y_2)$$
$$R_1 \le I(X; Y_1 | U)$$

The capacity region

- The auxiliary r.v. U has to be binary (cardinality bound)
- It has to be jointly distributed with X, so can be thought as being passed through a BSC to yield X
- Let the connection to X use a BSC with parameter β. Then,

$$\begin{aligned} I(U;Y_2) &= H(Y_2) - H(Y_2|U) \\ &= 1 - H(\beta * p_2) \end{aligned}$$

$$\begin{aligned} I(X;Y_1|U) &= H(Y_1|U) - H(Y_1|X,U) \\ &= H(Y_1|U) - H(Y_1|X) \\ &= H(\beta * p_1) - H(p_1) \end{aligned}$$

The capacity region

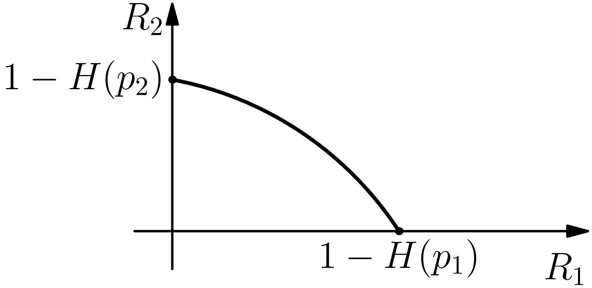
- What if $\beta = 0$?
- By definition, $\beta * p_2 = \beta(1 - p_2) + (1 - \beta)p_2 = p_2$
- Thus, $R_2 = 1 - H(p_2)$
- On the other hand, $R_1 = 0$
- Means only transmitting to Y_2, at maximum rate

The capacity region

- What if $\beta = 1/2$?
- By definition, $\beta * p_1 = \beta(1 - p_1) + (1 - \beta)p_1 = \frac{1}{2}$
- Then $R_2 = 0$
- On the other hand, $R_1 = 1 - H(p_1)$
- Means only transmitting to Y_1, at maximum rate

The capacity region

- The values $\beta = 0$ and $\beta = 1/2$ yield corner points in the rate region



Contents

Slepian-Wolf

- Encoding of X requires rate $R > H(X)$
- Encoding X and Y together requires rate $R > H(X, Y)$
- To encode X and Y separately, a rate $R_1 + R_2 > H(X) + H(Y)$ is sufficient
- Slepian-Wolf: encoding X, Y separately requires $R_1 + R_2 > H(X, Y)$

Slepian-Wolf

- Consider correlated r.v. U, V with distribution $p(u, v)$
- Consider i.i.d. copies (U_i, V_i) drawn according $p(u, v)$
- How many bits about U and how many about V are needed, so that the combined description conveys both U and V (with arbitrarily small error probability)?
- Encoders operate with knowledge of the other stream
- Both streams are available to the decoder

Slepian-Wolf

- Let (U_i, V_i) be i.i.d. discrete r.v.
- Let the encoder functions be $i_n : \mathcal{U}^n \longmapsto 2^{nR_1}$, $j_n : \mathcal{V}^n \longmapsto 2^{nR_2}$
- The reconstruction functions are $\hat{u}^n(i_n, j_n)$, $\hat{v}^n(i_n, j_n)$
- Then, $P\{(\hat{U}^n, \hat{V}^n) \neq (U^n, V^n)\} \to 0$ if and only if

$$
\begin{aligned}
R_1 &> H(U|V) \\
R_2 &> H(V|U) \\
R_1 + R_2 &> H(U, V)
\end{aligned}
$$

Slepian-Wolf

- Each encoder operates independently of the other
- Ignoring correlation, one achieves only $R_1 + R_2 > H(U) + H(V)$
- Exploring it, we get $R_1 + R_2 > H(U, V)$
- Thus, SW demonstrates that encoding of two correlated sources is possible at a rate equal to their joint entropy...
- ... even without communication between the two encoders
- Achieves the same compression rate as an optimal single encoder that has both correlated data streams as inputs

Proof ideas

- The binning scheme is used in SW but useful in other contexts
- Randomly map all sequences to 2^{nR} bins
- Each sequence is assigned an index in $1, 2, \ldots, 2^{nR}$
- The encoder represents the sequence by the index
- A set of sequences that share the same index form a bin
- Similar to hash tables
- Decoding: look for a typical sequence in the given bin
- No knowledge about the typical set at the encoder

Proof ideas

- In the SW case: every u^n is binned in 2^{nR_1} bins
- Same for v^n, using 2^{nR_2} bins
- This is the equivalent to a random code generation
- Encoding U^n: use its bin number $i(U^n)$
- Encoding V^n: use its bin number $j(U^n)$
- The receiver has access to the bin numbers of U^n and V^n
- If there is only one jointly typical pair in the bin, it will make no error
- The receiver needs to know about the jointly typical set

Why does it work?

- To each typical sequence in X there corresponds a set of jointly typical sequences in Y
- There are "many" typical sequences in X
- However, given one, the set of jointly typical sequences is "smaller"
- An encoder can use this to describe X first, then Y given X (within the "smaller" set)
- When encoding independently, it is still possible to describe Y by an arbitrary index
- If there are sufficiently many indexes, the jointly typical pairs can still be identified by the receiver

Contents

Deterministic broadcast

- Defined by

$$Y_1 = \begin{cases} 1, & x = 1 \\ 0, & x = 2, 3 \end{cases} \qquad Y_2 = \begin{cases} 1, & x = 1, 2 \\ 0, & x = 3 \end{cases}$$

- Deterministic channel!
- The symbol 1 means 1 for both Y_1 and Y_2
- The symbol 3 means 0 for both Y_1 and Y_2
- However, the symbol 2 means 0 for Y_1 and 1 for Y_2
- Capacity region known, uses binning as in Slepian-Wolf

Capacity region

- It is the convex closure of the pairs described by

$$
\begin{aligned}
R_1 &< H(Y_1) \\
R_2 &< H(Y_2) \\
R_1 + R_2 &< H(Y_1, Y_2)
\end{aligned}
$$

Proof ideas

- Goal: send (i, j) to receivers 1 and 2, respectively
- We want to control Y_1 and Y_2 using X
- Fix $p(x)$, consider $y_1 = f_1(x)$, $y_2 = f_2(x)$ with f_1 and f_2 deterministic
- $p(x)$ induces a joint distribution $p(x, y_1, y_2)$ and marginal $p(y_1, y_2)$
- Do a random binning into 2^{nR_1} and 2^{nR_2} bins
- Seek the rates for which the bins contain a jointly typical (y_1^n, y_2^n) w.h.p.
- Let there be a jointly typical pair in bin (i, j)
- Since y_1^n and y_2^n are deterministic functions of x^n, one can look for the x^n that yields y_1^n, y_2^n
- This would send i to Y_1 and j to Y_2, as required

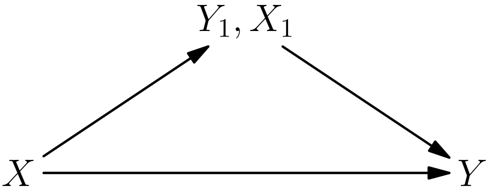
Contents

Relay channel

- One sender
- One receiver
- Intermediate nodes (relays)
- The relay nodes are intended to help the sender and receiver to communicate

Relay channel

- The simplest relay channel combines
  - A broadcast channel ($X \mapsto Y_1, Y$)
  - A multiple access channel ($X_1, X \mapsto Y$)



Relay channel

It is described by:

- $\mathcal{X}$, $\mathcal{X}_1$, $\mathcal{Y}_1$ and $\mathcal{Y}$
- a set $p(y, y_1|x, x_1)$, one for each $(x, x_1) \in \mathcal{X} \times \mathcal{X}_1$
- The problem is to find the capacity from X to Y

Codes for the relay channel

- A code $(2^{nR}, n)$ for the relay channel is described by:
- A set of integers:

$$\mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$$

- An encoder:

$$X : \{1, 2, \ldots, 2^{nR}\} \mapsto \mathcal{X}^n$$

- Relay functions, that map X_1 to Y_1
- A decoder:

$$g : \mathcal{Y}^n \mapsto \{1, 2, \ldots, 2^{nR}\}$$

Causality, memory

- The relay input is allowed to depend only on past observations
- This is important in practice, as relays cannot anticipate or predict the future
- The channel is also assumed memoryless
- (Y_i, Y_{1i}) depends on the past only through the current transmitted symbols (X_i, X_{1i})

Error probability

- Let the message $w \in \mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$ be sent
- The conditional error probability is

$$\lambda(w) = P\{g(Y) \neq w | w \text{ sent}\}$$

- The (average) error probability, assuming uniformly distributed codewords, is

$$P_e = \frac{1}{2^{nR}} \sum_{w} \lambda(w)$$

Achievable rate

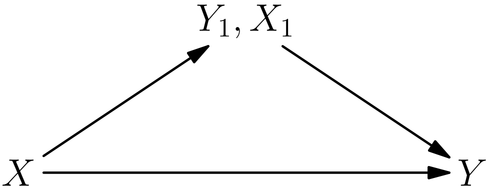
- The rate R is achievable if there exists a sequence of codes $(2^{nR}, n)$ such that $P_e \to 0$ (as $n \to \infty$, of course)
- The capacity of a relay channel is the supremum of the set of achievable rates

Capacity bound

- The capacity of the general relay channel is not known
- The following upper bound is known: the capacity is bounded by

$$C \leq \sup_{p(x,x_1)} \min\{I(X, X_1; Y), I(X; Y, Y_1|X_1)\}$$

- The first term bounds the maximum rate from X and X_1 to Y
- The second term bounds the rate from X to Y and Y_1



Contents

Broadcast capacity

- Consider a memoryless broadcast channel with r receivers, $Y_1, Y_2, \ldots, Y_r$
- It is possible to communicate to receiver k alone at capacity

$$C_k = \max_{p(x)} I(X; Y_k)$$

- An optimal code for Y_k will not be optimal to the remaining receivers
- Is it possible to communicate at C_k to all receivers?
- If so, does it contradict the definition of capacity?

Possible or impossible?

- There are a series of cycles $i = 1, 2, 3, \ldots$
- Within each cycle there are r segments intended to each of the r receivers
- The code used for the kth receiver is a $(2^{nC_k}, n)$ code
- The value of n depends on the cycle i and segment: $n = n(i, k)$
- In the cycle/segment (i, k) the kth receiver gets $n(i, k)C_k$ bits for its $n(i, k)$ transmissions
- The idea is to let $n(i, k)$ grow to infinity fast enough so that $n(i, k)$ approaches the total transmission time up to cycle/segment (i, k), denoted by $N(i, k)$
- The information rate for the kth receiver is

$$\frac{n(i, k)C_k}{N(i, k)}$$

- It approaches capacity! Can you explain the paradox?

Broadcast capacity

- The capacity region is the set of simultaneously achievable rates
- See: Thomas M. Cover. Comments on Broadcast Channels. *IEEE Trans. Inform. Theory* 44(6) p. 2524-2530 Oct. 1998