

- [12] —, "Cyclic Hadamard difference sets — Constructions and applications," in *Sequences and Their Applications: Proc. SETA '98*, C. Ding, T. Helleseeth, and H. Niederreiter, Eds. London, U.K.: Springer Verlag, 1999, pp. 39–48.
- [13] T. Helleseeth and P. V. Kumar, "Sequences with low correlation," in *Handbook of Coding Theory*, Pless and Huffman, Eds. Amsterdam, The Netherlands: Elsevier, 1998, vol. II, pp. 1765–1854.
- [14] D. Jungnickel and A. Pott, "Difference sets: Abelian," in *The CRC Handbook of Combinatorial Designs*, C. J. Colbourn and J. H. Dinitz, Eds. Boca Raton, FL: CRC, 1996, pp. 297–307.
- [15] J.-H. Kim and H.-Y. Song, "Existence of cyclic Hadamard difference sets and its relation to binary sequences with ideal autocorrelation," *J. Commun. and Networks*, vol. 1, no. 1, pp. 14–18, 1999.
- [16] A. Lempel, M. Cohn, and W. L. Eastman, "A class of binary sequences with optimal autocorrelation properties," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 38–42, Jan. 1977.
- [17] J.-S. No, H. Chung, H.-Y. Song, K. Yang, J.-D. Lee, and T. Helleseeth, "New construction for binary sequences of period $p^m - 1$ with optimal autocorrelation using $(z + 1)^d + az^d + b$," paper, submitted for publication.
- [18] J.-S. No, H. Chung, and M.-S. Yun, "Binary pseudorandom sequences of period $2^m - 1$ with ideal autocorrelation generated by the polynomial $z^d + (z + 1)^d$," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1278–1282, May 1998.
- [19] J.-S. No, S. W. Golomb, G. Gong, H.-K. Lee, and P. Gaal, "Binary pseudorandom sequences of period $2^n - 1$ with ideal autocorrelation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 814–817, Mar. 1998.
- [20] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications*. Rockville, MD: Computer Sci., 1985, vol. 1.
- [21] —, *Spread Spectrum Communications*, revised ed. New York: McGraw-Hill, 1994.

On Codes that Avoid Specified Differences

Bruce E. Moision, *Member, IEEE*,

Alon Orlitsky, *Senior Member, IEEE*, and Paul H. Siegel, *Fellow, IEEE*

Abstract—Certain magnetic recording applications call for a large number of sequences whose differences do not include certain disallowed binary patterns. We show that the number of such sequences increases exponentially with their length and that the growth rate, or capacity, is the logarithm of the joint spectral radius of an appropriately defined set of matrices. We derive a new algorithm for determining the joint spectral radius of sets of nonnegative matrices and combine it with existing algorithms to determine the capacity of several sets of disallowed differences that arise in practice.

Index Terms—Capacity, constrained coding, joint spectral radius, magnetic recording.

I. INTRODUCTION

The error probability of many magnetic-recording systems may be characterized in terms of the differences between the sequences that may be recorded [1]–[3]. In fact, the bit-error rate (BER) is often dominated by a small set of potential difference patterns. Recently, binary codes have been proposed which exploit this fact [4]–[9]. The codes are designed to avoid the most problematic difference patterns by constraining the set of allowed recorded sequences and have been shown to improve system performance.

In this correspondence, we study the largest number of sequences whose differences exclude a given set of disallowed patterns. We show that the number of such sequences increases exponentially with their length and that the growth rate, or capacity, is the logarithm of the joint spectral radius of an appropriately defined set of matrices. We derive new algorithms for determining the joint spectral radius of sets of nonnegative matrices and combine them with existing algorithms to determine the capacity of several sets of disallowed differences that arise in practice.

The correspondence is organized as follows. In the next section, we motivate the problem by summarizing known results showing that the error probability in models of magnetic recording systems is determined by the differences between recorded sequences. In Section III, we formally describe the resulting combinatorial problem, introduce the notation used, and present some simple examples. Section IV contains the main result of the correspondence, deriving the connection to

Manuscript received October 4, 1999; revised August 17, 2000. This work was supported by UC Micro Project 98-140 in cooperation with Marvell Semiconductor Incorporated, the Center for Magnetic Recording Research (UCSD), and under NSF Grant NCR-9612802. The material in this correspondence was presented in part at the 1999 IEEE Global Communications Conference (GLOBECOM'99), Rio de Janeiro, Brazil, December 5–9, 1999 and at the 2000 IEEE International Symposium on Information Theory (ISIT'00), Sorrento, Italy, June 25–30, 2000.

B. E. Moision was with the Mathematical Sciences Research Center, Lucent Technologies, Murray Hill, NJ 07974 USA. He is now with the Communications Systems and Research Section, Jet Propulsion Laboratory, Pasadena, CA 91101-8099 USA (e-mail: bmoision@shannon.jpl.nasa.gov).

A. Orlitsky and P. H. Siegel are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: alon@ece.ucsd.edu; psiegel@ucsd.edu).

Communicated by R. M. Roth, Associate Editor for Coding Theory.
 Publisher Item Identifier S 0018-9448(01)00374-1.

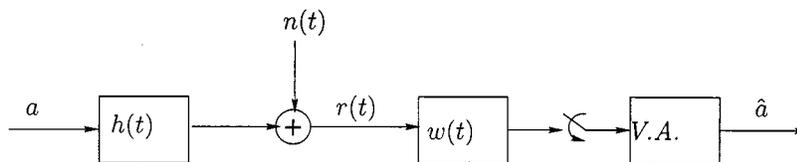


Fig. 1. Communications channel model.

the joint spectral radius. In Section V, we describe some known algorithms for determining the spectral radius, and derive a new algorithm. Finally, in Section VI, we determine the capacities of some simple sets of disallowed patterns.

II. MOTIVATION

Consider the binary communications channel in Fig. 1 where a sequence $a = (\dots, a_0, a_1, a_2, \dots)$ over the binary alphabet $\{+1, -1\}$ passes through a linear channel with impulse response $h(t)$ and $n(t)$ is additive white Gaussian noise. The received signal is given by

$$r(t) = \sum_k a_k h(t-k) + n(t).$$

Forney [10] showed that for $h(t)$ of finite duration, the receiver illustrated in Fig. 1 consisting of a whitened matched filter $w(t)$, a bit-rate sampler, and a minimum Euclidean distance estimate given by the Viterbi algorithm yields a maximum-likelihood (ML) estimate for the transmitted sequence a .

However, the complexity of implementing the Viterbi algorithm for the ML estimate grows exponentially with the bit-period duration of the channel pulse response. Hence, suboptimal partial-response schemes are often implemented in practice. For example, in hard-disk-drive magnetic recording channels, partial-response equalization and detection using the Viterbi algorithm is currently the accepted mode of operation, e.g., [11].

In a partial-response scheme, the receiving filter $w(t)$ is chosen such that the signal at the input to the Viterbi detector in the absence of noise approximates Xa , where X is the Toeplitz matrix corresponding to a finite target response x . The Viterbi algorithm is then used to obtain the sequence $\hat{X}\hat{a}$ closest in Euclidean distance to the sequence received at the input to the Viterbi detector.

Let the *difference sequence* $e = (\hat{a} - a)/2$ denote the difference between the decoded and transmitted sequences. The *effective distance* between a and \hat{a} is

$$d_{\text{eff}}(e) \stackrel{\text{def}}{=} \frac{(Xe)^T(Xe)}{((Xe)^T R(Xe))^{1/2}}$$

where R is the autocorrelation matrix of the noise at the input to the Viterbi detector and the superscript T denotes vector transposition.

At high signal-to-noise ratios (SNRs), the probability of a bit error for this detector is well approximated, e.g., [12], by

$$\Pr(\text{bit error}) \approx \sum_e P(e) \text{wgt}(e) Q\left(\frac{d_{\text{eff}}(e)}{2}\right) \quad (1)$$

where $\text{wgt}(e)$ is the number of ± 1 s in e , and

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$$

is the *error function*.

The sum in (1) is largely determined by a small set of dominant difference sequences—those with small effective distance. Letting D be the set of these difference sequences we can approximate the error probability using only a few terms

$$\Pr(\text{bit error}) \approx \sum_{e \in D} P(e) \text{wgt}(e) Q\left(\frac{d_{\text{eff}}(e)}{2}\right).$$

The magnetic recording channel may be modeled with the *Lorentzian* pulse response

$$h(t) = \frac{1}{1 + (2t/\tau)^2} - \frac{1}{1 + (2(t-1)/\tau)^2}$$

where the parameter τ measures the *density* of the recording. It was shown in [1] that for $x = (1, 1, -1, -1)$ and $1.5 \leq \tau \leq 2.75$, a target response and a density range of practical interest, the minimum effective distance difference pattern is contained in the set

$$D = \{000(+0)^k 00, 000 + (-+)^k 000\}$$

where $(\cdot)^k$ denotes $k \geq 1$ repetitions of (\cdot) , and we use a shorthand notation to represent the ternary difference patterns, i.e., $+0-$ is used to denote $+1, 0, -1$. The bit-error probability is well approximated over this range of densities by taking $k = 1$, i.e.,

$$D = \{000 + 000, 000 + - + 000\}.$$

The fact that a small set of difference patterns dominate the system performance has motivated the construction of codes designed to avoid the occurrence of the low-distance difference patterns [4]. The subsequent increase in the minimum effective distance is, however, offset by a loss in rate from the code. Further research [5]–[7], [9], [8] has investigated higher rate codes designed to avoid the difference pattern $000 + - + 000$, which is the minimum effective distance pattern over certain density ranges. This leads to the following question, which we address in this correspondence: what is the highest rate of a code that avoids a specified set of difference patterns?

III. NOTATION AND DEFINITIONS

The *difference* between two n -bit sequences $u = (u_1, \dots, u_n)$ and $v = (v_1, \dots, v_n)$ is the sequence

$$u - v \stackrel{\text{def}}{=} (u_1 - v_1, \dots, u_n - v_n) \in \{-1, 0, 1\}^n$$

where subtraction is over the reals.

Given a set D of finite-length disallowed difference patterns and a sequence length, n , we are interested in the largest number of n -bit sequences whose differences do not include any element of D .

An n -bit *code* \mathcal{C} is a collection of n -bit sequences, or *codewords*, thought of as potential recorded sequences. \mathcal{C} *avoids* D if, for all $u, v \in \mathcal{C}$ and all $i \leq j$ in $[1, n]$

$$u_{[i,j]} - v_{[i,j]} \notin D \quad (2)$$

where, for all $i \leq j$, we use the notation

$$[i, j] \stackrel{\text{def}}{=} \{i, \dots, j\}$$

and

$$u_{[i,j]} \stackrel{\text{def}}{=} u_i, \dots, u_j.$$

The largest number of sequences whose differences do not include any pattern in D is therefore

$$\delta_n(D) \stackrel{\text{def}}{=} \max\{|\mathcal{C}| : \mathcal{C} \text{ avoids } D\}.$$

It is easy to verify that $\delta_n(D)$ is *submultiplicative*

$$\delta_{n_1+n_2}(D) \leq \delta_{n_1}(D) \cdot \delta_{n_2}(D)$$

for all $n_1, n_2 > 0$. Hence, by the Sub-Additivity Lemma, e.g., [13], we can define the *capacity* of D as the limit

$$\text{cap}(D) \stackrel{\text{def}}{=} \log \left[\lim_{n \rightarrow \infty} (\delta_n(D))^{1/n} \right]. \quad (3)$$

We would like to determine the capacities of various difference sets D and find codes that achieve them.

We are primarily interested in finite difference sets. Without loss of generality we therefore assume from here on that all patterns in D have the same length m . Otherwise, let m be the length of the longest pattern in D and replace every pattern of length $m' < m$ by its $3^{m-m'}$ extensions of length m .

With this equal-length assumption, we restate constraint (2) and require that for all $u, v \in \mathcal{C}$ and all $i \in [1, n']$

$$u_{[i,i']} - v_{[i,i']} \notin D$$

where, for i and n only, we let

$$i' \stackrel{\text{def}}{=} i + m - 1$$

and

$$n' \stackrel{\text{def}}{=} n - m + 1.$$

Note also that we use the term *pattern* to refer to strings of length m and *sequence* for strings of length n .

The following examples illustrate these concepts for two simple disallowed difference sets. An n -bit code \mathcal{C} is represented as a $|\mathcal{C}| \times n$ array whose rows are the codewords.

Example 1: Consider the difference set $D = \{+-\}$ consisting of a single disallowed pattern. One can verify that the following 2-, 3-, and 4-bit codes

$$\mathcal{C}_2 \stackrel{\text{def}}{=} \begin{bmatrix} 00 \\ 01 \\ 11 \end{bmatrix} \quad \mathcal{C}_3 \stackrel{\text{def}}{=} \begin{bmatrix} 000 \\ 010 \\ 011 \\ 110 \\ 111 \end{bmatrix} \quad \mathcal{C}_4 \stackrel{\text{def}}{=} \begin{bmatrix} 0000 \\ 0001 \\ 0100 \\ 0101 \\ 0111 \\ 1100 \\ 1101 \\ 1111 \end{bmatrix}$$

avoid D and that there are no larger codes of lengths 2, 3, and 4. Hence

$$\delta_2(D) = 3, \quad \delta_3(D) = 5, \quad \text{and} \quad \delta_4(D) = 8.$$

We will show in Example 10 that for $n \geq 4$

$$\delta_n(D) = \delta_{n-1}(D) + \delta_{n-2}(D)$$

hence

$$\delta_n(D) = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{n+2} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+2} \right)$$

is the shifted Fibonacci sequence, and

$$\text{cap}(D) = \log((1 + \sqrt{5})/2). \quad \square$$

Example 2: Consider the difference set $D = \{0+0\}$. One can verify that the 3- and 4-bit codes

$$\mathcal{C}_3 \stackrel{\text{def}}{=} \begin{bmatrix} 000 \\ 100 \\ 011 \\ 111 \end{bmatrix} \quad \mathcal{C}_4 \stackrel{\text{def}}{=} \begin{bmatrix} 0000 \\ 0001 \\ 1000 \\ 1001 \\ 0110 \\ 0111 \\ 1110 \\ 1111 \end{bmatrix}$$

avoid D and there are no larger codes of length 3 and 4. Hence

$$\delta_3(D) = 4 \quad \text{and} \quad \delta_4(D) = 8.$$

More generally, for all n , the ‘‘repetition’’ code

$$\mathcal{C}_n \stackrel{\text{def}}{=} \begin{cases} \left\{ u_0 u_1 u_1 u_2 u_2 \cdots u_{\frac{n-1}{2}} u_{\frac{n-1}{2}} \right\}, & \text{for } n \text{ odd} \\ \left\{ u_0 u_1 u_1 \cdots u_{\frac{n}{2}-1} u_{\frac{n}{2}-1} u_{\frac{n}{2}} \right\}, & \text{for } n \text{ even} \end{cases}$$

avoids D . Hence

$$\delta_n(D) \geq 2^{\lfloor \frac{n}{2} \rfloor + 1}$$

and

$$\text{cap}(D) \geq 0.5.$$

It will be shown in Section VI that in fact

$$\text{cap}(D) = 0.5. \quad \square$$

Example 3: Consider the difference set $D = \{++, +-\}$ consisting of a pair of patterns. One can verify that the 2-, 3-, and 4-bit codes

$$\mathcal{C}_2 \stackrel{\text{def}}{=} \begin{bmatrix} 00 \\ 10 \end{bmatrix} \quad \mathcal{C}_3 \stackrel{\text{def}}{=} \begin{bmatrix} 000 \\ 001 \\ 100 \\ 101 \end{bmatrix} \quad \mathcal{C}_4 \stackrel{\text{def}}{=} \begin{bmatrix} 0000 \\ 0010 \\ 1000 \\ 1010 \end{bmatrix}$$

avoid D and that no larger codes for lengths 2, 3, and 4 avoid D , hence

$$\delta_2(D) = 2 \quad \text{and} \quad \delta_3(D) = \delta_4(D) = 4.$$

For general n , it can be easily seen that

$$\mathcal{C}_n \stackrel{\text{def}}{=} \begin{cases} \{u_1 0 u_3 0 \cdots 0 u_n\}, & \text{if } n \text{ is odd} \\ \{u_1 0 u_3 0 \cdots u_{n-1} 0\}, & \text{if } n \text{ is even} \end{cases}$$

avoids D , and that if a code \mathcal{C} avoids D then at least one of any two adjacent columns in \mathcal{C} must be constant. Hence

$$\delta_n(D) = 2^{\lceil \frac{n}{2} \rceil}$$

and

$$\text{cap}(D) = 0.5. \quad \square$$

IV. FROM DISALLOWED DIFFERENCES TO JOINT SPECTRAL RADIUS

In this section we describe the main result of the correspondence, showing that the capacity of a difference set is the joint spectral radius of an appropriately defined set of matrices. The proof is presented via a sequence of lemmas in the following sections.

A. Disallowed Joint Patterns

Let \mathcal{C} be an n -bit code. Let

$$P_i \stackrel{\text{def}}{=} \{u_{[i,i']} : u \in \mathcal{C}\}$$

be the set of m -bit patterns present in columns $[i, i']$, and let

$$M_i \stackrel{\text{def}}{=} \{0, 1\}^m - P_i \quad (4)$$

be the set of m -bit patterns missing from those columns.

A *joint pattern* is a set of two m -bit patterns. A joint pattern $\{p, p'\}$ is disallowed for a difference set D if

$$p - p' \in D \quad \text{or} \quad p' - p \in D.$$

Let $\mathcal{J}(D)$ denote the collection of all disallowed joint patterns. Observe that

$$|\mathcal{J}(D)| = \sum_{p \in D} 2^{z(p)}$$

where $z(p)$ is the number of zeros in the pattern p .

Lemma 1: Let \mathcal{C} be an n -bit code and let $M_1, \dots, M_{n'}$ be as defined in (4). Then \mathcal{C} avoids D iff for all $i \in [1, n']$ and all $J \in \mathcal{J}(D)$

$$J \cap M_i \neq \emptyset. \quad (5)$$

Proof: If $\{p, p'\} \in \mathcal{J}(D)$ is a disallowed joint pattern, then p and p' cannot both appear in any set P_i of a code that avoids D , for if they did, we would get a disallowed difference. Hence, $\{p, p'\} \cap M_i$ is not empty for all i .

Conversely, if (5) holds, then for every $u, v \in \mathcal{C}$ and all $i \in [1, n']$, $u_{[i, i']} - v_{[i, i']} \notin D$. \square

Example 4: For $D = \{+-\}$, the collection of disallowed joint patterns is

$$\mathcal{J}(D) = \{\{01, 10\}\}$$

and indeed, $|\mathcal{J}(D)| = 1 = 2^{z(+-)}$. One can verify that $\{01, 10\}$ intersects each M_i of the codes in Example 1.

For $D = \{0+0\}$, the collection of disallowed joint patterns is

$$\mathcal{J}(D) = \{\{010, 000\}, \{011, 001\}, \{110, 100\}, \{111, 101\}\}$$

satisfying $|\mathcal{J}(D)| = 4 = 2^{z(0+0)}$, and each set in $\mathcal{J}(D)$ intersects each M_i of the codes in Example 2.

For $D = \{+++, +- \}$

$$\mathcal{J}(D) = \{\{11, 00\}, \{10, 01\}\}.$$

Again $|\mathcal{J}(D)| = 2 = 2^{z(+++)} + 2^{z(+-)}$, and each set in $\mathcal{J}(D)$ intersects each M_i of the codes in Example 3. \square

B. Representing Sets

A set $M \subseteq \{0, 1\}^m$ *represents* or is a *representing set* for $\mathcal{J}(D)$ if it intersects every set in $\mathcal{J}(D)$. It is *minimal* if, in addition, none of its strict subsets represents $\mathcal{J}(D)$. Clearly, every representing set contains a minimal one. Let $\mathcal{M}(D)$ be the collection of all minimal representing sets for $\mathcal{J}(D)$.

In general, finding the smallest size of a minimal representing set, and therefore finding all of them, is NP-hard, e.g., [14, SP8]. However, in the cases we consider, m is fixed and typically small, hence finding $\mathcal{M}(D)$ is usually not difficult.

Equation (5) implies the following lemma.

Lemma 2: If a code \mathcal{C} avoids D then for every $i \in [1, n']$ the set M_i defined in (4) contains a set $M'_i \in \mathcal{M}(D)$.

Example 5:

$$\mathcal{M}(\{+-\}) = \{\{01\}, \{10\}\}$$

and

$$\mathcal{M}(\{+++, +- \}) = \{\{10, 11\}, \{01, 11\}, \{00, 10\}, \{00, 01\}\}.$$

One can verify that each of the sets M_i of the codes presented in Examples 1 and 3 contains an element of the corresponding $\mathcal{M}(D)$. \square

We can think of the minimal representing sets as the smallest candidate sets of patterns which must be missing from columns $[i, i']$ in a code which avoids D .

C. Disallowed Sets

Let $M_1, \dots, M_{n'} \subseteq \{0, 1\}^m$ be sets of m -bit patterns. An n -bit sequence s_1, \dots, s_n *avoids* the set sequence $M_1, \dots, M_{n'}$ if $s_{[i, i']} \notin M_i$ for all $i \in [1, n']$. We think of these sets as *missing* from s_1, \dots, s_n . Let $\mu(M_1, \dots, M_{n'})$ be the number of n -bit sequences that avoid $M_1, \dots, M_{n'}$. Note that if $M_i \supseteq M'_i$ for all $i \in [1, n']$, then

$$\mu(M_1, \dots, M_{n'}) \leq \mu(M'_1, \dots, M'_{n'}). \quad (6)$$

If \mathcal{M} is a collection of sets in $\{0, 1\}^m$, we let

$$\mu_n(\mathcal{M}) \stackrel{\text{def}}{=} \max\{\mu(M_1, \dots, M_{n'}) : M_i \in \mathcal{M} \forall i\}$$

be the largest number of n -bit sequences avoiding a sequence of sets in \mathcal{M} .

Note that unlike disallowed differences which constrain pairs of sequences, disallowed sets constrain individual sequences. We will show later that this type of constraint is easier to analyze, and we now prove that it leads to the same capacity.

Lemma 3: For every n

$$\delta_n(D) = \mu_n(\mathcal{M}(D)).$$

Proof: Consider a code \mathcal{C} that avoids D and achieves $\delta_n(D)$. For $i \in [1, n']$ let M_i be the set defined in (4), and let M'_i be the sets in $\mathcal{M}(D)$ indicated in Lemma 2. Then, using (6)

$$\begin{aligned} \delta_n(D) = |\mathcal{C}| &\leq \mu(M_1, \dots, M_{n'}) \\ &\leq \mu(M'_1, \dots, M'_{n'}) \leq \mu_n(\mathcal{M}(D)) \end{aligned}$$

where the first inequality follows as, by definition, each codeword in \mathcal{C} avoids $M_1, \dots, M_{n'}$, the second because, as Lemma 2 showed, for each M_i there exists M'_i in $\mathcal{M}(D)$ such that $M'_i \subseteq M_i$ hence

$$\mu(M_i, \dots, M_{n'}) \leq \mu(M'_i, \dots, M'_{n'})$$

and the third from the definition of $\mu_n(\mathcal{M}(D))$.

To establish the reverse inequality, note that if $M_1, \dots, M_{n'} \in \mathcal{M}(D)$, then, by Lemma 1, the set of all n -bit sequences avoiding $M_1, \dots, M_{n'}$ avoids D . \square

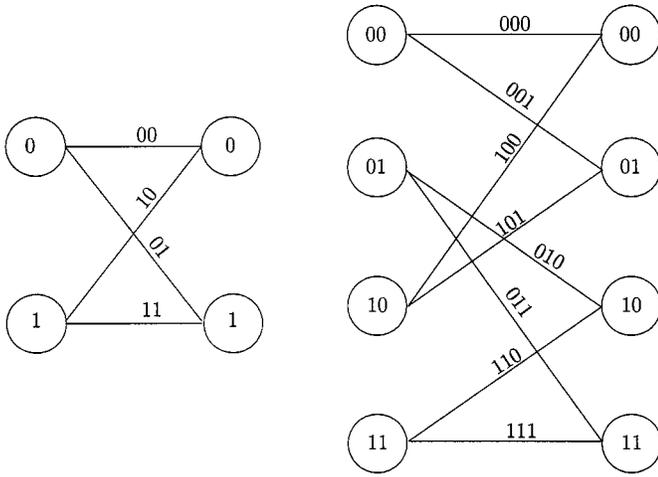
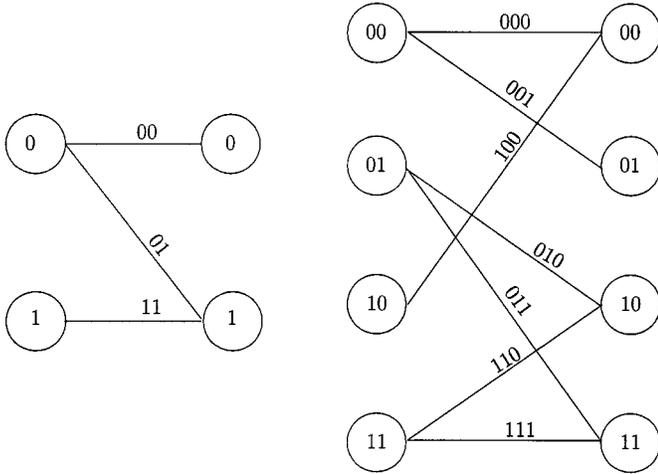
D. Bipartite and Cascade Graphs

In the previous subsection we reduced the difference constraint on pairs of sequences to a constraint on individual sequences. We now convert this problem to that of counting paths in graphs.

A bipartite graph (L, R, E) consists of a set L of *left vertices*, a set R of *right vertices*, and a set E of *edges*. Each edge $(l, r) \in E$ connects a left vertex $l \in L$ to a right vertex $r \in R$. Though we do not draw their direction explicitly, we think of the edges as directed from left to right.

For $m \geq 2$, let G_m be the bipartite graph where

$$L = R = \{0, 1\}^{m-1}$$


 Fig. 2. G_2 and G_3 .

 Fig. 3. $G_{\{10\}}$ and $G_{\{101\}}$.

and $(l_1, \dots, l_{m-1}) \in L$ is connected to $(r_1, \dots, r_{m-1}) \in R$ if $l_i = r_{i-1}$ for all $i = 2, \dots, m-1$. We identify this edge with the m -bit sequence

$$l_1, l_2, \dots, l_{m-1}, r_{m-1} = l_1, r_1, \dots, r_{m-1}.$$

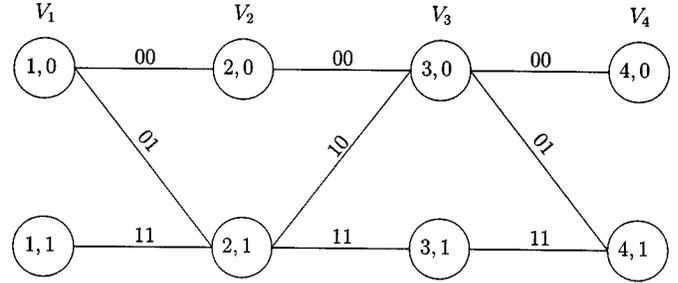
Fig. 2 illustrates G_2 and G_3 .

For $M \subseteq \{0, 1\}^m$, define G_M to be the bipartite graph obtained from G_m by removing the edges corresponding to elements of M . Fig. 3 illustrates $G_{\{10\}}$ and $G_{\{101\}}$.

If $G_1, \dots, G_{n'}$ are bipartite graphs with left vertex sets $L_1, \dots, L_{n'}$ and right vertex sets $R_1, \dots, R_{n'}$, respectively, such that $R_i = L_{i+1}$ for all $i \in [1, n' - 1]$, we let

$$V_i \stackrel{\text{def}}{=} \begin{cases} \{1\} \times L_1, & \text{if } i = 1 \\ \{i\} \times R_{i-1} = \{i\} \times L_i, & \text{if } 2 \leq i \leq n' \\ \{n'+1\} \times R_{n'}, & \text{if } i = n'+1 \end{cases}$$

and define the *cascade* $[G_1, \dots, G_{n'}]$ to be the graph whose vertex set is $V_1 \cup \dots \cup V_{n'+1}$ and where for $i \in [1, n']$, the edges between V_i and V_{i+1} are the edges of G_i , and there are no other edges. Drawing the vertices of each V_i vertically and to the left of the V_{i+1} vertices, we call the vertices V_1 and $V_{n'+1}$ *leftmost* and *rightmost*, respectively. Fig. 4 illustrates the cascade $[G_{\{10\}}, G_{\{01\}}, G_{\{10\}}]$.


 Fig. 4. $[G_{\{10\}}, G_{\{01\}}, G_{\{10\}}]$.

A *path* in a cascade $[G_1, \dots, G_{n'}]$ is a sequence $v_1, \dots, v_{n'+1}$ of vertices where each $v_i \in V_i$, and v_i is connected to v_{i+1} for all $i \in [1, n']$. Note that all paths connect a leftmost vertex to a rightmost vertex and proceed from left to right. We let $\psi([G_1, \dots, G_{n'}])$ be the total number of paths in the cascade. For example, in $[G_{\{10\}}, G_{\{01\}}, G_{\{10\}}]$ there are two paths from $(1, 0)$ to $(4, 0)$, three paths from $(1, 0)$ to $(4, 1)$, etc. Hence

$$\psi([G_{\{10\}}, G_{\{01\}}, G_{\{10\}}]) = 2 + 3 + 1 + 2 = 8.$$

If $M \subseteq \{0, 1\}^m$ and $s \in \{0, 1\}^m$, then $l = s_{[1, m-1]}$ and $r = s_{[2, m]}$ are vertices of G_M with an edge from left node l to right node r if and only if $s \notin M$, namely, s avoids M . More generally, for $n \geq m$, there is a bijection between n -bit sequences that avoid $M_1, \dots, M_{n'}$ and paths in the cascade $[G_{M_1}, \dots, G_{M_{n'}}]$, hence

$$\mu(M_1, \dots, M_{n'}) = \psi([G_{M_1}, \dots, G_{M_{n'}}]).$$

Letting

$$\psi_n(\mathcal{M}) \stackrel{\text{def}}{=} \max\{\psi([G_{M_1}, \dots, G_{M_{n'}}]) : M_i \in \mathcal{M} \forall i\}$$

we obtain the following lemma.

Lemma 4:

$$\mu_n(\mathcal{M}(D)) = \psi_n(\mathcal{M}(D)). \quad \square$$

Example 6: Let $D = \{+-\}$. One can verify that $\psi_4(\mathcal{M}(D)) = 8$, achieved by the cascade $[G_{\{10\}}, G_{\{01\}}, G_{\{10\}}]$, illustrated in Fig. 4. There is a bijection between the paths in the cascade and the codewords in Example 1. \square

E. Adjacency Matrices

Identifying the elements of L and R of a bipartite graph $G = (L, R, E)$ with the intervals $[1, |L|]$ and $[1, |R|]$, respectively, we let the *adjacency matrix* A_G be the $|L| \times |R|$ matrix whose (l, r) th element is 1 if $(l, r) \in E$, and 0 otherwise.

Note that the (l, r) th element of A_G is the number of edges from left node l to right node r in G . Similarly, it can be shown that in the cascade $[G_1, \dots, G_{n'}]$, the number of left-to-right paths from leftmost vertex l to rightmost vertex r is the (l, r) th element of the product $A_{G_1} A_{G_2} \dots A_{G_{n'}}$.

Letting

$$\|A\|_1 = \sum_{l,r} |A_{l,r}| \quad (7)$$

denote the L_1 norm of the matrix A , it follows that, for every $M_1, \dots, M_{n'} \subseteq \{0, 1\}^m$

$$\psi([G_{M_1}, \dots, G_{M_{n'}}]) = \|A_{G_{M_1}} \cdot \dots \cdot A_{G_{M_{n'}}}\|_1.$$

Let

$$\Sigma(D) \stackrel{\text{def}}{=} \{A_{G_M} : M \in \mathcal{M}(D)\}$$

denote the set of adjacency matrices corresponding to the collection $\mathcal{M}(D)$ of minimal representing sets for the disallowed joint patterns $\mathcal{J}(D)$ (see Sections IV-A and IV-B for definitions). Let

$$\Sigma^n \stackrel{\text{def}}{=} \left\{ \prod_{i=1}^n A_i : A_i \in \Sigma \right\}$$

denote the set of products of n matrices in Σ . Then, setting

$$\hat{\rho}_n(\Sigma, \|\cdot\|_1) \stackrel{\text{def}}{=} \max\{\|A\|_1 : A \in \Sigma^n\}$$

for an arbitrary set $\Sigma \subseteq \mathbb{C}^{m \times m}$, we get the following lemma.

Lemma 5:

$$\psi_n(\mathcal{M}(D) = \hat{\rho}_n(\Sigma(D), \|\cdot\|_1). \quad \square$$

This suggests looking for algebraic methods to determine the capacity.

F. Matrix Norms and Spectral Radius

A *matrix norm* for the set $\mathbb{C}^{m \times m}$ of complex square matrices is a mapping $\|\cdot\| : \mathbb{C}^{m \times m} \rightarrow [0, \infty)$ such that for all $A_1, A_2, A \in \mathbb{C}^{m \times m}$

- 1) $\|A\| = 0$, iff $A = 0$
- 2) $\|A_1 + A_2\| \leq \|A_1\| + \|A_2\|$
- 3) $\|A_1 \cdot A_2\| \leq \|A_1\| \cdot \|A_2\|$
- 4) $\|cA\| = |c| \cdot \|A\| \quad \forall c \in \mathbb{C}$.

Example 7: Let $A \in \mathbb{C}^{m \times m}$. The L_1 norm, $\|A\|_1$ of A was already defined in (7). The *maximum-column-sum norm* of A is

$$\|A\|_\gamma \stackrel{\text{def}}{=} \max_{1 \leq j \leq m} \sum_{i=1}^m |A_{i,j}|$$

and the *spectral norm* of $A \in \mathbb{C}^{m \times m}$ is

$$\|A\|_s \stackrel{\text{def}}{=} \max_{\|x\|_2=1} \|Ax\|_2$$

where

$$\|x\|_2 \stackrel{\text{def}}{=} \left(\sum_{i=1}^m |x_i|^2 \right)^{1/2}$$

is the *Euclidean norm* of a vector $x \in \mathbb{C}^m$. It can be shown that $\|\cdot\|_1$, $\|\cdot\|_\gamma$, and $\|\cdot\|_s$ are all matrix norms. \square

One can show, e.g., [15, Theorem 5.4.4], that for any two matrix norms $\|\cdot\|_\alpha$, $\|\cdot\|_\beta$ there are constants $0 < c_1 < c_2$ such that for all $A \in \mathbb{C}^{m \times m}$

$$c_1 \|A\|_\beta \leq \|A\|_\alpha \leq c_2 \|A\|_\beta. \quad (8)$$

By submultiplicativity of matrix norms, the limit

$$\hat{\rho}(A) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \|A^n\|^{1/n}$$

exists, and, by (8), is independent of the matrix norm $\|\cdot\|$. One can also define

$$\check{\rho}(A) \stackrel{\text{def}}{=} \max\{|\lambda| : \lambda \text{ an eigenvalue of } A\}.$$

For any matrix norm and $A \in \mathbb{C}^{m \times m}$ we have, e.g., [15, Theorem 5.6.9]

$$\check{\rho}(A) \leq \|A\|. \quad (9)$$

It is also well known, e.g., [15, Corollary 5.6.14], that

$$\check{\rho}(A) = \hat{\rho}(A).$$

They are called the *spectral radius* of A and denoted by $\rho(A)$.

G. Joint Spectral Radius

The quantities $\hat{\rho}$ and $\check{\rho}$ can be generalized to sets of matrices. We begin with $\hat{\rho}$. Letting

$$\hat{\rho}_n(\Sigma, \|\cdot\|) \stackrel{\text{def}}{=} \sup\{\|A\| : A \in \Sigma^n\}$$

for an arbitrary matrix norm $\|\cdot\|$ and set $\Sigma \subseteq \mathbb{C}^{m \times m}$, Rota and Strang [16] defined the *joint spectral radius* of Σ to be

$$\hat{\rho}(\Sigma) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \hat{\rho}_n(\Sigma, \|\cdot\|)^{1/n}$$

where the limit exists by submultiplicativity, and (8) implies that it is independent of the norm $\|\cdot\|$.

Daubechies and Lagarias [17] defined the *generalized spectral radius* of Σ to be

$$\check{\rho}(\Sigma) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \sup \check{\rho}_n(\Sigma)^{1/n}$$

where

$$\check{\rho}_n(\Sigma) \stackrel{\text{def}}{=} \sup\{\check{\rho}(A) : A \in \Sigma^n\}.$$

It follows from (9) that

$$\check{\rho}_n(\Sigma) \leq \hat{\rho}_n(\Sigma, \|\cdot\|)$$

for every n . Hence

$$\check{\rho}(\Sigma) \leq \hat{\rho}(\Sigma)$$

and Daubechies and Lagarias conjectured that equality holds, namely

$$\check{\rho}(\Sigma) = \hat{\rho}(\Sigma)$$

as was proven by Berger and Wang [18] for all finite Σ . We denote this quantity by $\rho(\Sigma)$, and refer to it as the *joint spectral radius*.

Combining (3) and Lemmas 3 to 5, we obtain our main result.

Theorem 1: For every finite D

$$\text{cap}(D) = \log(\rho(\Sigma(D))). \quad \square$$

Namely, the capacity is the logarithm of the joint spectral radius of $\Sigma(D)$.

This equality generalizes known results on *constrained systems* where, instead of differences, certain patterns are disallowed, and it is well known, e.g., [19, Theorem 3.9], that the growth rate of the number of sequences, or *Shannon capacity* of the constraint, is $\log(\hat{\rho}(A))$, the logarithm of the spectral radius of a corresponding adjacency matrix A .

The joint spectral radius measures the maximum growth rate of a product of matrices drawn from the set Σ . This concept appears in many applications. In addition to Rota and Strang's original work in matrix theory [16], it has been used to study convergence of infinite products

of matrices, e.g., [20], with applications to wavelets [17]. The concept is also related to the stability properties of discrete linear inclusions, e.g., [21], [22], wherein the logarithm of the joint spectral radius is referred to as the *Lyapunov indicator*.

In the next section we describe several existing algorithms for computing the joint spectral radius and introduce some new ones.

V. COMPUTING THE JOINT SPECTRAL RADIUS

A. Computing the Joint Spectral Radius Is Hard

Tsitsiklis and Blondel [23] have shown that approximating the joint spectral radius of a pair of matrices with $\{0, 1\}$ entries is NP-hard. In addition, they have shown [24] that determining whether $\rho(\Sigma) \leq 1$ when Σ is a set of nonnegative rational matrices is undecidable. Hence, the problem of determining the joint spectral radius of a set of nonnegative rational matrices is undecidable.

Note that $\{\Sigma(D) \mid D \text{ a difference set}\}$ is a subclass of the set of $\{0, 1\}$ matrices. It is currently unresolved whether or not determining the capacity of a difference set is NP-hard. Nonetheless, Tsitsiklis and Blondel's results point to the difficulty of finding efficient algorithms to determine the capacity of a given difference set.

Here we determine the capacity of several simple difference sets that arise in practice.

B. Existing Algorithms

Because of the submultiplicativity of $\hat{\rho}_n(\Sigma, \|\cdot\|)$

$$\rho(\Sigma) = \hat{\rho}(\Sigma) \leq \hat{\rho}_n(\Sigma, \|\cdot\|)^{1/n}$$

for every n . Furthermore, as n increases, this upper bound generally better approximates the joint spectral radius in the sense that for every n , there exists an $n' > n$ such that

$$\hat{\rho}_{n'}(\Sigma, \|\cdot\|)^{1/n'} \leq \hat{\rho}_n(\Sigma, \|\cdot\|)^{1/n}.$$

Similarly, every $\check{\rho}_n(\Sigma)$ lower-bounds $\rho(\Sigma)$, and as n increases, $\check{\rho}_n(\Sigma)$ generally better approximates the joint spectral radius from below, in the sense that, for every n

$$\check{\rho}_n(\Sigma)^{1/n} \leq \check{\rho}_{kn}(\Sigma)^{1/kn}$$

for any $k \geq 1$.

This suggests approximating the joint spectral radius $\rho(\Sigma)$ by computing the lower bounds

$$\max_{1 \leq k \leq n} \check{\rho}_k(\Sigma)^{1/k}$$

and upper bounds

$$\min_{1 \leq k \leq n} \hat{\rho}_k(\Sigma, \|\cdot\|)^{1/k}$$

for $n = 1, 2, \dots$. However, the number of matrix operations increases as $|\Sigma|^n$, such that determining $\rho(\Sigma)$ with an arbitrary error may be computationally prohibitive.

Several steps have been taken to reduce the growth rate of the number of computations required to approximate $\rho(\Sigma)$. Maesumi [25] has shown the number of matrix operations may be reduced from $|\Sigma|^n$ to $|\Sigma|^n/n$. Daubechies and Lagarias [17] proved the following result.

Lemma 6: If $\{A_j\}$ is a set of building blocks for Σ , i.e.,

- a) each A_j is the product of n_j matrices drawn from Σ ,
- b) there exists some $n_0 \geq 0$ such that, if A is a finite product of elements of Σ , then $A = A_{j_1} \cdots A_{j_k} Q$, where Q is a product of at most n_0 elements of Σ ,

then $\rho(\Sigma) \leq \sup \|A_j\|^{1/n_j}$. \square

Lemma 6 can be used to implement a recursive "branch-and-bound" algorithm to upper-bound $\rho(\Sigma)$, e.g., [17], [26], [27]. Gripenberg [28]

has provided an algorithm based on Lemma 6 which includes a sequence of lower bounds such that $\rho(\Sigma)$ may be specified to lie within an arbitrarily small interval.

The remainder of this section describes an algorithm which empirical results show has a computation time competitive with the algorithm in [28].

C. The Pruning Algorithm

We present an alternative method, the pruning algorithm, for bounding $\rho(\Sigma)$ when all the matrices in Σ are nonnegative. The method replaces the search for the largest norm among all (exponentially many) products of n matrices with a search over a smaller set with the same largest norm. It can be applied to compute $\hat{\rho}_n(\Sigma)$ and $\hat{\rho}_n(\Sigma, \|\cdot\|)$ for several norms.

We write $A \geq 0$ if every element of A is nonnegative and $A \geq B$ if every element of A is at least as large as the corresponding element of B . It can be shown, e.g., [15, Theorem 8.1.18], that if $A \geq B \geq 0$ then

$$\rho(A) \geq \rho(B). \quad (10)$$

A matrix A dominates matrix B with respect to the norm $\|\cdot\|$ if

$$\|AM\| \geq \|BM\|$$

for all $M \geq 0$. In particular

$$\|A\| \geq \|B\|.$$

A subset S of Σ^n is *dominating* if every matrix in Σ^n is dominated by some matrix in S . Let Ψ_n be any dominating subset of Σ^n . By definition

$$\hat{\rho}_n(\Sigma, \|\cdot\|) = \max\{\|A\| : A \in \Psi_n\}. \quad (11)$$

Furthermore, it is easy to verify that if all matrices in Σ are nonnegative then $\Psi_n \Sigma$ is a dominating subset of Σ^{n+1} , namely

$$\Psi_{n+1} \subseteq \Psi_n \Sigma.$$

Given a matrix norm one can, therefore, construct a recursive algorithm which computes a dominating set Ψ_n from Ψ_{n-1} by considering all products in $\Psi_{n-1} \Sigma$ and "pruning" those that are dominated by another product. The subsequent growth rate of $|\Psi_n|$ will depend on the condition for domination. The following lemmas provide sufficient conditions for domination with respect to the L_1 norm, the maximum-column-sum norm, and the spectral norm.

We write $A \geq_C B$ if every column-sum of A is at least as large as the corresponding column-sum of B .

Lemma 7: If $A \geq_C B$, then A dominates B with respect to $\|\cdot\|_1$ and $\|\cdot\|_\gamma$.

Proof: Clearly, if $A \geq_C B$, then $\|A\|_1 \geq \|B\|_1$ and $\|A\|_\gamma \geq \|B\|_\gamma$. Domination follows as $A \geq_C B$ implies $AM \geq_C BM$ for every $M \geq 0$. \square

It can be shown, e.g., [15, Sec. 5.6.6], that

$$\|A\|_s = \rho(A^*A)^{1/2} \quad (12)$$

where A^* denotes the Hermitian adjoint of A . This can be used to prove the following lemma.

Lemma 8: If $A^*A \geq B^*B$, then A dominates B with respect to $\|\cdot\|_s$.

Proof: Since A^*A and B^*B are nonnegative, (10) implies

$$\rho(A^*A) \geq \rho(B^*B)$$

hence

$$\|A\|_s \geq \|B\|_s.$$

Domination follows as for every $M \geq 0$

$$\begin{aligned} (AM)^*AM &= M^t A^*AM \\ &\geq M^t B^*BM = (BM)^*BM. \end{aligned} \quad \square$$

An analogous algorithm, based on (10) may be used to construct a sequence of convergent lower bounds on $\rho(\Sigma)$ and

VI. CAPACITIES OF CERTAIN DIFFERENCE SETS

A. Explicit Computation

In some cases, one can find the joint spectral radius, and hence the capacity, exactly.

Example 8: Let $D = \{0^m +\}$, $m \geq 0$. We show that $\delta_n(D) \leq 2^m$ for all $n \geq m$. For $m = 0$, the result is obvious. Suppose there exists a code \mathcal{C} which avoids D with $|\mathcal{C}| > 2^m$. Then there must exist $u, v \in \mathcal{C}$, $u \neq v$ such that $u_{[1,m]} = v_{[1,m]}$. For \mathcal{C} to avoid D , we must have $u_{[i]} = v_{[i]}$ for $m < i \leq n$. Hence $u = v$, a contradiction. Therefore, $\delta_n(D) \leq 2^m$ and

$$\text{cap}(\{0^m +\}) = 0$$

for all $m \geq 0$. □

Example 9: Let $D = \{0 + 0\}$. One can show that

$$\begin{aligned} \tilde{\rho}_2(\{0 + 0\}) &= \hat{\rho}_2(\{0 + 0\}, \|\cdot\|_s) \\ &= \rho\left(A_{G_{\{001, 010, 101, 110\}}} A_{G_{\{010, 011, 100, 101\}}}\right) \\ &= 2, \end{aligned}$$

and therefore

$$\text{cap}(\{0 + 0\}) = 0.5. \quad \square$$

The pruning algorithm may lead to a direct computation of the joint spectral radius via an inductive argument.

Example 10: For $D = \{+-\}$ we have

$$\Sigma(\{+-\}) = \left\{A_{G_{\{011\}}}, A_{G_{\{101\}}}\right\}.$$

By application of the domination condition in Lemma 7 and inspection of $\Sigma(\{+-\})$, one can show that for $n = 1, 2, \dots$

$$\begin{aligned} \Psi_{2n} &= \left\{\left(A_{G_{\{011\}}} A_{G_{\{101\}}}\right)^n, \left(A_{G_{\{101\}}} A_{G_{\{011\}}}\right)^n\right\} \\ \Psi_{2n+1} &= \left\{A_{G_{\{101\}}} \left(A_{G_{\{011\}}} A_{G_{\{101\}}}\right)^n, \right. \\ &\quad \left. A_{G_{\{011\}}} \left(A_{G_{\{101\}}} A_{G_{\{011\}}}\right)^n\right\} \end{aligned}$$

from which it follows that

$$\delta_n(\{+-\}) = \delta_{n-1}(\{+-\}) + \delta_{n-2}(\{+-\})$$

where $\delta_1 = 2$ and $\delta_2 = 3$. Hence, as shown in Example 1,

$$\text{cap}(\{+-\}) = \log_2((1 + \sqrt{5})/2) = 0.6942\dots \quad \square$$

An inductive argument may lead to a direct computation in cases where the convergence rate of the bounds is slow.

Example 11: For $D = \{++-\}$ we have

$$\Sigma(\{++-\}) = \left\{A_{G_{\{110\}}}, A_{G_{\{001\}}}\right\}$$

$$\begin{aligned} \max_{1 \leq k \leq 316} \log_2 \tilde{\rho}_k(\Sigma)^{1/k} &= 0.8113\dots \\ &\leq \text{cap}(D) \leq \min_{1 \leq k \leq 316} \log_2 \hat{\rho}_k(\|\cdot\|_s, \Sigma)^{1/k} = 0.8116\dots \end{aligned}$$

However, one can show by an inductive argument using the domination condition in Lemma 7 that for $n = 1, 2, \dots$

$$\begin{aligned} \Psi_{4n} &= \left\{\left(A_{G_{\{110\}}} A_{G_{\{110\}}} A_{G_{\{001\}}} A_{G_{\{001\}}}\right)^n, \right. \\ &\quad \left(A_{G_{\{001\}}} A_{G_{\{110\}}} A_{G_{\{110\}}} A_{G_{\{001\}}}\right)^n, \\ &\quad \left(A_{G_{\{001\}}} A_{G_{\{001\}}} A_{G_{\{110\}}} A_{G_{\{110\}}}\right)^n, \\ &\quad \left(A_{G_{\{110\}}} A_{G_{\{001\}}} A_{G_{\{001\}}} A_{G_{\{110\}}}\right)^n\right\} \end{aligned}$$

Hence

$$\begin{aligned} \rho(\Sigma) &= \lim_{n \rightarrow \infty} \left\| \left(A_{G_{\{110\}}} A_{G_{\{110\}}} A_{G_{\{001\}}} A_{G_{\{001\}}}\right)^n \right\|^{1/4n} \\ &= \rho\left(A_{G_{\{110\}}} A_{G_{\{110\}}} A_{G_{\{001\}}} A_{G_{\{001\}}}\right)^{1/4} \\ &= (2 + ((25 - 3\sqrt{69})/2)^{1/3} \\ &\quad + ((25 + 3\sqrt{69})/2)^{1/3})/3 \end{aligned}$$

and

$$\text{cap}(\{++-\}) = \log_2 \rho(\Sigma) = 0.8113\dots \quad \square$$

B. Simplifications

When all matrices in Σ are Hermitian, it follows from (12) that

$$\tilde{\rho}_1(\Sigma) = \hat{\rho}_1(\Sigma, \|\cdot\|_s)$$

hence

$$\rho(\Sigma) = \hat{\rho}_1(\Sigma, \|\cdot\|_s).$$

For example, this can be used to provide a simple proof of $\text{cap}(\{++\})$.

Example 12: For $D = \{++\}$ we have

$$\Sigma(\{++\}) = \left\{\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\right\},$$

hence $\text{cap}(\{++\}) = \log((1 + \sqrt{5})/2)$. □

The following lemma equates the capacities of certain pairs of difference sets for which there exists a bijection between codes which avoid the sets.

Lemma 9: Fix a difference set D . Let D' be the difference set obtained by inverting the symbols $+$, $-$ at odd positions of the patterns in D . Then

$$\text{cap}(D) = \text{cap}(D').$$

Proof: Suppose \mathcal{C} avoids D . Let \mathcal{C}' be the code constructed by inverting every other bit of the codewords of \mathcal{C} , for example, every bit corresponding to an even position index. (One could equally well invert every bit corresponding to an odd position index.) It is straightforward to show that \mathcal{C}' avoids D' , hence $\delta_n(\mathcal{C}') \geq \delta_n(D)$. Similarly, one can map any code which avoids D' to a code which avoids D by inverting

TABLE I
CAPACITY OF VARIOUS DIFFERENCE SETS D

$$\alpha = \log_2((1 + \sqrt{5})/2) = .6942\dots$$

$$\gamma = \log_2((2 + (\frac{25 - 3\sqrt{69}}{2})^{1/3} + (\frac{25 + 3\sqrt{69}}{2})^{1/3})/3) = .8113\dots,$$

$$\delta = \log_2((1 + (19 + 3\sqrt{33})^{1/3} + (19 - 3\sqrt{33})^{1/3})/3) = .8791\dots,$$

$$\epsilon = \log_2(\sqrt{(3 + \sqrt{17})/2}) = .9162\dots,$$

$$\eta = \log_2((3 + \sqrt{3\zeta} + \sqrt{99 - 3\zeta + 234\sqrt{3/\zeta}})/12) = .9467\dots,$$

WHERE $\zeta = 11 - 56\beta + 4/\beta$, AND $\beta = (2/(-65 + 3\sqrt{1689}))^{1/3}$.

m	D	$\text{cap}(D)$	\mathcal{O}
$m \geq 1$	$0^{m-1}+$	0	—
2	$+-$	α	$10^{(0)}, 01^{(1)}$
	$++$	α	11
3	$0+-$	$[\alpha, .6948]$	101, 010
	$0+0$.5	$00^{(1)}, 11^{(1)}$
	$0++$	$[\alpha, .6948]$	11
	$+0-$	α	$110^{(0)}, 001^{(0)},$ $011^{(1)}, 100^{(1)}$
	$+0+$	α	101, 111
	$++-$	γ	$110^{(0)}, 110^{(1)},$ $001^{(2)}, 001^{(3)}$
	$+++$	δ	111
	$+-+$	δ	$101^{(0)}, 010^{(1)}$
4	$0+--$	$[\delta, .8797]$	1010, 0101
	$0+++$	$[\delta, .8797]$	111
	$+--+$	$[\eta, .9468]$	$1010^{(0)}, 0101^{(1)}$
	$++++$	$[\eta, .9468]$	1111
5	$0+--+0$	$[\epsilon, .9164]$	$1010^{(1)}, 0101^{(1)}$
	$0++++0$	$[\epsilon, .9164]$	$1111^{(1)}, 0000^{(1)}$

every other bit of the codewords of C' . Hence, $\delta_n(D') = \delta_n(D)$ and $\text{cap}(D) = \text{cap}(D')$. \square

For example, it follows from Lemma 9 that

$$\begin{aligned} \text{cap}(\{++\}) &= \text{cap}(\{+-\}), \\ \text{cap}(\{+++ \}) &= \text{cap}(\{+-- \}), \quad \text{and} \\ \text{cap}(\{0+- \}) &= \text{cap}(\{0++ \}). \end{aligned}$$

C. Results

Table I summarizes known values or ranges of $\text{cap}(D)$ for all difference sets D consisting of a single pattern of length ≤ 3 and some patterns of larger length. Since the same number of sequences avoid a pattern p as its negation $-p$, we assume that the first nonzero element of p is $+$. Also, we do not list $\text{cap}(D)$ if the (identical) capacity of the string obtained by reversing the order of p has already been addressed.

Next to the capacity, we list a constraint describing a sequence of codes, $\{C_n\}$, such that each C_n avoids D and

$$\lim_{n \rightarrow \infty} \log |C_n|^{1/n}$$

achieves the lower bound on the capacity, or $\text{cap}(D)$ when it is known. In a notation similar to that used to describe shift spaces [29, Definition 1.2.1], the constraint is defined by a list of forbidden patterns \mathcal{O}

and the codes C_n can be taken to be the largest n -bit codes satisfying the constraint. If no superscript is listed with a pattern, the pattern is forbidden from appearing in all columns of the code. If superscripts appear, then the patterns are periodically forbidden and the period is one more than the largest superscript. The superscript then represents the column indexes (modulo the period) in which the pattern cannot appear. For example, 101, 010 means that these triples do not appear in any three consecutive columns, and $10^{(0)}, 01^{(1)}$ means that 10 does not appear in columns $[i, i+1]$ for even i and 01 does not appear in columns $[i, i+1]$ for odd i .

Several of these constraints have appeared in the magnetic recording literature. $\mathcal{O} = \{00^{(1)}, 11^{(1)}\}$ is referred to as the *biphase* constraint [30], $\mathcal{O} = \{1010, 0101\}$ as the *MTR* constraint [5], and $\mathcal{O} = \{1010^{(1)}, 0101^{(1)}\}$ as the *TMTR* constraint [6].

REFERENCES

- [1] R. Karabed and N. Nazari, "Analysis of error sequences for PRML and EPRML signaling performed over Lorentzian channel," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM'96)*, vol. 1, London, U.K., Nov. 1996, pp. 368–373.
- [2] S. A. Altekhar, "Detection and coding techniques for magnetic recording channels," Ph.D. dissertation, Univ. California, San Diego, CA, Aug. 1997.
- [3] B. E. Moision, P. Siegel, and E. Soljanin, "Error event characterization and coding for the equalized Lorentzian channel," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 77.
- [4] R. Karabed and P. H. Siegel, "Coding for higher order partial response channels," in *Proc. SPIE*, vol. 2605, M. R. Raghuvver, S. A. Dianat, S. W. McLaughlin, and M. Hassner, Eds., Philadelphia, PA, Oct. 1995, pp. 92–102.
- [5] J. Moon and B. Brickner, "Maximum transition run codes for data storage systems," *IEEE Trans. Magn.*, vol. 32, pp. 3992–3994, Sept. 1996.
- [6] W. Bliss, "An 8/9 rate time-varying trellis code for high density magnetic recording," *IEEE Trans. Magn.*, vol. 33, pp. 2746–2748, Sept. 1997.
- [7] K. K. Fitzpatrick and C. S. Modlin, "Time-varying MTR codes for high density magnetic recording," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM'97)*, vol. 3, Phoenix, AZ, Nov. 1997, pp. 1250–1253.
- [8] R. Karabed, P. Siegel, and E. Soljanin, "Constrained coding for binary channels with high intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1777–1797, Sept. 1999.
- [9] B. E. Moision, P. Siegel, and E. Soljanin, "Distance-enhancing codes for digital recording," *IEEE Trans. Magn.*, vol. 34, pp. 69–74, Jan. 1998.
- [10] G. D. Forney Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 18, pp. 363–378, May 1972.
- [11] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2260–2299, Oct. 1998.
- [12] L. C. Barbosa, "Maximum-likelihood sequence estimators: A geometric view," *IEEE Trans. Inform. Theory*, vol. 35, pp. 419–427, Mar. 1989.
- [13] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey Theory*. New York: Wiley, 1990.
- [14] M. D. Garey and D. S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*. New York: Freeman, 1979.
- [15] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1988.
- [16] G. C. Rota and G. Strang, "A note on the joint spectral radius," *Indagationes Mathematicae*, vol. 22, pp. 379–381, 1960.
- [17] I. Daubechies and J. C. Lagarias, "Two-scale difference equations II. Local regularity, infinite products of matrices and fractals," *SIAM J. Math. Anal.*, vol. 23, pp. 1031–1079, 1992.
- [18] M. A. Berger and Y. Wang, "Bounded semi-groups of matrices," *Linear Algebra Applications*, vol. 166, pp. 21–27, 1992.
- [19] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained systems and coding for recording channels," in *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, Eds. Amsterdam, The Netherlands: Elsevier Science, 1998, ch. 20.
- [20] I. Daubechies and J. C. Lagarias, "Sets of matrices all infinite products of which converge," *Linear Algebra Applications*, vol. 161, pp. 227–263, 1992.

- [21] L. Gurvits, "Stability of discrete linear inclusion," *Linear Algebra Applications*, vol. 231, pp. 47–85, 1995.
- [22] N. E. Barabanov, "Lyapunov indicators of discrete inclusions. I.," *Automat. Remote Contr.*, vol. 49, pp. 152–157, Feb. 1988.
- [23] J. N. Tsitsiklis and V. D. Blondel, "The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate," *Math. Control, Signals, Syst.*, vol. 10, pp. 31–40, 1997.
- [24] V. D. Blondel and J. N. Tsitsiklis, "Boundedness of finitely generated matrix semigroups is undecidable," preprint, Jan. 1999.
- [25] M. Maesumi, "An efficient lower bound for the generalized spectral radius of a set of matrices," *Linear Algebra Applications*, vol. 240, pp. 1–7, 1996.
- [26] —, "Spectral radius of sets of matrices," in *Proc. SPIE*, vol. 2303, San Diego, CA, July 1994, pp. 27–29.
- [27] D. Colella and C. Heil, "The characterization of continuous, four-coefficient scaling functions and wavelets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 876–881, Mar. 1992.
- [28] G. Gripenberg, "Computing the joint spectral radius," *Linear Algebra Applications*, vol. 234, pp. 43–60, 1996.
- [29] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. New York: Cambridge Univ. Press, 1995.
- [30] R. Karabed and P. H. Siegel, "Matched spectral-null codes for partial-response channels," *IEEE Trans. Inform. Theory*, vol. 37, pp. 818–855, May 1991.

On the Optimal Markov Chain of IS Simulation

Kenji Nakagawa, *Member, IEEE*

Abstract—We investigate the importance sampling (IS) simulation for the sample average of an output sequence from an irreducible Markov chain. The optimal Markov chain used in simulation is known to be a twisted Markov chain, however, the proofs in [2], [3] are very complicated and do not give us a good perspective. We give a simple and natural proof for the optimality of the simulation Markov chain in terms of the Kullback–Leibler (KL) divergence of Markov chains. The performance degradation of the IS simulation by using a not optimal simulation Markov chain, i.e., the difference between the obtained variance and the minimum variance is shown to be represented by the KL divergence. Moreover, we show a geometric relationship between a simulation Markov chain and the optimal one.

Index Terms—Importance sampling simulation, information geometry, Kullback–Leibler (KL) divergence, Markov chain.

I. INTRODUCTION

The importance sampling (IS) simulation technique has been used to obtain quickly an accurate estimate for a very small probability that is not tractable by the ordinary Monte Carlo (MC) simulation. The IS technique is widely used for various types of engineering problems, e.g., the estimation of a blocking probability in queuing system [2], [4], [7], an error rate in communications system [2], [6], etc. See [2] for an overview of the application of IS simulation.

If the target event (blocking in queuing systems, or error in communications systems) is a rare event with small probability of less than about 10^{-6} , it is impossible to obtain an estimate by the ordinary MC

simulation because of the limit of simulation time and the precision limit of pseudo-random numbers. To overcome these difficulties, a different probability distribution from the underlying probability distribution is used for simulation in order to generate more samples in the target event. Then the obtained value is modified by the likelihood ratio to obtain an unbiased estimate. This estimate is called an IS estimate. The probability distribution used for simulation is called a simulation distribution. If the simulation distribution is appropriately chosen, the variance of the IS estimate can be smaller than that of the MC estimate. The simulation distribution that yields the IS estimate of the minimum variance is referred to as the optimal simulation distribution. When we apply the IS technique to some simulation problem, it is critical to find the optimal simulation distribution.

In the case of a Markov chain, the Markov chain which is used in the IS simulation is called a simulation Markov chain, and the optimal one is called the optimal simulation Markov chain. It has been known [2], [3] that the optimal simulation Markov chain is unique and belongs to the class of twisted Markov chains (TMC), but the proofs in [2], [3] are complicated and do not give a good perspective. In [2], the perturbation technique is used to prove that the optimal simulation Markov chain is a TMC. In [3], Jensen's inequality is used in the proof and the line of argument is elementary but complicated. Both of the proofs in [2] and [3] are lengthy.

In [10], we studied geometric properties of IS simulation and showed that the Kullback–Leibler (KL) divergence of Markov chains plays an important role in this problem. In this correspondence, we give a geometric view to this problem and provide a simple and natural proof for the optimality of a simulation Markov chain in terms of KL divergence. The performance degradation of the IS simulation by using a not optimal simulation Markov chain, i.e., the difference between the obtained variance and the minimum variance is shown to be represented by the KL divergence. Moreover, we show a geometric relationship between a simulation Markov chain and the optimal one.

II. IMPORTANCE SAMPLING SIMULATION FOR MARKOV CHAINS

We investigate a simulation for the sample average of an output sequence from an irreducible finite-state Markov chain.

Let P_0 denote an irreducible Markov chain on the state space $\Omega \equiv \{0, 1, \dots, K\}$, $K > 0$. The state transition probability matrix of P_0 is denoted by $P_0 = (P_0(x'|x))_{x, x' \in \Omega}$, and the initial distribution is given by the stationary distribution $p_0 = (p_0(x))_{x \in \Omega}$ of P_0 . The joint probability of $x, x' \in \Omega$ is denoted by $P_0(x, x') \equiv p_0(x)P_0(x'|x)$. Consider a mapping $f: \Omega \times \Omega \rightarrow Z$, where Z denotes the set of integers. Denote by $E_{P_0}[f]$ the expectation of f with respect to P_0

$$E_{P_0}[f] \equiv \sum_{x, x' \in \Omega} P_0(x, x')f(x, x'). \quad (1)$$

Assume $E_{P_0}[f] = 0$ without loss of generality.

Let $\mathbf{x}^n = (x_1, x_2, \dots, x_n) \in \Omega^n$ be a sample sequence generated by the Markov chain P_0 . We consider the probability of the following set A_n :

$$A_n = \left\{ \mathbf{x}^n \in \Omega^n \left| \frac{1}{n-1} \sum_{i=1}^{n-1} f(x_i, x_{i+1}) > c \right. \right\}, \quad c > 0. \quad (2)$$

Write

$$\alpha_n \equiv P_0(A_n) = \sum_{\mathbf{x}^n \in A_n} P_0(\mathbf{x}^n)$$

where

$$P_0(\mathbf{x}^n) = p_0(x_1)P_0(x_2|x_1) \dots P_0(x_n|x_{n-1}).$$

Manuscript received December 28, 1998.

The author is with Nagaoka University of Technology, Nagaoka, Niigata 940-2188 Japan (e-mail: nakagawa@nagaokaut.ac.jp).

Communicated by R. Cruz, Associate Editor for Communication Networks.

Publisher Item Identifier S 0018-9448(01)00580-6.