Block Markov Encoding & Decoding

Deqiang Chen

I. Introduction

Various Markov encoding and decoding techniques are often proposed for specific channels, *e.g.*, the multi-access channel (MAC) with feedback, the relay channel, the MAC with cribbing encoders [1], [2], [3]. The goal of this report is to summarize the basic ideas of some block Markov coding techniques and present them using two simple channels, namely the point-to-point channel and the relay channel.

Transmission using block Markov coding operates over a number of blocks. In each block, with the exception of the first or the last block, a new message is sent. However, the codeword to send at each block depends on not only "fresh" information but also "past" information from one or more previous blocks. Hence the name "Markov encoding". The information from previous blocks can be refined information for the previous message or cooperative information for other users, *etc*. At the receiver, since the channel output at each block is related with messages from previous blocks, different decoding schemes have been proposed.

The rest of this summary is organized as follows. Section II discusses the application of Markov coding in the point-to-point, discrete memoryless channel (DMC). Block Markov coding is unnecessarily complicated for a simple point-to-point DMC, but studying this setting allows for a simplified treatment of the scheme's mechanics. To illustrate the benefits of block Markov coding in multi-user information theory, Section III applies block Markov coding techniques to prove an achievable rate for the relay channel. Section IV concludes the paper.

II. Markov Coding in the Point-to-Point Discrete Memoryless Channel

For the proof of the capacity theorem in the point-to-point DMC, block Markov coding techniques are not necessary. The benefit of these complicated techniques becomes clear when they are applied to more complicated channels, as we will see in Section III. Nevertheless, we use this application as a toy example to illustrate the basic ideas of block Markov coding.

Definition 1: A point-to-point DMC consists of an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$ and a probability transition function $\prod_{i=1}^{n} p(y_i|x_i)$ for all $x_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$, $i = 1, 2, ..., n$.

Definition 2: A $(\lceil 2^{nR} \rceil, n)$ code for a point-to-point DMC consists of the following:

- An index set $\mathcal{W} := \{1, 2, ..., \lceil 2^{nR} \rceil\}$.
- An encoding function $f_n : \mathcal{W} \to \mathcal{X}^n$.
- A decoding function $g_n : \mathcal{Y}^n \to \mathcal{W}$.

Definition 3: The average probability of error for a uniformly distributed message $\mathsf{w} \in \mathcal{W}$ is defined as the probability that the decoded message is not equal to the corresponding transmitted message [1],

$$P_e^{(n)} = P[g_n(\mathbf{y}) \neq \mathsf{w}]$$

Definition 4: A rate R is said to be achievable for a point-to-point DMC if there exits a sequence of $(\lceil 2^{nR} \rceil, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$.

Theorem 1: For a point-to-point DMC, a rate R is achievable if

$$R < \max_{p(x)} I(\mathsf{x}; \mathsf{y}).$$

Sketches of proofs of Theorem 1 using different schemes will be provided in the following sections. Throughout the whole summary, we consider B blocks, indexed by $1 \leq b \leq B$, each of n channel uses. A sequence of $B - 1$ messages $\mathsf{w}_b \in \mathcal{W}, b = 1, ..., B - 1$, will be transmitted using nB channel uses. For the convenience of presentation, we assume $\mathsf{w}_0 = \mathsf{w}_B = 1$ without loss of generality. The rate $R(B - 1)/B$ is arbitrarily close to R as $B \to \infty$.

A. Non-Markov coding

Recall that a standard non-Markov coding strategy is outlined as follows:

- Fix $p(x)$
- *Codebook generation*: Generate $\lceil 2^{nR} \rceil$ independent codewords of length n, $\mathbf{x}(w), w \in \mathcal{W}$, drawn according to $\prod_{i=1}^{n} p(x_i)$
- *Encoding*: In block b, send the codeword $\mathbf{x}(\mathsf{w}_b)$.

[1]Random variables are denoted using the sans serif font (*e.g.*, x) while random vectors and sequences are denoted with bold sans serif (*e.g.*, $\mathbf{x}$). A vector is assumed in column form unless otherwise stated. Instead of the more precise $p_{\mathsf{x}}(x)$, the probability density function (pdf) of the random variable x is denoted by $p(x)$.

- *Decoding*: At the end of block b, the decoder estimates the message of the transmitter $\hat{w}_b = w$ iff there exits a unique w such that $(\mathbf{x}(w), \mathbf{y}(b))$ is jointly typical.

Remarks:

- The above coding strategy is used in the now standard proof in [4].
- The codeword for block b only depends on the current message w_b. Hence, there is no Markov relationship between codewords of different blocks. Correspondingly, the receiver only utilizes the channel output in block b to decode the message w_b.

B. Irregular Encoding and Successive Decoding

Irregular encoding and successive decoding are first proposed for the relay channel in [1]. For a point-to-point DMC, the coding strategy is outlined as follows:

- Fix an auxiliary random variable u with distribution $p(u)$
- Fix $p(x|u)$
- *Codebook generation*: Generate $2^{nR_1}, R_1 \geq 0$, length n codewords, $\mathbf{u}(s), s \in \{1, 2, ..., \lceil 2^{nR_1} \rceil\}$, drawn according to $\prod_{i=1}^{n} p(u_i)$; for each $\mathbf{u}(s)$, generate $\lceil 2^{nR} \rceil$ independent codewords of length n, $\mathbf{x}(w|s), w \in \mathcal{W}$, drawn according to $\prod_{i=1}^{n} p(x_i|u_i(s))$.
- *Random partioning/binning*: Randomly throw each integer $w \in \mathcal{W}$ into one of 2^{nR_1} bins; denote the cell index of w as $s(w)$.
- *Encoding*: For block b, calculate $s_b = s(w_{b-1})$, send the codeword $\mathbf{x}(w_b|s_b)$.
- *Successive Decoding*: At the end of block b, assuming $\hat{s}_b$ is known, the decoder estimates the message $\hat{w}_b = w$ iff there exits a unique w such that $(\mathbf{u}(\hat{s}_b), \mathbf{x}(w|\hat{s}_b), \mathbf{y}(b))$ is jointly typical. Thereafter, $\hat{s}_{b+1} = s(\hat{w}_b)$ is available for decoding the next block.
- *Analysis of Error Probability*: It then can be shown that when $R < I(x; y|u)$, the error probability approaches zero as $n \to \infty$. Furthermore, considering the Markov chain $u \leftrightarrow x \leftrightarrow y$, we can show

$$\max_{p(u)p(x|u)} I(x; y|u) = \max_{p(x)} I(x; y).$$

To see this, we first notice that $p(y|x, u) = p(y|x)$, thus,

$$\begin{aligned} \max_{p(u)p(x|u)} I(x; y|u) &\geq \max_{p(x|u^*)} I(x; y|u^*) \\ &= \max_{p(x)} I(x; y), \end{aligned}$$

where the first inequality arises when we fix the distribution of u to be $P(u = u^*) = 1$. For the other direction, we can have

$$\begin{aligned} I(x;y|u) &= H(y|u) - H(y|x,u) \\ &= H(y|u) - H(y|x) \\ &\leq H(y) - H(y|x) \\ &= I(x;y), \end{aligned}$$

such that

$$\max_{p(u)p(x|u)} I(x;y|u) \leq \max_{p(x)} I(x;y).$$

Remarks:

- For a point-to-point DMC, another simple decoding strategy is to look for a unique w such that $(\mathbf{x}(w|\hat{s}_b), \mathbf{y}(b))$ is jointly typical given $\hat{s}_b = s(\hat{w}_{b-1})$.
- There is no constraint on R_1. Thus, we can choose $R_1 = 0$, from which we recover non-Markov coding.

C. Regular Encoding and Backwards Decoding

Regular encoding and backwards decoding are proposed for the MAC with generalized feedback in [3]. For a point-to-point DMC, the coding strategy is outlined as follows:

- Fix an auxiliary random variable u with distribution $p(u)$
- Fix $p(x|u)$
- *Codebook generation*: Generate $\lceil 2^{nR} \rceil$ length n codewords, $\mathbf{u}(s), s \in \mathcal{W}$, drawn according to $\prod_{i=1}^{n} p(u_i)$; for each $\mathbf{u}(s)$, generate $\lceil 2^{nR} \rceil$ independent codewords of length n, $\mathbf{x}(w,s), w \in \mathcal{W}$, drawn according to $\prod_{i=1}^{n} p(x_i|u_i(s))$.
- *Encoding*: For block b, send the codeword $\mathbf{x}(w_{b-1}, w_b)$.
- *Backwards Decoding*: The receiver starts decoding from the last block B and proceeds backwards. At block b, assuming $\hat{w}_b$ is known, the decoder estimates $\hat{w}_{b-1} = w$ iff there exits a unique w such that $(\mathbf{x}(\hat{w}_b, w), \mathbf{u}(w), \mathbf{y}(b))$ is jointly typical.
- *Analysis of Error Probability*: It then can be shown that when $R < I(x,u;y)$, the error probability approaches zero as $n \to \infty$. Furthermore, noticing the Markov chain $u \leftrightarrow x \leftrightarrow y$,

we can show that

$$I(x, u; y) \;=\; I(x; y).$$

Remarks:

- Compared to irregular encoding in Section II-B, this coding strategy does not require random binning. However, the number of the total codewords is $2^{nR} \times 2^{nR}$ compared to $2^{nR_1} \times 2^{nR}$ in irregular encoding. Recall that in Section II-B, there is no constraint on R_1. Thus, we can select $R_1 < R$ such that the irregular encoding requires less storage.
- Backwards decoding has a delay of B blocks. In contrast, successive decoding only imposes a delay of 2 blocks.

D. Regular Encoding and Sliding Window Decoding

In [2], another scheme is proposed that can decode regularly encoded information and reduce the delay. It utilizes the channel output in two blocks simultaneously for decoding. Hence the name "sliding window decoding". More specifically, at block b, assuming $\hat{w}_{b-1}$ is known from decoding of the previous block, the receiver estimates $\hat{w}_b = w$ iff there exists a unique w such that $(\mathbf{x}(w, \hat{w}_{b-1}), \mathbf{y}(b))$ is jointly typical. Recall that, in the successive decoding, the receiver estimates the "satellite" codeword given that the "cloud" center is decoded from the previous block. However, in the window decoding, the whole codeword is directly estimated without distinguishing between the "satellite" and the "cloud" center. In a point-to-point DMC, it is easy to see that both backwards decoding and window decoding yield the same rate in Theorem 1. In more general channels, sliding window decoding reduces the delay compared to backwards decoding, but can reduce the rate as well [5].

III. Markov coding for the relay channel

In the point-to-point DMC, Markov coding techniques are unnecessarily complicated and do not increase the rate. This section applies Markov coding techniques to the relay channel to illustrate the benefit of Markov coding in improving the rates of multi-user networks.

Definition 5: A relay channel consists of two input alphabets $\mathcal{X}_1, \mathcal{X}_2$, two output alphabets $\mathcal{Y}_2, \mathcal{Y}_3$ and a probability transition function $\prod_{i=1}^{n} p(y_{3,i}, y_{2,i}|x_{1,i}, x_{2,i})$ for all $x_{1,i} \in \mathcal{X}_1$, $x_{2,i} \in \mathcal{X}_2$, $y_{2,i} \in \mathcal{Y}_2$, $y_{3,i} \in \mathcal{Y}_3$ $i = 1, 2, ..., n$.

Definition 6: A $(\lceil 2^{nR} \rceil, n)$ code for a relay channel consists of the following:

- An index set $\mathcal{W} := \{1, 2, ..., \lceil 2^{nR} \rceil\}$.
- An encoding function $h_n : \mathcal{W} \to \mathcal{X}_1^n$.
- A set of relay functions $\{f_i\}_{i=1}^n$ such that

$$x_{2,i} = f_i(\mathsf{y}_{2,1}, \mathsf{y}_{2,2}, ..., \mathsf{y}_{2,i-1}) \quad , \quad 1 \leq i \leq n$$

- A decoding function $g_n : \mathcal{Y}_3^n \to \mathcal{W}$.

Definition 7: The average probability of error for a uniform message $\mathsf{w} \in \mathcal{W}$ is defined as the probability that the decoded message is not equal to the corresponding transmitted message, *i.e.*

$$P_e^{(n)} \quad = \quad P[g_n(\mathbf{y}_3) \neq \mathsf{w}]$$

Definition 8: A rate R is said to be achievable for a relay channel if there exits a sequence of $(\lceil 2^{nR} \rceil, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$.

The following theorem corresponds to [1, Theorem 1]. We will see that different Markov coding techniques can be used to prove this theorem. However, in general, different Markov coding techniques lead to different achievable rates. It is due to the special structure of the relay channel that the three Markov coding techniques lead to the same achievable rate.

Theorem 2: For a relay channel, a rate R is achievable if

$$R \quad < \quad \sup_{p(x_1, x_2)} \min\{I(\mathsf{x}_1; \mathsf{y}_2 | \mathsf{x}_2), I(\mathsf{x}_1, \mathsf{x}_2; \mathsf{y}_3)\}$$

A. Irregular Encoding and Successive Decoding

This section outlines the original proof of Theorem 2 using irregular encoding and successive decoding [1]:

- Fix $p(x_2)p(x_1|x_2)$
- *Codebook generation*: Generate 2^{nR_1} length n codewords, $\mathbf{x}_2(s), s \in \mathcal{W}$, drawn according to $\prod_{i=1}^n p(x_{2,i})$; for each $\mathbf{x}_2(s)$, generate $\lceil 2^{nR} \rceil$ independent codewords of length n, $\mathbf{x}_1(w|s), w \in \mathcal{W}$, drawn according to $\prod_{i=1}^n p(x_{1,i}|\mathsf{x}_{2,i}(s))$.
- *Random partioning/binning*: Randomly throw each integer $w \in \mathcal{W}$ into one of 2^{nR_1} bins; denote the cell index of w as $\mathsf{s}(w)$.
- *Encoding*:

- Node 1: Calculate $s_b = s(w_{b-1})$, send the codeword $\mathbf{x}_1(w_b|s_b)$.
- Node 2: Given $\hat{\hat{s}}_b = s(\hat{\hat{w}}_{b-1})$ is estimated from the previous block $b-1$, send the codeword $\mathbf{x}_2(\hat{\hat{s}}_b)$

- *Relay Decoding*: At the end of block $b-1$, given $\hat{\hat{w}}_{b-2}$ and $\hat{\hat{s}}_{b-1} = s(\hat{\hat{w}}_{b-2})$ are estimated from the previous block $b-1$, node 2 estimates $\hat{\hat{w}}_{b-1} = w$ iff there exists unique w such that $(\mathbf{x}_2(\hat{\hat{s}}_{b-1}), \mathbf{x}_1(w|\hat{\hat{s}}_{b-1}), \mathbf{y}_2(b-1))$ is jointly typical;
- *Successive Decoding*: At the end of block b, node 3
  - Estimates $\hat{s}_b = s$ iff there exists a unique s such that $(\mathbf{x}_2(s), \mathbf{y}_3(b))$ is jointly typical;
  - Estimates $\hat{w}_{b-1} = w$ iff there exists a unique w such that $(\mathbf{x}_1(w|\hat{s}_{b-1}), \mathbf{x}_2(\hat{s}_{b-1}), \mathbf{y}_3(b-1))$ is jointly typical and $\hat{s}_b = s(\hat{w}_{b-1})$.

Remarks:

- An intermediate result shows $R_1 < I(x_2; y_3)$.
- $R < I(x_1; y_2|x_2)$ is required to guarantee successful decoding at the relay, and $R < I(x_1, x_2; y_2)$ is required to guarantee successful decoding at the destination.

B. Regular Encoding and Backwards Decoding

This section uses regular encoding and backwards decoding to prove Theorem 2. It can be viewed as a simplified version of the proof in [3]. An outline of the proof is as follows:

- Fix $p(x_2)p(x_1|x_2)$
- *Codebook generation*: Generate $\lceil 2^{nR} \rceil$ length n codewords, $\mathbf{x}_2(s), s \in \mathcal{W}$, drawn according to $\prod_{i=1}^n p(x_{2,i})$; for each $\mathbf{x}_2(s)$, generate $\lceil 2^{nR} \rceil$ independent codewords of length n, $\mathbf{x}_1(w, s), w \in \mathcal{W}$, drawn according to $\prod_{i=1}^n p(x_{1,i}|x_{2,i}(s))$.
- *Encoding*:
  - Node 1: At block b, send the codeword $\mathbf{x}_1(w_{b-1}, w_b)$.
  - Node 2: Given $\hat{\hat{w}}_{b-1}$ is estimated from the previous block $b-1$, send the codeword $\mathbf{x}_2(\hat{\hat{w}}_{b-1})$
- *Relay Decoding*: At the end of block b, given $\hat{\hat{w}}_{b-1}$, node 2 estimates $\hat{\hat{w}}_b = w$ iff there exists unique w such that $(\mathbf{x}_2(\hat{\hat{w}}_{b-1}), \mathbf{x}_1(w, \hat{\hat{w}}_{b-1})), \mathbf{y}_2(b))$ is jointly typical;
- *Backwards Decoding*: The receiver starts decoding from the last block B and proceeds backwards. At block b, assume $\hat{w}_b$ is known following decoding of block $b+1$, the decoder

estimates $\hat{w}_{b-1} = w$ iff there exits a unique w such that $(\mathbf{x}_1(\hat{w}_b, w), \mathbf{x}_2(w), \mathbf{y}_3(b))$ is jointly typical.

C. Regular Encoding and Window Decoding

Again, we can improve on the decoding delay of backwards decoding by using the sliding window decoding [2]. The codebook generation and encoding procedure is the same as that of Section III-B. The decoding procedure is as follows:

- *Sliding Window Decoding*: At the end of block b, assume $\hat{w}_{b-2}$ is known, the decoder estimates $\hat{w}_{b-1} = w$ iff there exits a unique w such that $(\mathbf{x}_1(w, \hat{w}_{b-2}), \mathbf{x}_2(\hat{w}_{b-2}), \mathbf{y}_3(b-1))$ is jointly typical and $(\mathbf{x}_2(w), \mathbf{y}_3(b))$ is jointly typical.

IV. Conclusion

This report summarizes various Markov encoding and decoding techniques [1], [2], [3], focusing on the point-to-point DMC and the relay channel to illustrate the mechanics and an important application. In the point-to-point DMC and the relay channel, various Markov encoding and decoding techniques yield the same achievable rates. However, in general, in more complicated networks, different Markov encoding and decoding techniques yield different achievable rates and have various trade-offs in terms of delay and complexity.

References

[1] T. M. Cover and A. A. El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[2] A. B. Carleial, "Multiple-access channel with different generalized feedback signals," *IEEE Trans. Inform. Theory*, vol. 28, no. 6, pp. 841–850, Nov. 1982.

[3] F. M. J. Willems, "Informationtheoretical Results for the Discrete Memoryless Multiple Access Channel," Ph.D. dissertation, Doctor in de Wetenschappen Proefschrift, Katholieke Universiteit Leuven, Leuven, Belgium, 1982.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[5] J. N. Laneman and G. Kramer, "Window decoding for the multiaccess channel with generalized feedback," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2004.