Lecture 6: Broadcast Channels

Jun Chen

1 Gelfand-Pinsker Coding:

The capacity of a discrete memoryless channel with random state $\mathbf{S}$ known to the encoder is given by

$$C = \max_{p(u,x|s)} [I(U;Y) - I(U;S)],$$

where the maximum is over all joint distributions of the form $p(s)p(u,x|s)p(y|x,s)$, where U is a finite alphabet auxiliary random variable.

The random coding argument, assuming discrete, finite alphabets and unconstrained input, can be outlined as follows. First generate $\exp\{n(I(U,Y) - \epsilon)\}$ i.i.d. sequences U, according to the uniform distribution over the set of typical U. Next, distribute these sequences uniformly over 2^{nR} bins. For each sequence $\mathbf{u}$ let $i(\mathbf{u})$ be the index of the bin containing $\mathbf{u}$. For encoding, given the state vector $\mathbf{S}$ and the message W, look in bin W for a sequence $\mathbf{U}$ such that $(\mathbf{U}, \mathbf{S})$ is jointly typical. Declare an error if no such $\mathbf{U}$ can be found. If the number of sequences in bin W is larger than $\exp\{n(I(U;S) + \epsilon)\}$, the probability of finding no such $\mathbf{U}$ decreases to zero exponentially as n increases. Next, choose $\mathbf{X}$ such that $(\mathbf{X}, \mathbf{U}, \mathbf{S})$ is jointly typical and send it through the channel. At the decoder look for the unique sequence $\mathbf{U}$ such that $(\mathbf{U}, \mathbf{Y})$ is jointly typical. Declare an error if more than one or no such sequence exist. Then set the estimate $\hat{W}$ equal to the index of the bin containing the obtained sequence $\mathbf{U}$. If $R < I(U;Y) - I(U;S) - \epsilon - \delta$, the probability of error averaged over all codes decreases exponentially to zero as $n \to \infty$. This shows the existence of a code that achieves rate R with arbitrarily small probability of error.

2 Costa's Dirty Paper Coding:

Model: $Y = X + S + Z$, where S and Z are independent random variables distributed according to $\mathcal{N}(0, Q)$ and $\mathcal{N}(0, N)$.

We consider $U = X + \alpha S$, where X and S are independent random variables distributed according to $\mathcal{N}(0, P)$ and $\mathcal{N}(0, Q)$, respectively, and α is a parameter to be determined. Note that there could be a loss of generality in restricting attention to such U, but as we shall see, the derived answer is clearly optimal. Recalling that $Y = X + S + Z$ with Z distributed according to $\mathcal{N}(0, N)$, the relevant mutual informations can be calculated to yield

$$\begin{aligned}
I(U;Y) &= h(Y) - h(Y|U) \\
&= h(Y) + h(U) - h(Y,U) \\
&= h(X+S+Z) + h(X+\alpha S) - h(X+S+Z, X+\alpha S) \\
&= \frac{1}{2}\log\left[(P+Q+N)(P+\alpha^2 Q)\right] - \frac{1}{2}\log\left[(P+Q+N)(P+\alpha^2 Q) - (P+\alpha Q)^2\right] \\
&= \frac{1}{2}\log\left[\frac{(P+Q+N)(P+\alpha^2 Q)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}\right]
\end{aligned}$$

and

$$\begin{aligned} I(U;S) &= h(U) - h(U|S) \\ &= h(X + \alpha S) - h(X) \\ &= \frac{1}{2}\log\left(\frac{P + \alpha^2 Q}{P}\right). \end{aligned}$$

Let

$$\begin{aligned} R(\alpha) &= I(U;Y) - I(U;S) \\ &= \frac{1}{2}\log\left[\frac{P(P+Q+N)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}\right]. \end{aligned}$$

$R(\alpha)$ is maximized at $\alpha^* = P/(P+N)$ and

$$R(\alpha^*) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right).$$

3 Marton's Region:

For any set of random variables $\{Z_1, Z_2, \cdots, Z_k\}$, let

$$\psi(Z_1, \cdots, Z_k) = \sum_{i=1}^{k-1} I(Z_1, \cdots, Z_i; Z_{i+1})$$

if $k \geq 2$ and let $\psi(Z_1, \cdots, Z_k) = 0$ if $k = 1$.

Theorem 3.1 (Marton's region). *The rate tuple $(R_1, \cdots, R_L)$ is achievable for discrete memoryless broadcast channel $p(y_1, \cdots, y_L|x)$ if there exist random variables $(W_1, \cdots, W_L)$ with $(W_1, \cdots, W_L) \to X \to (Y_1, \cdots, Y_L)$ such that*

$$\sum_{i\in\mathcal{A}} R_i \leq \sum_{i\in\mathcal{A}} I(W_i; Y_i) - \psi(W_i, i \in \mathcal{A})$$

for all $\emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L$.

Let

$$\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L) = \left\{(R_1, \cdots, R_L) \in \mathcal{R}_+^L : \sum_{i\in\mathcal{A}} R_i \leq \sum_{i\in\mathcal{A}} I(W_i; Y_i) - \psi(W_i, i \in \mathcal{A}),\ \forall \emptyset \subset \mathcal{A} \subseteq \mathcal{I}_L\right\}.$$

Theorem 3.2. *$\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$ is a polymatroid if $I(W_i; Y_i) - I(W_i; W_{\{i\}^c}) \geq 0$ for all i.*

If $\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$ is a polymatroid, then it has $L!$ vertices. Specifically, if π is a permutation on $\mathcal{I}_L$, define the vector $(R_1(\pi), \cdots, R_L(\pi))$ by

$$\begin{aligned} R_{\pi(1)}(\pi) &= I(W_{\pi(1)}; Y_{\pi(1)}), \\ R_{\pi(i)}(\pi) &= I(W_{\pi(i)}; Y_{\pi(i)}) - I(W_{\pi(1)}, \cdots, W_{\pi(i-1)}; W_{\pi(i)}) \end{aligned}$$

for $i = 2, \cdots, L$. Then $(R_1(\pi), \cdots, R_L(\pi))$ is a vertex of $\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$ for every permutation π. From the above expressions we can see that these vertices are achievable via L-step successive Gelfand-Pinsker coding.

Let

$$\mathcal{D}_{BC}(W_i, i \in \mathcal{I}_L) = \left\{ (R_1, \cdots, R_L) \in \mathcal{R}_{BC}(W_1, \cdots, W_L) : \sum_{i=1}^{L} R_i = \sum_{i=1}^{L} I(W_i; Y_i) - \psi(W_i, i \in \mathcal{I}_L) \right\}.$$

We call $\mathcal{D}_{BC}(W_i, i \in \mathcal{I}_L)$ the dominant face of $\mathcal{R}_{BC}(W_i, i \in \mathcal{I}_L)$. For $\emptyset \subset \mathcal{A} \subset \mathcal{I}_L$, define the hyperplane

$$\mathcal{H}'(\mathcal{A}) = \{(R_1, \cdots, R_L) : \sum_{i \in \mathcal{A}} R_i = \sum_{i \in \mathcal{A}} I(W_i; Y_i) - \psi(W_i, i \in \mathcal{A})\}$$

and let $\mathcal{F}'_{\mathcal{A}} = \mathcal{H}'(\mathcal{A}) \cap \mathcal{D}_{BC}(W_i, i \in \mathcal{I}_L)$. If $\emptyset \subset \mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{A}_k \subset \mathcal{I}_L$, is a telescopic sequence of subsets, then $\bigcap_{j=1}^{k} \mathcal{F}'_{\mathcal{A}_i}$ is a face of $\mathcal{D}_{BC}(W_i, i \in \mathcal{I}_L)$; conversely, every face of $\mathcal{D}_{BC}(W_i, i \in \mathcal{I}_L)$ can be written in this form. Let $\mathcal{B}_j = \mathcal{A}_j - \mathcal{A}_{j-1}$, $j = 1, 2, \cdots, k+1$, where we set $\mathcal{A}_0 = \emptyset$ and $\mathcal{A}_{k+1} = \mathcal{I}_L$. Any rate tuple $(R_1, \cdots, R_L) \in \bigcap_{j=1}^{k} \mathcal{F}'_{\mathcal{A}_i}$ can be achieved via a $k+1$-step successive group dirty paper coding scheme. Specifically, there are L bin arrays with 2^{nR_i} bins in array i, $i = 1, \cdots, L$. At step j, encoder searches for codewords $\mathbf{W}_i$ $(i \in \mathcal{B}_j)$ jointly in target bins (determined by messages for receivers $\mathcal{B}_j$) in arrays $\mathcal{B}_j$ such that $(\mathbf{W}_i, i \in \mathcal{B}_j)$ are jointly typical with $(\mathbf{W}_i, i \in \bigcup_{l=1}^{j-1} \mathcal{B}_l)$, $j = 1, \cdots, k+1$. Successive coding is implemented among different groups while joint coding is implemented within each group.

The splitting method can also be used here to achieve a general point on $\mathcal{D}(W_i, i \in \mathcal{I}_L)$ via successive Gelfand-Pinsker coding of at most $2L-1$ steps. More accurately, for any rate tuple $(R_1, \cdots, R_L) \in \bigcap_{i=1}^{k} \mathcal{F}'_{\mathcal{A}_i}$, the number of successive coding steps is $L + dim(\bigcap_{i=1}^{k} \mathcal{F}'_{\mathcal{A}_i}) \leq 2L - k - 1$.

For simplicity, we only discuss the 2-receiver case. In this case we have

$$\begin{aligned} R_i &\leq I(W_i; Y_i), \quad i = 1, 2, \\ R_1 + R_2 &= I(W_1; Y_1) + I(W_2; Y_2) - I(W_1; W_2). \end{aligned}$$

For any rate pair (R_1^*, R_2^*) we can find W_1' such that $W_1' \to W_1 \to (W_2, Y_1, Y_2)$ form a Markov chain and

$$\begin{aligned} R_1 &= I(W_1'; Y_1) + I(W_1; Y_1 | W_1') - I(W_1; W_2 | W_1'), \\ R_2 &= I(W_2; Y_2) - I(W_2; W_1'). \end{aligned}$$

Note: $I(W_1'; Y_1) + I(W_1; Y_1 | W_1') = I(W_1; Y_1)$, $I(W_1; W_2 | W_1') + I(W_2; W_1') = I(W_1; W_2)$.

Codebook Generation: First generate $2^{nI(W_1'; Y_1)}$ codewords $\mathbf{W}_1'$. For each $\mathbf{W}_1'$ generate $2^{n(I(W_1; Y_1 | W_1'))}$ codewords $\mathbf{W}_1$ according to $p(w_1 | w_1')$ and distribute them into $2^{n(I(W_1; Y_1 | W_1') - I(W_1; W_2 | W_1'))}$ bins. Generate $2^{nI(W_2; Y_2)}$ codewords $\mathbf{W}_2$ and distribute them into $2^{n(I(W_2; Y_2) - I(W_2; W_1'))}$ bins.

Encoding: First find $\mathbf{W}_1'$, then $\mathbf{W}_2$, and finally $\mathbf{W}_1$.

Decoding: Receiver 1 first decodes $\mathbf{W}_1'$ then decodes $\mathbf{W}_1$. Receiver 2 decodes $\mathbf{W}_2$.

4 Sato's Cooperative Bound:

An important observation is that the capacity region of a broadcast channel depends only on the conditional marginal distributions $p(y_i | x)$, $i = 1, \cdots, L$. This observation leads directly to the following upper-bound on the sum-rate,

$$\sum_{i=1}^{L} R_i \leq \min_{q(y_1, \cdots, y_L | x) \in \mathcal{Q}^*} I(X; Y_1, \cdots, Y_L),$$

where $\mathcal{Q}^*$ is the set of conditional distribution $q(y_1, \cdots, y_L|x)$ such that $q(y_i|x) = p(y_i|x)$ for all i.

Interpretation: Suppose that receivers can cooperate, then the broadcast channel problem reduces to a point to point communication problem. Clearly the maximum achievable rate of this point to point communication problem gives an upper bound on the sum-rate of the broadcast channel. Furthermore, we have the freedom to choose the conditional distribution as long as the conditional marginal distributions are kept fixed.

Interestingly, this upper bound is tight for the (scalar/vector) Gaussian broadcast channel.

5 Gaussian Broadcast Channels:

5.1 Scalar Case:

Channel model: $Y_i = X + Z_i$, where $Z_i \sim \mathcal{N}(0, N_i)$, $i = 1, \cdots, L$. Power constraint: $\mathbb{E}X^2 \leq P$. Without loss of generality, we can assume that $N_1 \leq N_2 \leq \cdots \leq N_L$. So we can write $Z_{i+1} = Z_i + \Delta Z_i$, where $\Delta Z_i \in \mathcal{N}(0, N_{i+1} - N_i)$, $i = 1, \cdots, L-1$. We can see that the signal at receiver $i+1$ is a degraded version of the signal at receiver i.

Consider the following coding scheme. For simplicity we only discuss the 2-receiver case. Generalization to the L-receiver case is straightforward. We can write

$$
\begin{aligned}
Y_1 &= X_1 + X_2 + Z_1, \\
Y_2 &= X_1 + X_2 + Z_2,
\end{aligned}
$$

where $X_1 \sim \mathcal{N}(0, \alpha P)$, $X_2 \sim \mathcal{N}(0, (1-\alpha)P)$, and $\alpha \in [0, 1]$. We can think that X_1 is the signal to receiver 1, and X_2 is the signal to receiver 2. At receiver 2, X_1 is viewed as the interference, and thus the following rate is achievable,

$$
R_2 \leq \frac{1}{2} \log \left(1 + \frac{(1-\alpha)P}{\alpha P + N_2} \right).
$$

For receiver 1, we use dirty paper coding, and the achievable rate is the same as the case where X_2 does not exist,

$$
R_1 \leq \frac{1}{2} \log \left(1 + \frac{\alpha P}{N_1} \right).
$$

In general, the ordering of dirty paper coding is determined by the ordering of noise variances.

Bergmans proved that this region is indeed the capacity region. Note: this capacity region can also be achieved via superposition coding[1]. Actually superposition coding is optimal for all degraded broadcast channels. But as we will see, dirty paper scheme can be easily applied in the non-degraded scenarios while superposition coding can not.

5.2 Vector Case:

We first generalized the dirty paper coding to the vector case. Let $\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}$, where $\mathbf{S} \sim \mathcal{N}(0, K_{\mathbf{S}})$, $\mathbf{Z} \sim \mathcal{N}(0, K_{\mathbf{Z}})$, and $\mathbf{X} \sim \mathcal{N}(0, K_{\mathbf{X}})$.

It can be shown that $I(\mathbf{U}; \mathbf{Y}) - I(\mathbf{U}; \mathbf{S})$ is maximized when $\mathbf{U} = \mathbf{X} + F\mathbf{S}$, and $\mathbf{X}$ is independent of $\mathbf{S}$. Here $F = K_{\mathbf{X}}(K_{\mathbf{X}} + K_{\mathbf{Z}})^{-1}$. It is curious that F is the Wiener filter to estimate

[1] Superposition coding can be viewed as a dual of successive refinement in source coding.

$\mathbf{X}$ from $\mathbf{X}+\mathbf{Z}$. It will be nice if someone can give an intuitive interpretation. Furthermore, when $\mathbf{U}=\mathbf{X}+F\mathbf{S}$, we have

$$I(\mathbf{U};\mathbf{Y})-I(\mathbf{U};\mathbf{S})=\frac{1}{2}\log\left(\frac{|K_{\mathbf{X}}+K_{\mathbf{Z}}|}{|K_{\mathbf{Z}}|}\right),$$

which is the maximum achievable rate when $\mathbf{S}$ does not exist. This fact can be explained by the following simple argument.

First consider the channel model without interference,

$$\mathbf{Y}=\mathbf{X}+\mathbf{Z}.$$

We can decompose this vector channel into paralleled scalar channels. Specifically, let $K_{\mathbf{Z}}=U_{\mathbf{Z}}^T\Sigma_{\mathbf{Z}}U_{\mathbf{Z}}$ be the eigenvalue decomposition of $K_{\mathbf{Z}}$. We can create the following equivalent channel

$$\mathbf{Y}'=\Sigma_{\mathbf{Z}}^{-1/2}U_{\mathbf{Z}}\mathbf{X}+\mathbf{Z}',$$

where $\mathbf{Y}'=\Sigma_{\mathbf{Z}}^{-1/2}U\mathbf{Y}$, and $\mathbf{Z}'=\Sigma_{\mathbf{Z}}^{-1/2}U\mathbf{Z}$. Let $K_{\mathbf{X}}=U_{\mathbf{X}}^T\Sigma_{\mathbf{X}}U_{\mathbf{X}}$ be the eigenvalue decomposition of $K_{\mathbf{X}}$. We have

$$\mathbf{Y}'=H\mathbf{X}'+\mathbf{Z}',$$

where $\mathbf{X}'=\Sigma_{\mathbf{X}}^{-1/2}U_{\mathbf{X}}\mathbf{X}$, and $H=\Sigma_{\mathbf{Z}}^{-1/2}U_{\mathbf{Z}}U_{\mathbf{X}}^T\Sigma_{\mathbf{X}}^{1/2}$. Let $H=V\Lambda W^T$ be the singular value decomposition of H. We can further define the following equivalent channel

$$\mathbf{Y}''=\Lambda\mathbf{X}''+\mathbf{Z}'',$$

where $\mathbf{Y}''=V^T\mathbf{Y}'$, $\mathbf{Z}''=V^T\mathbf{Z}'$, and $\mathbf{X}''=W^T\mathbf{X}'$. Now coding over sub-channels, we can achieve the maximum rate $\frac{1}{2}\log\left(\frac{|K_{\mathbf{X}}+K_{\mathbf{Z}}|}{|K_{\mathbf{Z}}|}\right)$. We can apply the exact same transform to the channel $\mathbf{Y}=\mathbf{X}+\mathbf{S}+\mathbf{Z}$. The interference $\mathbf{S}$, after transform, may be still correlated across the sub-channels. Such correlation is of no consequence to channel capacity because each sub-channel can be coded separately and the capacity of the scalar channel with known interference does not depend on the distribution of interference. Thus, the capacity of the memoryless vector channel is the same as if $\mathbf{S}$ is not present.

Now let's consider the vector Gaussian broadcast channel: $\mathbf{Y}_i=H_i\mathbf{X}+\mathbf{Z}_i$, $i=1,2,\cdots,L$. Power constraint: $\mathbb{E}\mathbf{X}\mathbf{X}^T\preccurlyeq K_{\mathbf{X}}$. In general, the vector Gaussian broadcast channel is non-degraded. Let $\mathbf{X}=\mathbf{X}_1+\cdots+\mathbf{X}_L$, where $\mathbb{E}\mathbf{X}\mathbf{X}^T=B_i$, $\sum_{i=1}^L B_i\preccurlyeq K_{\mathbf{X}}$, and $\mathbf{X}_1,\cdots,\mathbf{X}_L$ are independent. We have

$$\mathbf{Y}_i=H_i(\mathbf{X}_1+\cdots,\mathbf{X}_{i-1})+H_i\mathbf{X}_i+H_i(\mathbf{X}_{i+1}+\cdots+\mathbf{X}_L)+\mathbf{Z}_i,\quad i=1,\cdots,L.$$

By applying successive dirty paper coding in the order $L\to L-1\to\cdots\to 1$, we get the following achievable region

$$\begin{aligned}
R_1 &\le \frac{1}{2}\log\left(\frac{|H_1B_1H_1^T+K_{\mathbf{Z}_1}|}{|K_{\mathbf{Z}_1}|}\right),\\
R_i &\le \frac{1}{2}\log\left(\frac{\left|H_i\left(\sum\limits_{j=1}^{i}B_j\right)H_i^T+K_{\mathbf{Z}_i}\right|}{\left|H_i\left(\sum\limits_{j=1}^{i-1}B_j\right)H_i^T+K_{\mathbf{Z}_i}\right|}\right),\quad i=2,\cdots,L.
\end{aligned}$$

Since the vector Gaussian broadcast channel is non-degraded, we can choose arbitrary coding order, each corresponding to a permutation π on $\{1, \cdots, L\}$. Furthermore, the power allocation $\{B_1, \cdots, B_L\}$ can also be arbitrary as long as the constraint $\sum_{i=1}^{L} B_i \preccurlyeq K_{\mathbf{X}}$ is satisfied, and we can also use timesharing to convexify the resulting region. Remarkably, Weingarten *et al.* proved that this region is the capacity region.

References

[1] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. and Inform. Theory*, vol. 9(1), pp. 19-31, 1980.

[2] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1912-1921, Aug. 2003.

[3] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658-2668, Oct. 2003.

[4] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1875-1892, Sept. 2004.

[5] K. Marton, "A coding theorem for the discrete memoryless broadcast channels," *IEEE Trans. Inform. Th.*, vol. 25(3), pp. 306-311, May 1979.

[6] G. Caire and S. Shamai, "On the achievable throughput in multantenna Gaussian broadcast channel," *IEEE Trans. Infor. Theory*, vol. 49, pp. 1691-1706, July 2003.

[7] M. Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. IT-29, pp. 439-441, May 1983.

[8] Patrick P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise," *IEEE Trans. on Inform. Theory*, pp. 279-280, Mar. 1974.

[9] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2524-2530, Sept. 1998.

[10] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 24, pp. 374-377, May 1978.

[11] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. on Inform. Theory*, submitted for publication.

[12] T. M. Cover, "Broadcast channels," *IEEE Trans. on Inform. Theory*, vol. IT-18, pp. 2-14, Jan. 1972.

[13] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, 10(3):3-14, 1974.