to 0.5–0.75 bits/pixel results in good to very-good quality images that are sufficient for many applications. At 0.75–1.5 bits/pixel, excellent quality images are obtained sufficient for most applications. Finally, at rates of 1.5–2 bits/pixel, the resulting image is practically indistinguishable from the original. These rates are sufficient for the most demanding applications.

6.10 FURTHER READING

Any standard text on information theory covers source-coding theorems and algorithms in detail. Gallager (1968), Blahut (1987), and particularly Cover and Thomas (1991) provide nice and readable treatments of the subject. Our treatment of the Lempel-Ziv algorithm follows that of Cover and Thomas (1991). Berger (1971) is devoted entirely to rate-distortion theory. Jayant and Noll (1984) and Gersho and Gray (1992) examine various quantization and waveform-coding techniques in detail. Gersho and Gray (1992) includes detailed treatment of vector quantization. Analysis-synthesis techniques and linear-predictive coding are treated in books on speech coding such as Markel and Gray (1976), Rabiner and Schafer (1978), and Deller, Proakis, and Hansen (2000). The JPEG standard is described in detail in the book by Gibson, et al. (1998).

Among the original works contributing to the material covered in this chapter, we mention Shannon (1948a, 1959), Huffman (1952), Lloyd (1957), Max (1960), Ziv and Lempel (1978), and Linde, Buzo, and Gray (1980).

PROBLEMS

6.1 A source has an alphabet $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ with corresponding probabilities $\{0.1, 0.2, 0.3, 0.05, 0.15, 0.2\}$. Find the entropy of this source. Compare this entropy with the entropy of a uniformly distributed source with the same alphabet.

6.2 Let the random variable X be the output of the source that is uniformly distributed with size N. Find its entropy.

6.3 Show that $H(X) \geq 0$ with equality holding if and only if X is deterministic.

6.4 Let X be a geometrically distributed random variable; i.e.,

$$P(X = k) = p(1-p)^{k-1} \qquad k = 1, 2, 3, \ldots$$

1. Find the entropy of X.
2. Knowing that $X > K$, where K is a positive integer, what is the entropy of X?

6.5 Let $Y = g(X)$, where g denotes a deterministic function. Show that, in general, $H(Y) \leq H(X)$. When does equality hold?

6.6 An information source can be modeled as a bandlimited process with a bandwidth of 6000 Hz. This process is sampled at a rate higher than the Nyquist rate to provide a guard band of 2000 Hz. It is observed that the resulting samples take values in the set $\mathcal{A} = \{-4, -3, -1, 2, 4, 7\}$ with probabilities 0.2, 0.1, 0.15, 0.05, 0.3, 0.2.

What is the entropy of the discrete-time source in bits/output (sample)? What is the entropy in bits/sec?

6.7 Let X denote a random variable distributed on the set $\mathscr{A} = \{a_1, a_2, \ldots, a_N\}$ with corresponding probabilities $\{p_1, p_2, \ldots, p_N\}$. Let Y be another random variable defined on the same set but distributed uniformly. Show that

$$H(X) \leq H(Y)$$

with equality if and only if X is also uniformly distributed. (Hint: First prove the inequality $\ln x \leq x - 1$ with equality for $x = 1$, then apply this inequality to $\sum_{n=1}^{N} p_n \ln(\frac{\frac{1}{N}}{p_n})$).

6.8 A random variable X is distributed on the set of all positive integers $1, 2, 3, \ldots$ with corresponding probabilities $p_1, p_2, p_3, \ldots$. We further know that the expected value of this random variable is given to be m; i.e.,

$$\sum_{i=1}^{\infty} i p_i = m$$

Show that among all random variables that satisfy the above condition, the geometric random variable which is defined by

$$p_i = \frac{1}{m}\left(1 - \frac{1}{m}\right)^{i-1} \qquad i = 1, 2, 3, \ldots$$

has the highest entropy. (Hint: Define two distributions on the source, the first one being the geometric distribution given above and the second one an arbitrary distribution denoted by q_i, and then apply the approach of Problem 6.7.)

6.9 Two binary random variables X and Y are distributed according to the joint distribution $p(X = Y = 0) = p(X = 0, Y = 1) = p(X = Y = 1) = \frac{1}{3}$. Compute $H(X)$, $H(Y)$, $H(X \mid Y)$, $H(Y \mid X)$, and $H(X, Y)$.

6.10 Show that if $Y = g(X)$ where g denotes a deterministic function, then $H(Y \mid X) = 0$.

6.11 A memoryless source has the alphabet $\mathscr{A} = \{-5, -3, -1, 0, 1, 3, 5\}$ with corresponding probabilities $\{0.05, 0.1, 0.1, 0.15, 0.05, 0.25, 0.3\}$.

1. Find the entropy of the source.
2. Assume that the source is quantized according to the quantization rule

$$\begin{cases} q(-5) = g(-3) = -4, \\ q(-1) = q(0) = q(1) = 0 \\ q(3) = q(5) = 4 \end{cases}$$

Find the entropy of the quantized source.

6.12 Using both definitions of the entropy rate of a process, prove that for a DMS the entropy rate and the entropy are equal.

3. What is the probability that in this sequence the first k symbols are ones and the next $n - k$ symbols are zeros?
4. What is the probability that this sequence has k ones and $n - k$ zeros?
5. How would your answers change if instead of a BSS we were dealing with a general binary DMS with $p(X_i = 1) = p$.

6.20 Give an estimate of the number of binary sequences of length 10,000 with 3000 zeros and 7000 ones.

6.21 A memoryless ternary source with output alphabet a_1, a_2, and a_3 and corresponding probabilities 0.2, 0.3, 0.5 produces sequences of length 1000.

1. Approximately what is the number of typical sequences in the source output?
2. What is the ratio of typical sequences to nontypical sequences?
3. What is the probability of a typical sequence?
4. What is the number of bits required to represent all output sequences?
5. What is the number of bits required to represent only the typical output sequences?
6. What is the most probable sequence and what is its probability?
7. Is the most probable sequence a typical sequence?

6.22 A source has an alphabet $\{a_1, a_2, a_3, a_4\}$ with corresponding probabilities $\{0.1, 0.2, 0.3, 0.4\}$.

1. Find the entropy of the source.
2. What is the minimum required average code word length to represent this source for error-free reconstruction?
3. Design a Huffman code for the source and compare the average length of the Huffman code with the entropy of the source.
4. Design a Huffman code for the second extension of the source (take two letters at a time). What is the average code word length? What is the average required binary letters per each source output letter?
5. Which one is a more efficient coding scheme, Huffman coding of the original source or Huffman coding of the second extension of the source?

6.23 Design a Huffman code for a source with n output letters and corresponding probabilities $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^{n-1}}, \frac{1}{2^{n-1}}\}$. Show that the average code word length for such a source is equal to the source entropy.

6.24 Show that $\{01, 100, 101, 1110, 1111, 0011, 0001\}$ cannot be a Huffman code for *any* source probability distribution.

6.25 Design a *ternary* Huffman code, using 0, 1, 2 as letters, for a source with output alphabet probabilities given by $\{0.05, 0.1, 0.15, 0.17, 0.18, 0.22, 0.13\}$. What is the resulting average code word length? Compare the average code word length with the entropy of the source. (In what base would you compute the logarithms in the expression for the entropy for a meaningful comparison?)

6.26 Design a ternary Huffman code for a source with output alphabet probabilities given by $\{0.05, 0.1, 0.15, 0.17, 0.13, 0.4\}$. (Hint: You can add a dummy source output with zero probability.)

6.27 Find the Lempel-Ziv source code for the binary source sequence

$$000100100000011000010000000100000001010000010000000110100000000110$$

Recover the original sequence back from the Lempel-Ziv source code. (Hint: You require two passes of the binary sequence to decide on the size of the dictionary.)

6.28 Using the definition of $H(X)$ and $H(X \mid Y)$ show that

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Now by using the approach of Problem 6.7 show that $I(X;Y) \geq 0$ with equality if and only if X and Y are independent.

6.29 Show that

1. $I(X;Y) \leq \min\{H(X), H(Y)\}$.
2. If $|\mathscr{X}|$ and $|\mathscr{Y}|$ represent the size of sets $\mathscr{X}$ and $\mathscr{Y}$, respectively, then $I(X;Y) \leq \min\{\log|\mathscr{X}|, \log|\mathscr{Y}|\}$.

6.30 Show that $I(X;Y) = H(X)+H(Y)-H(X,Y) = H(Y)-H(Y \mid X) = I(Y;X)$.

6.31 Let X denote a binary random variable with $p(X=0) = 1 - p(X=1) = p$ and let Y be a binary random variable that depends on X through $p(Y=1 \mid X=0) = p(Y=0 \mid X=1) = \epsilon$.

1. Find $H(X)$, $H(Y)$, $H(Y \mid X)$, $H(X,Y)$, $H(X \mid Y)$, and $I(X;Y)$.
2. For a fixed ϵ, which p maximizes $I(X;Y)$?
3. For a fixed p, which ϵ minimizes $I(X;Y)$?

6.32 Show that

$$I(X;YZW) = I(X;Y) + I(X:Z \mid Y) + I(X;W \mid ZY)$$

Can you interpret this relation?

6.33 Let X, Y, and Z be three discrete random variables.

1. Show that if $p(x,y,z) = p(z)p(x \mid z)p(y \mid x)$, we have
$$I(X;Y \mid Z) \leq I(X;Y)$$
2. Show that if $p(x,y,z) = p(x)p(y)p(z \mid x,y)$, then
$$I(X;Y) \leq I(X;Y \mid Z)$$
3. In each case give an example where strict inequality holds.

1. What is $D_{\max}$ in the compression of the source?
2. Find the rate-distortion function for the source.
3. If we want to reproduce $X(t)$ with a distortion equal to 10, what transmission rate is required?
4. Find the channel capacity-cost function, where cost is assumed to be the power. What is the required power such that the source can be transmitted via the channel with a distortion not exceeding 10?

9.20 It can be shown that the capacity of a discrete-time power constrained additive noise channel described by $Y = X + Z$, where X and Y are the input and the output and Z is the noise, satisfies the inequalities

$$\frac{1}{2}\log\left(1+\frac{P}{N}\right) \leq C \leq \frac{1}{2}\log[2\pi e(P+N)] - h(Z)$$

where P is the input power constraint, N is the variance (power) of the noise process, and $h(.)$ denotes the differential entropy as defined in Chapter 6. Using the result of Problem 6.36, plot the lower and upper bounds to the capacity for a channel with Laplacian noise (see Problem 6.36) as a function of the noise variance (noise power).

9.21 Plot the capacity of an AWGN channel that employs binary antipodal signaling, with optimal bit-by-bit detection at the receiver, as a function of $\frac{\mathcal{E}_b}{N_0}$. On the same axis, plot the capacity of the same channel when binary orthogonal signaling is employed.

9.22 In Example 9.5.1, find the minimum distance of the code. Which code word(s) is(are) minimum weight?

9.23 In Example 9.5.3, verify that all code words of the original code satisfy

$$\mathbf{c}\mathbf{H}^t = \mathbf{0}$$

9.24 By listing all code words of the $(7, 4)$ Hamming code verify that its minimum distance is equal to 3.

9.25 Find the parity check matrix and the generator matrix of a $(15, 11)$ Hamming code in the systematic form.

9.26 Show that the minimum Hamming distance of a linear block code is equal to the minimum number of columns of its parity check matrix that are linearly dependent. From this conclude that the minimum Hamming distance of a Hamming code is always equal to 3.

9.27 A simple repetition code of blocklength n is a simple code consisting of only two code words one $\underbrace{(0, 0, \ldots, 0)}_{n}$ and the other $\underbrace{(1, 1, \ldots, 1)}_{n}$. Find the parity check matrix and the generator matrix of this code in the systematic form.

9.28 $\mathbf{G}$ is the generator matrix of a $(6, 3)$ linear code. This code is *extended* by adding an overall parity check bit to each code word so that the Hamming weight of each resulting code word is even.

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

1. Find the parity check matrix of the extended code.
2. What is the minimum distance of the extended code?
3. Find the coding gain of the extended code?

9.29 Compare the block error probability of an uncoded system with a system that uses a (15, 11) Hamming code. The transmission rate is $R = 10^4$ bps and the channel is AWGN with a received power of 1 μW and noise power-spectral density of $\frac{N_0}{2}$. The modulation scheme is binary PSK and soft-decision decoding is employed. Answer the question when hard decision is employed.

9.30 Generate the standard array for a (7, 4) Hamming code and use it to decode the received sequence (1, 1, 1, 0, 1, 0, 0).

9.31 For what values of k does an (n, k) cyclic code exist? List all possible k's with corresponding generator polynomial(s).

9.32 Find a generator polynomial and the corresponding code words for a (7, 3) cyclic code. What is the minimum distance of this code?

9.33 Design an encoder for a (15, 11) cyclic code.

9.34 Using the generator polynomial $g(p) = 1 + p + p^4$, find the generator matrix and the parity check matrix (in systematic form) of a (15, 11) cyclic code.

9.35 Let $g(p) = p^8 + p^6 + p^4 + p^2 + 1$ denote a polynomial over the binary field.

1. Find the lowest rate cyclic code whose generator polynomial is $g(p)$; what is the rate of this code?
2. Find the minimum distance of the code found in part 1.
3. What is the coding gain for the code found in part 1?

9.36 The polynomial $g(p) = p + 1$ over the binary field is considered.

1. Show that this polynomial can generate a cyclic code for any choice of n. Find the corresponding k.
2. Find the systematic form of **G** and **H** for the code generated by $g(p)$.
3. Can you say which type of code this generator polynomial generate?

9.37 Design a (6, 2) cyclic code by choosing the shortest possible generator polynomial.

1. Determine the generator matrix **G** (in the systematic form) for this code and find all possible code words.
2. How many errors can be corrected by this code?
3. If this code is used in conjunction with binary PSK over an AWGN channel with $P = 1$ W, $N_0 = 2 \times 10^{-6}$ W/Hz, and $W = 6 \times 10^4$ Hz and the

5. Show that

$$\underbrace{\int\cdots\int}_{R^N}\sqrt{p(\mathbf{r}\mid\mathbf{x}_m)p(\mathbf{r}\mid\mathbf{x}_{m'})}\,d\mathbf{r}=e^{-\frac{\|\mathbf{x}_m-\mathbf{x}_{m'}\|^2}{4N_0}}$$

and, therefore,

$$P(\text{error}\mid x_m(t)\text{ sent})\le\sum_{\substack{1\le m'\le M\\ m'\ne m}}e^{-\frac{\|\mathbf{x}_m-\mathbf{x}_{m'}\|^2}{4N_0}}$$

9.41 A convolutional code is described by

$$\mathbf{g}_1=[1\quad 0\quad 0]$$
$$\mathbf{g}_2=[1\quad 0\quad 1]$$
$$\mathbf{g}_3=[1\quad 1\quad 1]$$

1. Draw the encoder corresponding to this code.
2. Draw the state-transition diagram for this code.
3. Draw the trellis diagram for this code.
4. Find the transfer function and the free distance of this code.
5. Verify whether this code is catastrophic or not.

9.42 Show that in the trellis diagram of a convolutional code, 2^k branches enter each state and 2^k branches leave each state.

9.43 The block diagram of a binary convolutional code is shown in Figure P-9.43.

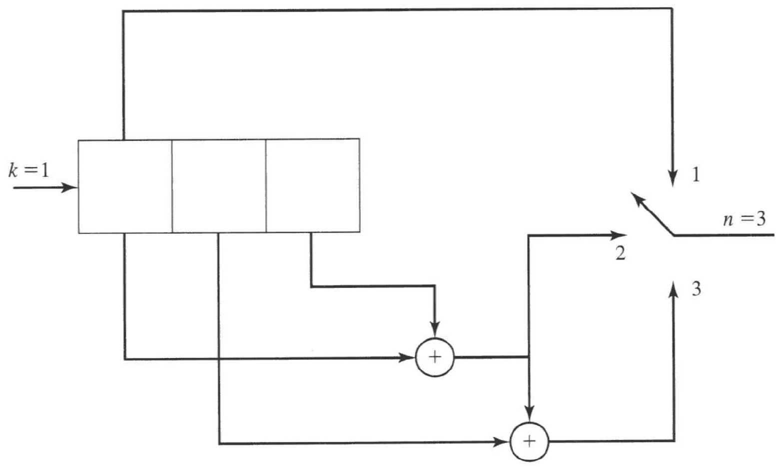

Figure P-9.43

1. Draw the state diagram for the code.
2. Find $T(D)$, the transfer function of the code.
3. What is d_{free}, the minimum free distance of the code?
4. Assume that a message has been encoded by this code and transmitted over a binary-symmetric channel with an error probability of $p = 10^{-5}$. If the received sequence is $\mathbf{r} = (110, 110, 110, 111, 010, 101, 101)$, using the Viterbi algorithm find the transmitted bit sequence.
5. Find an upper bound to bit-error probability of the code when the above binary-symmetric channel is employed. Make any reasonable approximations.

9.44 The block diagram of a (3, 1) convolutional code is shown in Figure P-9.44.

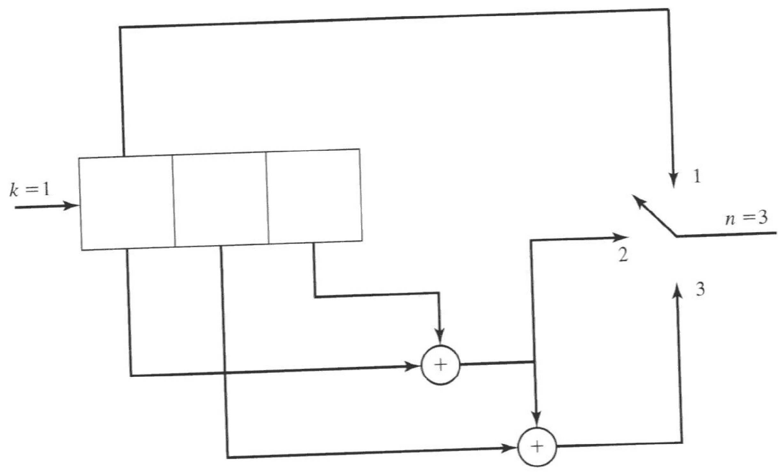

Figure P-9.44

1. Draw the state diagram of the code.
2. Find the transfer function $T(D)$ of the code.
3. Find the minimum free distance (d_{free}) of the code and show the corresponding path (at distance d_{free} from the all-zero code word) on the trellis.
4. Assume that four information bits, (x_1, x_2, x_3, x_4), followed by two zero bits, have been encoded and sent via a binary-symmetric channel with crossover probability equal to 0.1. The received sequence is (111, 111, 111, 111, 111, 111). Use the Viterbi decoding algorithm to find the most likely data sequence.

9.45 The convolutional code of Problem 9.41 is used for transmission over a AWGN channel with hard-decision decoding. The output of the demodulator-detector is (101001011110111...). Using the Viterbi algorithm, find the transmitted sequence.

9.46 Repeat Problem 9.41 for a code with

$$\mathbf{g}_1 = [1 \quad 1 \quad 0]$$
$$\mathbf{g}_2 = [1 \quad 0 \quad 1]$$
$$\mathbf{g}_3 = [1 \quad 1 \quad 1]$$

9.47 Show the paths corresponding to all code words of weight 6 in Example 9.7.3.

9.48 In the convolutional code generated by the encoder shown in Figure P-9.48.

1. Find the transfer function of the code in the form $T(N, D)$.
2. Find d_{free} of the code.
3. If the code is used on a channel using hard-decision Viterbi decoding, assuming the crossover probability of the channel is $p = 10^{-6}$, use the hard-decision bound to find an upper bound on the average bit-error probability of the code.

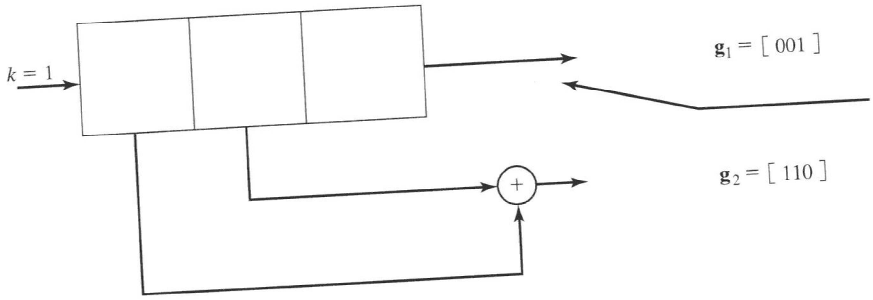

Figure P-9.48

9.49 Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two code words of length n with distance d, and assume that these two code words are transmitted via a binary-symmetric channel with crossover probability p. Let $P(d)$ denote the error probability in transmission of these two code words.

1. Show that

$$P(d) \leq \sum_{i=1}^{2^n} \sqrt{p(\mathbf{y}_i \mid \mathbf{x}_1) p(\mathbf{y}_i \mid \mathbf{x}_2)}$$

where the summation is over all binary sequences $\mathbf{y}_i$.