ECE 729 Information Theory **Jan. 20, 2009**

Lecture 1

Lecturer: John A. Gubner **Scribe:** John Gubner

Lecture Outline

- Course Organization
- Course Introduction

CHAPTER 1
Introduction

1.1. Communication Systems

Communication is the conveying of a message (or its approximation) from one place, called the **source location**, to another place, called the **destination** [2]. In order to accomplish this task, we have available a **channel** that accepts inputs $\mathbf{x}$ at the source location, and produces a related output $\mathbf{y}$ at the destination; the output $\mathbf{y}$ de-

pends on the input $\mathbf{x}$, usually in some random way. The first difficulty we have is that the message, call it U, is usually not the same type of object as the acceptable inputs to the channel. For example, the message may be a baseband voice waveform, but the channel may accept only waveforms in a certain passband. Or the channel may be a fiber-optic cable that accepts only pulses of light. So we usually need some sort of converter to transform a message object into a channel-input object. Depending on the context, this converter may be called an **encoder**, a **modulator**, or a **transmitter**. The second difficulty we have is that the channel output $\mathbf{y}$ may not be the same type of object as the message U that is desired at the destination. For this reason, we need a deconverter that transforms the channel output $\mathbf{y}$ into an approximation of U, which we denote by $\widehat{U}$. Depending on the context, the deconverter may be called a **decoder**, a **demodulator**, or a **receiver**. The concatenation of the converter, the channel, and the deconverter constitutes a **communication system** as shown in Figure 1.1.

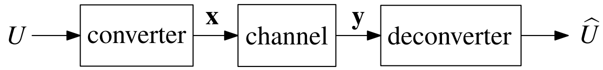

Figure 1.1. A communication system.

1.2. Digital Communication Systems

In a ***digital*** **communication system**, although there are many admissible values of the channel input $\mathbf{x}$, we intentionally use only a finite number of them, say N. This means that the converter in Figure 1.1 can be broken down into two subsystems.[a] The first one is a **source encoder** that takes a message and maps it into one of the integers $1,\ldots,N$. The second subsystem is a **channel encoder** that maps each of these integers into one of our N selected channel inputs, say $\mathbf{x}_1,\ldots,\mathbf{x}_N$. The deconverter can be similarly broken down into a **channel decoder** that takes channel outputs into the integers $1,\ldots,N$ and a **source decoder** that takes these integers into message values or their approximations.

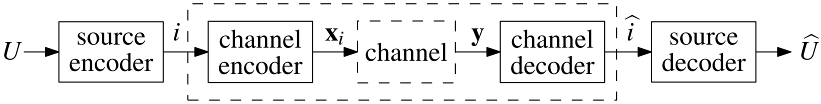

Figure 1.2. A digital communication system.

Another consequence of using only N channel inputs is that if the message U takes more than N distinct values, we cannot recover U in all cases. This situation can occur whether message is discrete or analog. Source encoders for binary data are discussed in the next subsection. As for analog data, for now we just mention that to approximate real-valued messages, we use **quantization**. If the message is a continuous-time waveform or a spatially continuous image, we sample it and then quantize the samples. Sampling is justified by the **sampling theorem** in the case of bandlimited data [2], [3]. As a practical matter, all sources are approximately bandlimited — a nice discussion of this can be found in the first three pages of Slepian [4] (easy reading — I recommend it).

1.2.1. Source Encoders for Binary Data

Consider an information source that generates an infinite stream of zeros and ones, which we denote by $U_1, U_2, \ldots$. Suppose we group the information symbols

[a] Just because we can break a system into two subsystems does not necessarily mean that it is a good idea to do so. We will say more about this shortly when we discuss the **Source-Channel Separation Theorem**.

into blocks of size k, e.g., $U := (U_1, \ldots, U_k)$. Then there are *at most* 2^k possible values of U. A **source encoder** is a function that maps each possible k-tuple of bits into an integer $i \in \{1, \ldots, N\}$, where $N \leq 2^k$. The **rate** of a such a source encoder is

$$R_s := \frac{\log_2 N}{k},$$

measured in *bits per source symbol.*

Why Might We Use $N < 2^k$?

Suppose that $U_{2i} = 0$ for $i = 1, 2, \ldots$. Then $(U_1, U_2, \ldots, U_k)$ can take only $2^{k/2}$ possible values. More generally, not all length-k blocks have equal probability. Some may be so unlikely that we can ignore them and still have a very low probability of error. **Shannon's Source Coding Theorem** says that this will be the case if R_s greater than or equal to the **entropy** H of the source.

1.2.2. Channel Codes

A **channel encoder** is a function that maps each $i \in \{1, \ldots, N\}$ into a vector $\mathbf{x}_i$ in a vector space X of dimension n. For example, we might have $\mathsf{X} = \{0, 1\}^n$; i.e., the set of n-tuples of bits. In any case, if a source block $U = (U_1, \ldots, U_k)$ is mapped into the integer i, then the signal $\mathbf{x}_i$ is transmitted over a channel, whose output $\mathbf{y}$ is determined by the transmitted signal $\mathbf{x}_i$ as well as noise. The **channel decoder** is a mapping that assigns to each possible channel output $\mathbf{y}$, an element in $\{1, \ldots, N\}$. The **rate** of the channel code is

$$R_c := \frac{\log_2 N}{n},$$

measured in *bits per dimension* or *bits per channel use*, depending on the space X. **Shannon's Channel Coding Theorem** says that under suitable conditions, if R_c is less than or equal to the **capacity** C of the channel, then $\mathsf{P}(i \neq \widehat{i}\,)$ can be made as small as we like by increasing n. In other words, the dashed box in Figure 1.2 can be made almost noiseless!

It is important to stress here that C does not depend on n. Since all we need is $(\log N)/n \leq C$, as n increases, we can let N increase as well!

1.2.3. Source Decoder

A **source decoder** is a mapping that assigns to each element in $\{1, \ldots, N\}$, a unique binary k-tuple.

1.2.4. Rate-Distortion Theory

When the dashed box in Figure 1.2 is almost noiseless, what remains is the **rate-distortion** problem. If N is not large enough to code each message separately, and this is always the case for analog data, how can we make $\widehat{U}$ the best approximation of U?

1.2.5. Source-Channel Separation

The conditions $R_s \geq H$ and $R_c \leq C$ together imply that

$$kH \leq \log_2 N \leq nC.$$

If source symbols arrive at rate ρ_s per second and channel symbols are sent at rate ρ_c symbols (or dimensions) per second, then during a time window of duration T, $k = \rho_s T$ and $n = \rho_c T$. Hence, the above condition implies $\rho_s TH \leq \rho_c TC$, or $\rho_s H \leq \rho_c C$. If source symbols arrive at the same rate as channel symbols (or dimensions) leave, then we need $H \leq C$.

Under suitable assumptions, there is *no* loss in designing the source encoder and the channel encoder separately. This is known as the **Source-Channel Separation Theorem** [1].

1.3. Course Outline

As suggested by the foregoing discussion, key topics in the course include the sampling theorem, quantization, fixed-length source codes and Shannon's Source Coding Theorem along with Shannon's Channel Coding Theorem. In addition, we will study the sampling theorem, quantization, variable-length source codes, and rate-distortion theory. Various other topics may be included as time permits.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006. Link to chapters from 1st ed.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July, Oct. 1948.

[3] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp, 10-21, Jan. 1949. Reprinted in *Proc. IEEE*, vol. 86, no. 2, pp. 447–457, Feb. 1998.

[4] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, no. 3, pp. 292–300, Mar. 1976.