Improved Gilbert-Varshamov Bound for Constrained Systems

Brian H. Marcus* Ron M. Roth†

December 10, 1991

Abstract

Nonconstructive existence results are obtained for block error-correcting codes whose codewords lie in a given constrained system. Each such system is defined as a set of words obtained by reading the labels of a finite directed labelled graph. For a prescribed constrained system and relative minimum distance δ, the new lower bounds on the rate of such codes improve on those derived recently by Kolesnik and Krachkovsky. The better bounds are achieved by considering a special subclass of sequences in the constrained system, namely, those having certain empirical statistics determined by δ.

Key words: Constrained systems; Gilbert-Varshamov bound; Markov chains; Run-length-limited codes; Sofic systems.

---

*IBM Research Division, Almaden Research Center K65/802, 650 Harry Road, San Jose, CA 95120.

†Computer Science Department, Technion — Israel Institute of Technology, Haifa 32000, Israel. This work was done while the author was with IBM Research Division, Almaden Research Center, San Jose, California.

1 Introduction

Let Σ be an alphabet of size q and let Σ^n denote the set of all words of length n over Σ. For a word $\mathbf{w} \in \Sigma^n$, let $\mathcal{B}_{\Sigma^n}(\mathbf{w}; r)$ denote the Hamming sphere in Σ^n of radius r centered at $\mathbf{w}$. The volume (size) $\mathcal{V}_{\Sigma^n}(r)$ of $\mathcal{B}_{\Sigma^n}(\mathbf{w}; r)$ is equal to $\sum_{i=0}^{r} \binom{n}{i}(q-1)^i$ and is independent of $\mathbf{w}$. By a (Σ^n, M, d)-code we mean a subset of Σ^n of size M and minimum distance $\geq d$. The classical Gilbert-Varshamov bound asserts the existence of a (Σ^n, M, d) code whenever $M - 1 < q^n / \mathcal{V}_{\Sigma^n}(d-1)$. This is proven simply by taking any word $\mathbf{w}_1 \in \Sigma^n$ as the first codeword in C, and then picking up iteratively the l-th codeword $\mathbf{w}_l$ out of $\Sigma^n - \bigcup_{i=1}^{l-1} \mathcal{B}_{\Sigma^n}(\mathbf{w}_i; d-1)$, $l = 2, 3, \ldots, M$ (see, for instance, [4, pp. 321–322]). If we continue this process until Σ^n is exhausted, we end up with a (Σ^n, M, d)-code of rate

$$R \triangleq \frac{\log_q M}{n} \;\geq\; 1 - \frac{\log_q \mathcal{V}_{\Sigma^n}(d-1)}{n} \;\geq\; 1 - H_q(\delta) \;,$$

where $\delta \triangleq d/n$ is the relative minimum distance of the code and $H_q(\delta) \triangleq -\delta \cdot \log_q \delta - (1 - \delta) \cdot \log_q (1-\delta) + \delta \cdot \log_q (q-1)$, $0 \leq \delta \leq 1 - \frac{1}{q}$.

The situation becomes quite different when, in addition, we require the codewords of C to lie in a given constrained system — most notable of which are run-length-limited systems and charge constrained systems. A constrained system S is defined as a set of words over Σ obtained by reading the labels of a finite directed graph. And counting only words in $S_n \triangleq \Sigma^n \cap S$ (i.e., words of length n in S), the volumes of constrained spheres $\mathcal{B}_{\Sigma^n}(\mathbf{w}; r) \cap S_n$ do usually vary with their center $\mathbf{w}$.

Recently, Kolesnik and Krachkovsky [8] generalized the above Gilbert-Varshamov bound to constrained systems by estimating the *average* volume of such constrained spheres, with their centers ranging over the whole space S_n. They then used a generating function for the distribution of pairwise distances of words in the constrained system, together with the Perron-Frobenius Theorem, in order to obtain asymptotic existence results relating the attainable code rate R to a prescribed relative minimum distance δ when $n \to \infty$. For related work see also [1][3][14].

In this paper, we sharpen the lower bounds on the attainable rates for given δ and constrained system S. Our improvement is based on considering subsets T of S_n, and averaging the volumes of the more severely constrained spheres, $\mathcal{B}_{\Sigma^n}(\mathbf{w}; r) \cap T$, over $\mathbf{w} \in T$.

In particular, we consider subsets T of all words of length n in S whose empirical statistics satisfy a given requirement: the frequency with which any symbol of Σ appears in the word is restricted to lie within a certain range which depends on the designed relative minimum distance δ. The asymptotic bounds on the attainable rates obtained by this approach are almost as easy to compute as in the unrestricted case (i.e., where the whole center space is considered); the computation is carried out by classical convex duality based on Lagrange multipliers. In particular, we get a slightly different proof of the main result of [8]; and moreover, our bounds significantly improve those obtained in the unrestricted case.

Our approach has been motivated by experience in trying to compute the capacity of the binary symmetric channel with input constraints specified by a constrained system (see, for instance [15]). In particular, for such channels, the input probability distributions which yield high mutual information tend to be far away from the most uniform distribution on the constrained system i.e., the one which is possessed by the 'typical sequences' of the constrained system. Thus, it seemed reasonable to guess that 'good' codes (i.e., large codes for a prescribed relative minimum distance δ) should be obtained by restricting the codewords to have statistics that resemble probability distributions which yield high mutual information for the corresponding input-constrained binary symmetric channel i.e., the one with crossover probability $< \delta/2$ (observe that virtually all error patterns of such a channel can be corrected by an error-correcting code of relative minimum distance δ).

Finally, we mention that the main result of [8] makes use of *upper* bounds on the growth rate of the sizes of certain word sets; we need to use upper bounds for more elaborate sets as well as *lower* bounds for related sets.

2 Constrained systems and constrained codes

In the sequel we will make use of a few terms which are related to both aspects of the codes considered in this paper — i.e., the fact that they lie in a given constrained system and the fact that they satisfy certain distance requirements. We start with a precise definition of constrained systems.

A *labelled graph* (or simply a *graph*) $G = (V, E, L)$ is a finite directed graph with states

$V = V_G$, edges $E = E_G$, and labelling $L = L_G : E \to \Sigma$ for some finite alphabet Σ. The set of edges from state u to state v is denoted by $E_G(u, v)$, and the set of outgoing (incoming) edges from (to) state u is denoted by $E_G^+(u)$ $(E_G^-(u))$. The initial state of an edge e in G is denoted by $\imath(e)$.

A *constrained system* is the set of all words (i.e., finite sequences) obtained by reading the labels of paths in a graph G (*path* means finite path). We call the set of words $S = S(G)$, and we say that G *presents* S. The set of all words in S of length n is denoted by S_n. The *capacity* of S is defined as the growth rate of the number of words in S i.e., $\lim_{n\to\infty} \frac{1}{n} \log |S_n|$, with $|S_n|$ standing for the size of S_n (hereafter all logarithms and exponents are taken to base $q = |\Sigma|$). Every constrained system can be presented by a graph which is *deterministic*, that is, for each state, the outgoing edges are labelled distinctly.

Constrained systems are essentially the same as regular languages in automata theory, or sofic systems in symbolic dynamics (except that in the latter case we consider bi-infinite sequences obtained from a graph, rather than finite ones).

A graph G is *irreducible* (or *strongly connected*) if for every pair of states $u, v \in V_G$ there is a path in G from u to v. An *irreducible component* of a graph G is a maximal irreducible subgraph of G (maximal with respect to inclusion). Every deterministic graph G has a (deterministic) irreducible component G' such that the capacity of the constrained system presented by G' is the same as that presented by G. Therefore, the study of general constrained systems reduces fairly easily to the study of constrained systems that are presented by deterministic irreducible graphs.

A *cycle* in a graph G is a path that starts and terminates at the same state. The greatest common divisor of all cycle lengths in an irreducible graph G is called the *period* of G and is denoted $\pi(G)$. An irreducible graph G of period 1 is called *primitive*. It is easy to verify that a graph G is primitive if and only if there is an integer N_G such that for any two states $u, v \in V_G$ there is a path of length N_G in G from u to v. We can reduce the general deterministic irreducible case to the deterministic primitive case by looking at the graph $H = G^{\pi(G)} = (V_H, E_H, L_H)$, where $V_H = V_G$, and each edge $\overline{e}$ from state u to state v in H corresponds to a path of length $\pi(G)$ from state u to state v in G. Letting $e_1 e_2 \ldots e_{\pi(G)}$ denote the edges in G along that path, the labelling $L_H(\overline{e})$ over the alphabet $\Sigma^{\pi(G)}$ is set to $L_G(e_1)L_G(e_2)\ldots L_G(e_{\pi(G)})$. That is, H generates all words of S of lengths which are multiples of $\pi(G)$. The irreducible components of H are all primitive and H is the disjoint union of these components. Because of the way we have chosen the base for the logarithm, the capacity of $S(H)$ is equal to that of $S(G)$ which, in turn, is equal to the capacity of a constrained system presented by any one of the (primitive) irreducible components of H.

Therefore, from now on we assume that our graphs are deterministic and primitive. By a *primitive constrained system* we mean a constrained system which is presented by a deterministic primitive graph.

For $\mathbf{w}, \mathbf{w}' \in \Sigma^n$, denote by $\Delta(\mathbf{w}, \mathbf{w}')$ the Hamming distance between $\mathbf{w}$ and $\mathbf{w}'$. Given a subset $T \subseteq \Sigma^n$ and $\mathbf{w} \in T$, denote by $\mathcal{B}_T(\mathbf{w}; r)$ the Hamming sphere of radius r in T centered at $\mathbf{w}$ i.e.,

$$\mathcal{B}_T(\mathbf{w}; r) \;\triangleq\; \mathcal{B}_{\Sigma^n}(\mathbf{w}; r) \cap T \;=\; \Big\{\mathbf{w}' \in T \;:\; \Delta(\mathbf{w}, \mathbf{w}') \le r\Big\} \,,$$

and let $\mathcal{V}_T(r)$ denote the average volume of spheres of radius r in T, namely,

$$\mathcal{V}_T(r) \;\triangleq\; \frac{1}{|T|} \sum_{\mathbf{w} \in T} |\mathcal{B}_T(\mathbf{w}; r)| \,.$$

In the sequel we will also make of use of the set $\mathcal{B}_T(r) \subseteq T \times T$ of all r-close pairs of words in T, that is,

$$\mathcal{B}_T(r) \;\triangleq\; \Big\{(\mathbf{w}, \mathbf{w}') \in T \times T \;:\; \Delta(\mathbf{w}, \mathbf{w}') \le r\Big\} \,.$$

Clearly,

$$\mathcal{B}_T(r) \;=\; \Big\{(\mathbf{w}, \mathbf{w}') \;:\; \mathbf{w} \in T \;\text{ and }\; \mathbf{w}' \in \mathcal{B}_T(\mathbf{w}; r)\Big\}$$

and, therefore, $|\mathcal{B}_T(r)| = \mathcal{V}_T(r) \cdot |T|$.

Given a subset T of Σ^n, a (T, M, d)*-code* is a (Σ^n, M, d)-code which is also a subset of T. In this work we will be interested in the special case $T = S_n$. The lower bounds obtained in [8] on the attainable rates of (S_n, M, d)-codes with prescribed minimum distance d are based on the following result which is essentially [8, Lemma 2.1]. We give a proof for completeness.

Lemma 1. (Kolesnik and Krachkovsky [8]). *Let T be a subset of Σ^n and let d be a positive integer. Then there exists a (T, M, d)-code such that*

$$M \;\ge\; \frac{|T|}{4\,\mathcal{V}_T(d-1)} \;=\; \frac{|T|^2}{4|\mathcal{B}_T(d-1)|} \,.$$

Proof. Let T' denote the set of all words $\mathbf{w} \in T$ such that $|\mathcal{B}_T(\mathbf{w}; d-1)| \le 2\,\mathcal{V}_T(d-1)$. It is easy to check that the size of T' must be at least half the size of T, or else the average volume of the spheres of radius $d-1$ in T would be more than $\mathcal{V}_T(d-1)$. Now, define the code C as in the unconstrained case by first selecting an arbitrary element $\mathbf{w}_1$ of T', and then picking up iteratively the l-th codeword $\mathbf{w}_l$ out of $T' - \bigcup_{i=1}^{l-1} \mathcal{B}_{T'}(\mathbf{w}_i; d-1)$, $l = 2, 3, \ldots$, until T' is exhausted. This yields a code C which is a subset of T and clearly has minimum distance at least d. Furthermore,

$$M \;\ge\; \frac{|T'|}{2\,\mathcal{V}_T(d-1)} \;\ge\; \frac{\frac{1}{2}|T|}{2\,\mathcal{V}_T(d-1)} \;=\; \frac{|T|^2}{4|\mathcal{B}_T(d-1)|}\;,$$

as claimed. $\square$

The lower bound in [8] was obtained by setting $T = S_n$; that bound was then used to derive lower bounds on the achievable asymptotic rates of constrained codes with prescribed relative minimum distance. Our bounds are obtained by considering more restrictive subsets T of S_n so as to obtain larger ratios $\frac{1}{4}|T|^2/|\mathcal{B}_T(d-1)|$. The computation of the bounds requires consideration of stationary Markov chains which are discussed in the next section.

3 Markov chains

A *stationary Markov chain* on a graph G is a function $P : E_G \to [0, 1]$ such that

- $\sum_{e \in E_G} P(e) = 1$; that is, P is a probability distribution on the edges of G;
- $\sum_{e \in E_G^+(u)} P(e) = \sum_{e \in E_G^-(u)} P(e)$ for every $u \in V_G$ (the stationary condition); that is, the sum of the probabilities of the outgoing edges equals the sum of the probabilities of the incoming edges for any state in G.

We denote by $\mathcal{M}(G)$ the set of all stationary Markov chains on G.

For a stationary Markov chain $P \in \mathcal{M}(G)$ and a function $f : E_G \to \Re^k$, we denote by $\mathbf{E}_P(f)$ the expected value of f with respect to P; that is,

$$\mathbf{E}_P(f) \;=\; \sum_{e \in E_G} P(e) f(e)\;.$$

Let $P \in \mathcal{M}(G)$ and, for each $e \in E_G$, define the *conditional probability* $P(e \,|\, \imath(e))$ by

$$P(e \,|\, \imath(e)) \;\triangleq\; \frac{P(e)}{\sum_{e' \in E_G^+(\imath(e))} P(e')} \; .$$

The *entropy rate* of P is defined by

$$h(P) \;\triangleq\; \mathbf{E}_P\Big(-\log\, P(e \,|\, \imath(e))\Big) \;=\; -\sum_{u \in V_G}\,\sum_{e \in E_G^+(u)} P(e) \cdot \log\, P(e \,|\, u) \; .$$

Let $\Gamma_n(G)$ denote the set of all cycles in G of length n that start and terminate at some arbitrary fixed state u. For a cycle $\gamma = e_1 e_2 \ldots e_n \in \Gamma_n(G)$, let P_γ denote the stationary Markov chain defined by

$$P_\gamma(e) \;\triangleq\; \frac{1}{n} \left| \Big\{ i \in \{1, 2, \ldots, n\} \;:\; e_i = e \Big\} \right| \; . \tag{1}$$

We refer to P_γ as the *empirical distribution* of γ, and to

$$\mathbf{E}_{P_\gamma}(f) \;=\; \sum_{e \in E_G} P_\gamma(e) f(e)$$

as the *empirical average* of f on γ. The functions f used in the sequel are typically indicator functions of symbols, in which case $\mathbf{E}_{P_\gamma}(f)$ becomes the frequency with which a particular symbol appears in the labelling of the cycle γ.

For a subset $U \subseteq \Re^k$, let $\mathcal{M}(G; f, U)$ denote the set of all stationary Markov chains on G such that $\mathbf{E}_P(f) \in U$. Also, let $\Gamma_n(G; f, U)$ be defined by

$$\Gamma_n(G; f, U) \;\triangleq\; \Big\{ \gamma \in \Gamma_n(G) \;:\; \mathbf{E}_{P_\gamma}(f) \in U \Big\} \; .$$

The following is a consequence of well-known results and is central to our lower bounds (this is essentially contained in [2, Lemma 2]; see also [7][11]).

Lemma 2. *Let G be a primitive graph and $f : E_G \to \Re^k$ be a function on the edges of G. Let U be an open subset of $\Re^k$ (e.g., U is an open rectangular parallelepiped $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_k, b_k)$). Then,*

$$\lim_{n \to \infty} \frac{1}{n} \log\, |\Gamma_n(G; f, U)| \;=\; \sup_{P \in \mathcal{M}(G; f, U)} h(P) \; . \tag{2}$$

Proof. First, we show that the left-hand side of (2) is not greater than its right-hand side. The proof is based on the known fact (see Appendix A) that for any stationary Markov chain P, the number of cycles $\gamma \in \Gamma_n(G)$ of empirical distribution $P_\gamma = P$ is at most $q^{n\,h(P)}$, where $q = |\Sigma|$.

Let Q_n be a stationary Markov chain which is the empirical distribution of a maximal number of cycles in $\Gamma_n(G; f, U)$; this maximal number of cycles, in turn, is at most $q^{n\,h(Q_n)}$. Now, the number of distinct empirical distributions of cycles in $\Gamma_n(G)$ — and therefore of cycles in $\Gamma_n(G; f, U)$ — is bounded from above by $(n+1)^{|E_G|} = q^{n\cdot o(1)}$, where $o(1)$ stands for an expression which tends to zero as $n \to \infty$ (with G held fixed). Hence, the size of $\Gamma_n(G; f, U)$ is at most $q^{n\,(h(Q_n)+o(1))}$ and the desired inequality follows from the fact that $Q_n \in \mathcal{M}(G; f, U)$ for every n.

We now turn to showing the inequality in the other direction. Here we base our proof on the fact that, for any stationary Markov chain P, there are at least $q^{n\,(h(P)-o(1))}$ cycles $\gamma \in \Gamma_n(G)$ of empirical distribution P_γ such that $\sum_{e \in E_G} |P_\gamma(e) - P(e)| = o(1)$ (these are the so-called 'typical cycles' with respect to P). We give a proof of this fact, which is a well-known consequence of the law of large numbers, in Appendix A.

Fix some $\epsilon > 0$ and let $Q_0 \in \mathcal{M}(G; f, U)$ be such that $h(Q_0) \geq \sup_{P \in \mathcal{M}(G;f,U)} h(P) - \epsilon$. Since U is an open subset of $\Re^k$, there is some $\rho = \rho(Q_0) > 0$ such that the whole ρ-neighborhood of $\mathbf{E}_{Q_0}(f)$ is contained in U. By the continuity of the mapping $Q \mapsto \mathbf{E}_Q(f)$ it follows that there are $q^{n\,(h(Q_0)-o(1))}$ cycles in $\Gamma_n(G)$ with empirical average within $o(1)$-neighborhood of $\mathbf{E}_{Q_0}(f)$ and, as such, these cycles are contained in $\Gamma_n(G; f, U)$ for n greater than some $N(Q_0)$. The lemma now follows by letting ϵ go to zero. $\square$

4 Asymptotic lower bounds

Let $S = S(G)$ be a primitive constrained system over Σ presented by the primitive graph G. For a symbol $w \in \Sigma$, define the indicator function $\mathcal{I}_w : E_G \to \Re$ of w by

$$\mathcal{I}_w(e) \triangleq \begin{cases} 1 & \text{if } L_G(e) = w \\ 0 & \text{otherwise} \end{cases} .$$

Similarly, for a subset $W = \{w_1, w_2, \ldots, w_t\}$ of Σ of size t, define the function $\mathcal{I}_W : E_G \to \Re^t$ by $\mathcal{I}_W(e) \triangleq \Big[\mathcal{I}_{w_1}(e),\, \mathcal{I}_{w_2}(e),\, \ldots,\, \mathcal{I}_{w_t}(e)\Big]$.

Given a subset $W \subseteq \Sigma$ of size t, let U be a t-dimensional open rectangular parallelepiped $\prod_{i=1}^t (a_i, b_i)$ in the unit hypercube $[0,1]^t$ and define the subset $S_{W;n}(U)$ of S_n by

$$S_{W;n}(U) \triangleq \Big\{L_G(e_1)L_G(e_2)\ldots L_G(e_n) \in S_n \;:\; e_1e_2\ldots e_n \in \Gamma_n(G;\mathcal{I}_W,\, U)\Big\} .$$

In other words, $S_{W;n}(U)$ consists of all words $\mathbf{w}$ in S_n which are obtained by reading the labelling of cycles in $\Gamma_n(G)$ and for which the frequency with which the symbol $w_i \in W$ appears in $\mathbf{w}$ lies in (a_i, b_i) for $i = 1, 2, \ldots, t$. Note that since G is deterministic, each word in $S_{W;n}(U)$ is generated by exactly one cycle in $\Gamma_n(G;\mathcal{I}_W,\, U)$.

Our improvement on the attainable rates of constrained codes with prescribed relative minimum distance δ will be achieved by considering subsets of the form $T = S_{W;n}(U)$ in Lemma 1 for sufficiently large n.

Let $G \times G$ be the graph defined by $V_{G\times G} = V_G \times V_G = \Big\{\langle u, u'\rangle \;:\; u, u' \in V_G\Big\}$ and $E_{G\times G} = E_G \times E_G$; that is, if e is an edge in G from state u to state v, and e' is an edge in G from state u' to state v', then $G \times G$ has an edge $\langle e, e'\rangle$ from state $\langle u, u'\rangle$ to state $\langle v, v'\rangle$ with labelling $\Big\langle L_G(e), L_G(e')\Big\rangle$. Note that $G \times G$ is primitive and deterministic whenever G is.

Given the function $\mathcal{I}_W$ defined on the edges of G, we define the functions $\mathcal{I}_W^{(1)}$ and $\mathcal{I}_W^{(2)}$ on $E_{G\times G}$ by $\mathcal{I}_W^{(1)}\Big(\langle e, e'\rangle\Big) = \mathcal{I}_W(e)$ and $\mathcal{I}_W^{(2)}\Big(\langle e, e'\rangle\Big) = \mathcal{I}_W(e')$. Also, define the mapping $\mathcal{D} : E_{G\times G} \to \Re$ by

$$\mathcal{D}\Big(\langle e, e'\rangle\Big) \triangleq \Delta\Big(L_G(e), L_G(e')\Big) ;$$

that is, $\mathcal{D}\Big(\langle e, e'\rangle\Big)$ is 1 when $L_G(e) = L_G(e')$, and zero elsewhere. We now put $\mathcal{I}_W^{(1)}$, $\mathcal{I}_W^{(2)}$, and $\mathcal{D}$ all together into a function

$$\mathcal{T}_W \triangleq \Big[\mathcal{I}_W^{(1)}, \mathcal{I}_W^{(2)}, \mathcal{D}\Big] \;:\; E_{G\times G} \to \Re^t \times \Re^t \times \Re .$$

For $W \subseteq \Sigma$ and $U \subseteq \Re^t$ set

$$\mathcal{F}_W(U) \triangleq \sup_{P \in \mathcal{M}(G;\mathcal{I}_W, U)} h(P)$$

and

$$\mathcal{G}_W(U, \delta) \triangleq \sup_{Q \in \mathcal{M}(G\times G;\, \mathcal{T}_W,\, U\times U\times[0,\delta)\,)} h(Q) .$$

Lemma 3. *Given a primitive constrained system S over Σ and $\delta > 0$, let W be a subset of Σ of size t and let $U = \prod_{i=1}^{t}[a_i, b_i]$ be a closed rectangular parallelepiped (possibly of volume zero) in the unit hypercube $[0,1]^t$. Then there exist $(S_n, M, \delta n)$-codes satisfying*

$$\frac{\log M}{n} \;\geq\; 2\mathcal{F}_W(U) - \mathcal{G}_W(U,\delta) - o(1) \; .$$

Proof. Write $d_n \triangleq \lceil \delta n \rceil$ and, for $\epsilon > 0$, let U_ϵ denote the open rectangular parallelepiped $\prod_{i=1}^{t}(a_i - \epsilon, b_i + \epsilon)$. By Lemma 1, there exists an (S_n, M, d_n)-code satisfying

$$\frac{\log M}{n} \;\geq\; 2\,\frac{\log|S_{W;n}(U_\epsilon)|}{n} \;-\; \frac{\log|\mathcal{B}_{S_{W;n}(U_\epsilon)}(d_n - 1)|}{n} \;-\; o(1) \; . \tag{3}$$

Noting that $|S_{W;n}(U_\epsilon)| = |\Gamma_n(G;\, \mathcal{I}_W,\, U_\epsilon)|$, by Lemma 2 we thus have

$$\lim_{n\to\infty} \frac{1}{n} \log|S_{W;n}(U_\epsilon)| \;=\; \mathcal{F}_W(U_\epsilon) \; . \tag{4}$$

For the same initial state u that was used to define $\Gamma_n(G)$, we now define $\Gamma_n(G \times G)$ as the set of all cycles in $G \times G$ (i.e., the set of all cycle pairs of G) that start and terminate at $\langle u, u\rangle$. It can be easily verified that the set $\mathcal{B}_{S_{W;n}(U_\epsilon)}(d_n - 1)$ consists of all word pairs obtained by reading the labelling of the cycles in $\Gamma_n(G \times G;\, \mathcal{T}_W,\, U_\epsilon \times U_\epsilon \times [0,\delta)\,)$. Hence,

$$\Big|\mathcal{B}_{S_{W;n}(U_\epsilon)}(d_n - 1)\Big| \;=\; \Big|\Gamma_n(G \times G;\, \mathcal{T}_W,\, U_\epsilon \times U_\epsilon \times (-\epsilon,\delta)\,)\Big|$$

and, therefore, by Lemma 2 we have

$$\lim_{n\to\infty} \frac{1}{n} \log\Big|\mathcal{B}_{S_{W;n}(U_\epsilon)}(d_n - 1)\Big| \;=\; \mathcal{G}_W(U_\epsilon, \delta) \; . \tag{5}$$

Now, by the continuity of the functions $P \mapsto h(P)$ and $P \mapsto \mathbf{E}_P(f)$ we have both

$$\lim_{\epsilon\to 0} \mathcal{F}_W(U_\epsilon) \;=\; \mathcal{F}_W(U) \qquad \text{and} \qquad \lim_{\epsilon\to 0} \mathcal{G}_W(U_\epsilon,\delta) \;=\; \mathcal{G}_W(U,\delta) \; . \tag{6}$$

The lemma now follows by combining (3), (4), (5), and (6). $\square$

In particular, Lemma 3 holds for the case where U is the trivial parallelepiped $\{\mathbf{p}\}$ for some $\mathbf{p} \in [0,1]^t$. In fact, in the next lemma we show that the lower bound in Lemma 3 is maximized at such trivial parallelepipeds.

Hereafter we use $\mathcal{F}_W(\mathbf{p})$ and $\mathcal{G}_W(\mathbf{p},\delta)$ as short-hand notations for $\mathcal{F}_W(\{\mathbf{p}\})$ and $\mathcal{G}_W(\{\mathbf{p}\},\delta)$, respectively.

Lemma 4. *Let S be a primitive constrained system over Σ, let $\delta > 0$, and let W be a subset of Σ of size t. Denote by Υ_t the set of all closed rectangular parallelepipeds in $[0,1]^t$. Then,*

$$\sup_{U \in \Upsilon_t} \Big\{ 2\,\mathcal{F}_W(U) - \mathcal{G}_W(U,\delta) \Big\} \;=\; \sup_{\mathbf{p} \in [0,1]^t} \Big\{ 2\,\mathcal{F}_W(\mathbf{p}) - \mathcal{G}_W(\mathbf{p},\delta) \Big\} \,.$$

Proof. The inequality

$$\sup_U \Big\{ 2\,\mathcal{F}_W(U) - \mathcal{G}_W(U,\delta) \Big\} \;\ge\; \sup_{\mathbf{p} \in [0,1]^t} \Big\{ 2\,\mathcal{F}_W(\mathbf{p}) - \mathcal{G}_W(\mathbf{p},\delta) \Big\}$$

follows from the fact that the trivial parallelepipeds $\{\mathbf{p}\}$ are members of Υ_t. The opposite inequality follows from continuity arguments: Fix some $\epsilon > 0$ and choose $U_0 \in \Upsilon_t$ so that $2\,\mathcal{F}_W(U_0) - \mathcal{G}_W(U_0,\delta)$ is within ϵ of $\sup_U \Big\{ 2\,\mathcal{F}_W(U) - \mathcal{G}_W(U,\delta) \Big\}$. Then choose a stationary Markov chain $P \in \mathcal{M}(G;\, \mathcal{I}_W,\, U_0)$ such that $h(P)$ is within ϵ of $\mathcal{F}_W(U_0)$. Let $\mathbf{p} = \mathbf{E}_P(\mathcal{I}_W) \in U_0$. Then $\mathcal{F}_W(\mathbf{p})$ is within ϵ of $\mathcal{F}_W(U_0)$ and $\mathcal{G}_W(\mathbf{p},\delta) \le \mathcal{G}_W(U_0,\delta)$. Thus, $2\,\mathcal{F}_W(\mathbf{p}) - \mathcal{G}_W(\mathbf{p},\delta)$ is at least $2\,\mathcal{F}_W(U_0) - \mathcal{G}_W(U_0,\delta) - 2\epsilon$, which in turn is at least $\sup_U \Big\{ 2\,\mathcal{F}_W(U) - \mathcal{G}_W(U,\delta) \Big\} - 3\epsilon$. Therefore,

$$\sup_U \Big\{ 2\,\mathcal{F}_W(U) - \mathcal{G}_W(U,\delta) \Big\} \;\le\; \sup_{\mathbf{p} \in [0,1]^t} \Big\{ 2\,\mathcal{F}_W(\mathbf{p}) - \mathcal{G}_W(\mathbf{p},\delta) \Big\} \,. \qquad \square$$

Combining the preceding result with Lemma 3, our best bounds are obtained by choosing codewords all with approximately the same frequency of appearance of symbols of a given set W. We summarize this in the next theorem.

Define

$$\begin{aligned} R_W(\delta) \;&\triangleq\; \sup_{\mathbf{p} \in [0,1]^t} \Big\{ 2\,\mathcal{F}_W(\mathbf{p}) - \mathcal{G}_W(\mathbf{p},\delta) \Big\} \\ &=\; \sup_{\mathbf{p} \in [0,1]^t} \left\{ 2 \sup_{\substack{P \in \mathcal{M}(G)\,: \\ \mathbf{E}_P(\mathcal{I}_W) = \mathbf{p}}} h(P) \;-\; \sup_{\substack{Q \in \mathcal{M}(G \times G)\,: \\ \mathbf{E}_Q(\mathcal{I}_W^{(i)}) = \mathbf{p}\,,\; i=1,2 \\ \mathbf{E}_Q(\mathcal{D}) \in [0,\delta)}} h(Q) \right\} . \end{aligned}$$

Theorem 1. *Let $S = S(G)$ be a primitive constrained system over Σ, let $\delta > 0$, and let W be a subset of Σ of size t. Then there exist $(S_n, M, \delta n)$-codes satisfying*

$$\frac{\log M}{n} \;\ge\; R_W(\delta) \;-\; o(1) \,.$$

The finest version of Theorem 1 is obtained by setting $W = \Sigma$. In fact, one can get even better lower bounds by considering the frequency of appearance of words $\mathbf{w} \in S$ of arbitrary length, rather than just symbols of Σ. This is done as follows.

Given a graph G that presents a constrained system S, the *ℓ-th order path graph* $G^{(\ell)}$ is the graph whose states are pairs $(w_1 w_2 \ldots w_\ell; u) \in S_\ell \times V_G$, where u ranges over all states in G and $w_1 w_2 \ldots w_\ell$ ranges over all words that can be generated by paths in G that terminate at u. We put an edge $\overline{e}$ from state $\overline{u} = (w_1 w_2 \ldots w_\ell; u)$ to state $\overline{u}' = (w'_1 w'_2 \ldots w'_\ell; u')$ in $G^{(\ell)}$ if $w_2 w_3 \ldots w_\ell = w'_1 w'_2 \ldots w'_{\ell-1}$ and there is an edge e from state u to state u' in G with $L_G(e) = w'_\ell$. We also define $L_{G^{(\ell)}}(\overline{e}) \triangleq L_G(e) = w'_\ell$. Note that $S(G) = S(G^{(\ell)})$ and that $G^{(\ell)}$ is deterministic and primitive whenever G is.

Example 1. Consider the graph G over $\Sigma = \{0, 1\}$ of Figure 1. The graph G presents the $(0,1)$-run-length-limited (RLL) system i.e., the set of binary words whose runs of zeros are of length 1 at most.

G :

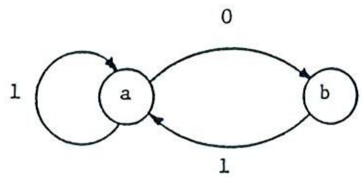

Figure 1: $(0,1)$-RLL system.

The path graphs $G^{(1)}$ and $G^{(2)}$ are depicted in Figures 2 and 3, respectively.

$G^{(1)}$:

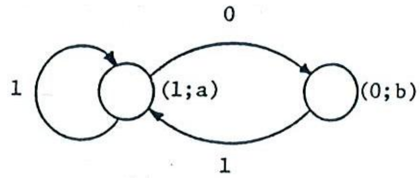

Figure 2: Graph $G^{(1)}$ for the $(0,1)$-RLL system.

$G^{(2)}$:

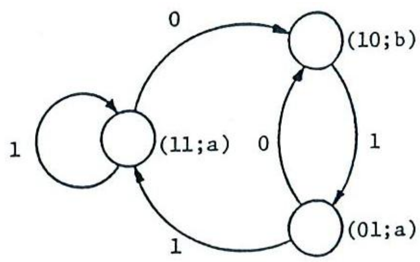

Figure 3: Graph $G^{(2)}$ for the $(0,1)$-RLL system.

Note that in this particular example, the graph $G^{(1)}$ is essentially the same as $G^{(0)} \triangleq G$. •

Let $\mathbf{w}$ be a word of length $\ell+1$ in $S(G)$ and let $\mathcal{I}_{\mathbf{w}}$ be the function on the edges of $G^{(\ell)}$ defined by

$$\mathcal{I}_{\mathbf{w}}(\overline{e}) \triangleq \begin{cases} 1 & \text{if } \imath(\overline{e}) = (w'_1 w'_2 \ldots w'_\ell; u) \text{ and } w'_1 w'_2 \ldots w'_\ell L_{G^{(\ell)}}(\overline{e}) = \mathbf{w} \\ 0 & \text{otherwise} \end{cases} .$$

The mapping $\mathcal{D} : E_{G^{(\ell)} \times G^{(\ell)}} \to \Re$ is defined for the path graph the same way as for G i.e.,

$$\mathcal{D}\Big(\langle \overline{e}, \overline{e}' \rangle\Big) \triangleq \Delta\Big(L_{G^{(\ell)}}(\overline{e}), L_{G^{(\ell)}}(\overline{e}')\Big) .$$

The even-better bounds are now obtained by applying Theorem 1 with the above mapping definitions for a set $W \subseteq S_{\ell+1}$ on the path graph $G^{(\ell)}$.

Remark 1. The codes guaranteed by Theorem 1 are *repeatable* i.e., all concatenations of codewords lie in the primitive constrained system S. This is because the codewords are generated by cycles passing through a fixed state. In case the given constrained system is nonprimitive (e.g., charge-constrained systems, where the period is 2), Theorem 1 can be used to obtain repeatable codes of lengths n which are multiples of the period $\pi(G)$; namely, we apply the theorem on a constrained system S' which is presented by an irreducible component G' of $G^{\pi(G)}$ which has the same capacity as that of S. Note that S' is, in effect, a constrained system over the alphabet $\Sigma^{\pi(G)}$; however, since we are interested in minimum distances over the base alphabet Σ, we need to re-define the function $\mathcal{D}$ as

$$\mathcal{D}\Big(\langle e, e' \rangle\Big) \triangleq \tfrac{1}{\pi(G)} \Delta\Big(L_G(e), L_G(e')\Big) ,$$

where the labels $L_G(\cdot)$ are regarded as $\pi(G)$-tuples over Σ and $\Delta(\cdot,\cdot)$ stands for Hamming distance over Σ. •

5 Dual formula

In order to compute the bound given by Theorem 1, we need to compute the functions $\mathcal{F}_W(\mathbf{p})$ and $\mathcal{G}_W(\mathbf{p}, \delta)$. Each of these quantities is computed as a constrained optimization problem with concave objective function, $P \mapsto h(P)$, and linear constraints. In this section, we show how, using convex duality based on Lagrange multipliers (see [9][13]), these quantities can be computed as an essentially unconstrained optimization problem with convex objective function.

Let $\mathbf{0}$ denote the all-zero vector. For vectors $\mathbf{x} = [x_i]_i$ and $\mathbf{y} = [y_i]_i$ in $\Re^k$, $\mathbf{x} \le \mathbf{y}$ means $x_i \le y_i$ for $i = 1, 2, \ldots, k$, and $(\Re^+)^k$ stands for the set $\{\mathbf{y} \in \Re^k \,:\, \mathbf{0} \le \mathbf{y}\}$.

For a function $f : E_G \to \Re^k$, let $A_{G;f}(\mathbf{x}), \mathbf{x} \in \Re^k$, be the matrix function indexed by the states of G with entries

$$\Big[\, A_{G;f}(\mathbf{x}) \,\Big]_{u,v} \;\triangleq\; \sum_{e \in E_G(u,v)} q^{-\mathbf{x} \cdot f(e)}$$

and let $\lambda_{G;f}(\mathbf{x}) = \lambda_f(\mathbf{x})$ denote the spectral radius of $A_{G;f}(\mathbf{x})$. Note that $A_{G;f}(\mathbf{0})$ is simply the adjacency matrix of G for any function f.

The following is a consequence of well-known results on optimizing convex (concave) functions subject to linear equality and linear inequality constraints.

Lemma 5. *Let G be a graph and $f : E_G \to \Re^k$, $g : E_G \to \Re^\ell$ be functions on the edges of G. Write $\psi = [f, g] : E_G \to \Re^{k+\ell}$. Then, for any $\mathbf{r} \in \Re^k$ and $\mathbf{s} \in \Re^\ell$,*

$$\sup_{\substack{P \in \mathcal{M}(G)\,: \\ \mathbf{E}_P(f) = \mathbf{r} \\ \mathbf{E}_P(g) \le \mathbf{s}}} h(P) \;=\; \inf_{\substack{\mathbf{x} \in \Re^k \\ \mathbf{z} \in (\Re^+)^\ell}} \Big\{ \mathbf{x} \cdot \mathbf{r} + \mathbf{z} \cdot \mathbf{s} + \log\, \lambda_{G;\psi}(\mathbf{x}, \mathbf{z}) \Big\}\;.$$

Proof. The lemma follows from standard results in convex duality (see [9][13]). Define the *dual function*

$$\phi(\mathbf{x}, \mathbf{z}) \;\triangleq\; \sup_{P \in \mathcal{M}(G)} \Big\{ h(P) - \mathbf{x} \cdot \mathbf{E}_P(f) - \mathbf{z} \cdot \mathbf{E}_P(g) \Big\}\;.$$

It follows from [9, pp. 312–316][13, Section 28] that

$$\sup_{\substack{P \in \mathcal{M}(G)\,: \\ \mathbf{E}_P(f) = \mathbf{r} \\ \mathbf{E}_P(g) \le \mathbf{s}}} h(P) \;=\; \inf_{\substack{\mathbf{x} \in \Re^k \\ \mathbf{z} \in (\Re^+)^\ell}} \Big\{ \mathbf{x} \cdot \mathbf{r} + \mathbf{z} \cdot \mathbf{s} + \phi(\mathbf{x}, \mathbf{z}) \Big\}\;.$$

Thus, it remains only to see that

$$\phi(\mathbf{x}, \mathbf{z}) = \log\, \lambda_{G;\psi}(\mathbf{x}, \mathbf{z})\;.$$

A proof of this fact is contained in [12, Ch. II, Theorem 25]. For completeness, we have included a proof in Appendix B. $\square$

Remark 2. It is easy to verify directly that the dual function ϕ mentioned in the proof above is convex in its arguments. Thus, $\log \lambda_{G;\psi}(\mathbf{x}, \mathbf{z})$ is convex. In general, for any function f, $\log \lambda_{G;f}(\mathbf{x})$ is convex. $\bullet$

The special case in which $\psi = f$ (i.e., the constraints are equality constraints) is contained in [10], and the special case in which $\psi = g$ (i.e., the constraints are inequality constraints) is contained in [6].

We now apply Lemma 5 to the bound of Theorem 1. For the sake of simplicity, we consider a special case of the bound where W consists of only one symbol w. The forthcoming discussion generalizes also for any symbol set W, except that the number of variables involved in the optimization process becomes larger.

Recalling the definitions of $\mathcal{I}_w^{(i)} \triangleq \mathcal{I}_{\{w\}}^{(i)}$, $\mathcal{D}$, and $\mathcal{T}_w \triangleq \mathcal{T}_{\{w\}}$ from Section 4, define

$$\mathcal{J}_w \triangleq \Big[\mathcal{I}_w^{(1)} + \mathcal{I}_w^{(2)},\, \mathcal{D} \Big] \;:\; E_{G\times G} \to \Re^2 \;.$$

Also, write $\mathcal{F}_w(p) \triangleq \mathcal{F}_{\{w\}}\big([p]\big)$, $\mathcal{G}_w(p,\delta) \triangleq \mathcal{G}_{\{w\}}\big([p],\delta\big)$, and $R_w(\delta) \triangleq R_{\{w\}}(\delta)$. The next theorem gives a dual formula for the lower bound $R_w(\delta)$ on the attainable rates of $(S_n, M, \delta n)$-codes.

Theorem 2. *Let S be a primitive constrained system over Σ, let $\delta > 0$, and let $w \in \Sigma$. Then,*

(i) $\mathcal{F}_w(p) \;=\; \inf_{x\in\Re} \Big\{ p\,x + \log\, \lambda_{G;\mathcal{I}_w}(x) \Big\}$;

(ii) $\mathcal{G}_w(p,\delta) \;=\; \inf_{x\in\Re,\; z\in\Re^+} \Big\{ 2p\,x + \delta\, z + \log\, \lambda_{G\times G;\mathcal{J}_w}(x,z) \Big\}$;

(iii) $R_w(\delta) \;=\; \sup_{p\in[0,1]} \left\{ 2 \inf_{x\in\Re} \Big\{ p\,x + \log\, \lambda_{\mathcal{I}_w}(x) \Big\} - \inf_{x\in\Re,\; z\in\Re^+} \Big\{ 2p\,x + \delta\, z + \log\, \lambda_{\mathcal{J}_w}(x,z) \Big\} \right\}$.

Proof. First note that part (iii) is a consequence of parts (i) and (ii) and the definition of $R_w(\delta)$.

Part (i) follows by application of Lemma 5 to the function $\psi = \mathcal{I}_w$ with $\mathbf{r} = [p]$. As for part (ii), set $\psi = \mathcal{T}_w = \left[\mathcal{I}_w^{(1)},\, \mathcal{I}_w^{(2)},\, \mathcal{D}\right]$, $\mathbf{r} = [p,p]$, and $\mathbf{s} = [\delta]$ to obtain

$$\begin{aligned} \mathcal{G}_w(p,\delta) &= \inf_{x,y\in\Re,\; z\in\Re^+} \Big\{ p\,x + p\,y + \delta\,z + \log\,\lambda_{\mathcal{T}_w}(x,y,z) \Big\} \\ &\le \inf_{x\in\Re,\; z\in\Re^+} \Big\{ 2p\,x + \delta\,z + \log\,\lambda_{\mathcal{T}_w}(x,x,z) \Big\} \\ &= \inf_{x\in\Re, z\in\Re^+} \Big\{ 2p\,x + \delta\,z + \log\,\lambda_{\mathcal{J}_w}(x,z) \Big\}\,. \end{aligned} \tag{7}$$

(Note that at this point we have already established an inequality in part (iii) which yields a lower bound on $R_w(\delta)$.)

In order to complete the proof, it remains to show that

$$\inf_{\substack{x\in\Re\\ z\in\Re^+}} \Big\{ 2p\,x + \delta\,z + \log\,\lambda_{\mathcal{T}_w}(x,x,z) \Big\} \;\le\; \inf_{\substack{x,y\in\Re\\ z\in\Re^+}} \Big\{ p\,x + p\,y + \delta\,z + \log\,\lambda_{\mathcal{T}_w}(x,y,z) \Big\}\,. \tag{8}$$

For $\epsilon > 0$, let the infimum in the right-hand side of (8) be achieved, to within ϵ, at (x_0,y_0,z_0) and let $\overline{x} \triangleq (x_0{+}y_0)/2$. Since $\lambda_{\mathcal{T}_w}(x,y,z) = \lambda_{\mathcal{T}_w}(y,x,z)$ and $\log\,\lambda_{\mathcal{T}_w}$ is convex (see Remark 2), we have

$$\begin{aligned} 2p\,\overline{x} + \delta\,z_0 + \log\,\lambda_{\mathcal{T}_w}(\overline{x},\overline{x},z_0) &\le 2p\,\overline{x} + \delta\,z_0 + \log\,\lambda_{\mathcal{T}_w}(x_0,y_0,z_0) \\ &= p\,x_0 + p\,y_0 + \delta\,z_0 + \log\,\lambda_{\mathcal{T}_w}(x_0,y_0,z_0)\;, \end{aligned}$$

as desired. $\square$

Note that, as promised, for fixed p, this lower bound can be computed as essentially unconstrained optimization of convex functions in at most two variables.

Remark 3. In order to compute the function $\lambda_{\mathcal{J}_w}(x,z)$, we need to compute the spectral radius (the largest eigenvalue) of the $|V_G|^2 \times |V_G|^2$ matrix $A_{G\times G;\mathcal{J}_w}(x,z)$. We can apply the reduction technique used in [8] to obtain a smaller matrix, $A'_{G\times G;\mathcal{J}_w}(x,z)$, of order $\frac{1}{2}|V_G| \cdot \Big(|V_G|+1\Big)$, which has the same spectral radius as $A_{G\times G;\mathcal{J}_w}(x,z)$. The matrix $A'_{G\times G;\mathcal{J}_w}(x,z)$ is obtained in the manner described below.

Two states $\overline{u} = \langle u_1, u_2\rangle$ and $\overline{u}' = \langle u_1', u_2'\rangle$ in $G \times G$ are said to be *equivalent* if either $\overline{u} = \overline{u}'$, or $u_1 = u_2'$ and $u_2 = u_1'$. It is easy to verify that

$$\Big[A_{G\times G;\mathcal{J}_w}(x,z)\Big]_{\overline{u},\overline{v}} \;=\; \Big[A_{G\times G;\mathcal{J}_w}(x,z)\Big]_{\overline{u}',\overline{v}'}$$

whenever $\overline{u}$ and $\overline{u}'$ form an equivalent pair, and so do $\overline{v}$ and $\overline{v}'$. The matrix $A'_{G\times G;\mathcal{J}_w}$ is now obtained by iteratively picking a pair of distinct equivalent states $\overline{u}, \overline{u}' \in V_{G\times G}$, adding the column in $A_{G\times G;\mathcal{J}_w}$ which is indexed by $\overline{u}'$ to that which is indexed by $\overline{u}$, and then erasing the row in $A_{G\times G;\mathcal{J}_w}$ which corresponds to $\overline{u}'$. Doing this for each such pair of states in $G\times G$, we end up with a matrix $A'_{G\times G;\mathcal{J}_w}(x,z)$ of the claimed order and of spectral radius $\lambda_{\mathcal{J}_w}(x,z)$. We point out that in specific constrained systems, like the ones discussed in [8], the order of matrices involved may be reduced even further. Note that this procedure is the same as the standard one used to reduce the number of states in a presentation of a given constrained system S by merging states which are equivalent in the sense that the paths from which generate the same sets of words in S. •

Remark 4. Recall that by Equations (5) and (6) we have

$$\mathcal{G}_w(p,\delta) \;=\; \lim_{\epsilon\to 0}\lim_{n\to\infty} \frac{1}{n}\log\left|\,\mathcal{B}_{S_{w;n}(\,(p-\epsilon,p+\epsilon)\,)}(\lceil \delta n\rceil - 1)\,\right| ,$$

which is the growth rate of the number of pairs of words of length n in $S(G)$ that are generated from a fixed state u in G, that are of Hamming distance $< \delta n$ apart, and such that the frequency of appearance of w in each of which is approximately p. Hence, the inequality (7) in the last proof can be regarded, in a way, as a generalization of the known Chernoff bound [5, pp. 127–131].

If P is the stationary Markov chain on G of entropy rate $h(P) = \sup_{Q\in\mathcal{M}(G)} h(Q)$ and $p = \mathbf{E}_P(\mathcal{I}_w)$, then

$$\mathcal{F}_w(p) \;=\; \log\,\lambda_G\ ,$$

where λ_G is the spectral radius of the adjacency matrix of G. Also, by substituting $x=0$, we see from Theorem 2(ii) that

$$\mathcal{G}_w(p,\delta) \;\le\; \inf_{z\in\Re^+}\left\{\delta\, z + \log\,\lambda_{\mathcal{J}_w}(0,z)\right\} .$$

Write

$$R_{\mathrm{KK}}(\delta) \;\triangleq\; 2\,\log\lambda_G \;-\; \inf_{z\in\Re^+}\left\{\delta\, z + \log\,\lambda_{\mathcal{J}_w}(0,z)\right\} . \tag{9}$$

By Theorem 1 it thus follows that there exist $(S_n, M, \delta n)$-codes satisfying

$$\frac{\log\,M}{n} \;\ge\; R_{\mathrm{KK}}(\delta) \;-\; o(1)\ .$$

This is precisely the bound obtained in [8].

Applying Theorem 2 on path graphs $G^{(\ell)}$, we obtain the following result.

Theorem 3. *Let $S = S(G)$ be a primitive constrained system, let $\delta > 0$, and let $\mathbf{w}$ be a word of length $\ell + 1$ in S. Then there exist $(S_n, M, \delta n)$-codes satisfying $(\log M)/n \geq R_{\mathbf{w}}(\delta) - o(1)$, where*

$$\begin{aligned} R_{\mathbf{w}}(\delta) &= \sup_{p\in[0,1]} \left\{ 2 \inf_{x\in\Re} \left\{ p\,x + \log\,\lambda_{G^{(\ell)};\mathcal{I}_{\mathbf{w}}}(x) \right\} \right. \\ &\qquad \left. - \inf_{x\in\Re, z\in\Re^{+}} \left\{ 2p\,x + \delta\,z + \log\,\lambda_{G^{(\ell)}\times G^{(\ell)};\mathcal{J}_{\mathbf{w}}}(x,z) \right\} \right\} . \end{aligned}$$

As we pointed out earlier, Theorem 3 can be further improved by considering subsets $W \subseteq S_{\ell+1}$, rather than only one word $\mathbf{w}$, at the expense of introducing more variables to the optimization process.

6 Example: the (0,1)-RLL system

In this section we present a table of the attainable rates of $(S_n, M, \delta n)$-codes for the $(0,1)$-RLL system S presented by the graph of Figure 1 (see Example 1). Table 1 presents the values of $R_1(\delta)$, $R_{11}(\delta)$, and $R_{111}(\delta)$ obtained by setting $\mathbf{w}$ to 1, 11, and 111 in Theorem 3. The computation of $R_1(\delta)$ involves calculating the spectral radii $\lambda_{G;\mathcal{I}_1}(x)$ and $\lambda_{G\times G;\mathcal{J}_1}(x,z)$ of the matrices

$$A_{G;\mathcal{I}_1}(x) = \begin{bmatrix} 2^{-x} & 1 \\ 2^{-x} & 0 \end{bmatrix}$$

and

$$A_{G\times G;\mathcal{J}_1}(x,z) = \begin{bmatrix} 2^{-2x} & 2^{-x-z} & 2^{-x-z} & 1 \\ 2^{-2x} & 0 & 2^{-x-z} & 0 \\ 2^{-2x} & 2^{-x-z} & 0 & 0 \\ 2^{-2x} & 0 & 0 & 0 \end{bmatrix}$$

(we take the base of exponents — and therefore the base of logarithms — to be $|\Sigma| = 2$). By Remark 3 we can find $\lambda_{G\times G;\mathcal{J}_1}(x,z)$ also by calculating the spectral radius of the smaller matrix

$$A'_{G\times G;\mathcal{J}_1}(x,z) \;=\; \begin{bmatrix} 2^{-2x} & 2\cdot 2^{-x-z} & 1 \\ 2^{-2x} & 2^{-x-z} & 0 \\ 2^{-2x} & 0 & 0 \end{bmatrix}.$$

In order to compute $R_{11}(\delta)$ we need to find the spectral radii of matrices of the same orders, and the computation of $R_{111}(\delta)$ requires evaluating the spectral radius of

$$A_{G^{(2)};\mathcal{I}_{111}}(x) \;=\; \begin{bmatrix} 2^{-x} & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

and that of the 6×6 matrix $A'_{G^{(2)}\times G^{(2)};\mathcal{J}_{111}}(x,z)$ (which has the same spectral radius as the 9×9 matrix $A_{G^{(2)}\times G^{(2)};\mathcal{J}_{111}}(x,z)$).

For comparison, we also included in Table 1 the values of $R_{\mathrm{KK}}(\delta)$ (see (9)) that were obtained in [8].

δ	$R_{\mathrm{KK}}(\delta)$	$R_1(\delta)=R_{11}(\delta)$	$R_{111}(\delta)$
0.00	0.6942	0.6942	0.6942
0.05	0.4492	0.4504	0.4507
0.10	0.3055	0.3096	0.3109
0.15	0.2014	0.2094	0.2119
0.20	0.1241	0.1361	0.1399
0.25	0.0679	0.0831	0.0877
0.30	0.0295	0.0461	0.0506
0.35	0.0073	0.0218	0.0254
0.40	0.0000	0.0077	0.0097
0.45	0.0000	0.0013	0.0016
0.50	0.0000	0.0000	0.0000

Table 1: Attainable rates for the $(0,1)$*-RLL system.*

Note that $R_1(\delta)$ equals $R_\Sigma(\delta)$, since the frequencies with which the symbols 0 and 1 appear in each codeword must add to 1. In fact, in this particular constrained system we also have $R_1(\delta)=R_{11}(\delta)=R_{S_2}(\delta)$. This is due to the fact that the frequencies with which the words 01 and 10 appear in each codeword must be the same (assuming that $n \to \infty$, or that we join the two end symbols of the codeword cyclically). Therefore there is only one degree of freedom for choosing the frequencies with which the words in $S_2 = \{01, 10, 11\}$ appear in a codeword. These frequencies, in turn, are determined uniquely by the frequency of appearance of the symbol 1.

Acknowledgment

We thank P. Algoet for introducing us to convex duality and to H. Moorthy for helpful discussions in the course of work on this paper.

Appendix A

Here we prove the following known result that was used in the proof of Lemma 2 (see also [5, 56–70]).

Proposition 1. *Let P be a stationary Markov chain defined on a primitive graph G. Then,*

(i) there are at most $q^{n\,h(P)}$ cycles $\gamma \in \Gamma_n(G)$ of empirical distribution $P_\gamma = P$;

(ii) there are at least $q^{n\,(h(P)-o(1))}$ cycles $\gamma \in \Gamma_n(G)$ that satisfy

$$\sum_{e \in E_G} |P_\gamma(e) - P(e)| = o(1) \ . \tag{10}$$

We remark that the statement of Proposition 1 does not depend on the specific starting state used to define $\Gamma_n(G)$.

Proof of Proposition 1. We first show part (i). Let u be the starting state used to define $\Gamma_n(G)$ and let $\Pi_n(G)$ denote the set of all paths of length n in G that start at u. For each $\gamma \in \Pi_n(G)$, we define the empirical distribution P_γ as in (1) (except that here P_γ need not be a stationary Markov chain). For a given stationary Markov chain P, let the (conditional) probability of a path $\gamma = e_1 e_2 \ldots e_n \in \Pi_n(G)$ be defined by $P(\gamma \,|\, u) \triangleq \prod_{i=1}^{n} P(e_i \,|\, \imath(e_i))$. Note that $P(\gamma \,|\, u)$ is the probability of obtaining a path γ by a random walk of length n in G, given that we start at state u, according to the probability measure P on the edges of G. It is easy to verify that

$$P(\gamma \,|\, u) \;=\; \prod_{e \in E_G} P(e \,|\, \imath(e))^{nP_\gamma(e)} \tag{11}$$

which, for $P_\gamma = P$ becomes

$$P(\gamma \,|\, u) \;=\; \prod_{e \in E_G} P(e \,|\, \imath(e))^{nP(e)} \;=\; q^{-n\,h(P)} \;.$$

On the other hand,

$$\sum_{\substack{\gamma \in \Pi_n(G)\,: \\ P_\gamma = P}} P(\gamma \,|\, u) \;\leq\; 1$$

and, therefore, the number of paths $\gamma \in \Pi_n(G)$ with $P_\gamma = P$ is at most $q^{n\,h(P)}$. Part (i) now follows from the fact that $\Gamma_n(G) \subseteq \Pi_n(G)$.

We now turn to part (ii). A stationary Markov chain P is of *full support* if $P(e) > 0$ for every $e \in E_G$. It suffices to prove the claim for full-support stationary Markov chains because one can always perturb a stationary Markov chain to obtain one of full support. The nice thing about full-support stationary Markov chains is that the law of large numbers applies.

Given a full-support stationary Markov chain P, a path $\gamma \in \Pi_n(G)$ is called 'good' if it satisfies (10). By (11) we have

$$P(\gamma \,|\, u) \;=\; \prod_{e \in E_G} P(e \,|\, \imath(e))^{n\,(P(e)-o(1))} \;=\; q^{-n\,(h(P)-o(1))} \quad \text{for every good path } \gamma \in \Pi_n(G)\;. \tag{12}$$

The law of large numbers, applied on the indicator functions of each of the edges of G, implies that each path γ in $\Pi_n(G)$ is good with probability $1 - o(1)$. And, by (12) we thus conclude that the number of good paths in $\Pi_n(G)$ is at least $q^{n\,(h(P)-o(1))}$.

We have now completed the proof of part (ii) except that the good paths need not be in $\Gamma_n(G)$. Here is where we use the primitivity assumption, namely, that there is a path of fixed length N_G between any two states in G: take any good path of length $n - N_G$, and extend it to a path of length n that terminates at u. This yields at least $q^{(n-N_G)\,(h(P)-o(1))} = q^{n\,(h(P)-o(1))}$ good cycles in $\Gamma_n(G)$. $\square$

Appendix B

Here we establish the following statement

$$\log \lambda_{G;\psi}(\mathbf{x},\mathbf{z}) \;=\; \phi(\mathbf{x},\mathbf{z}) \;=\; \sup_{P\in\mathcal{M}(G)} \Big\{ h(P) - \mathbf{x}\cdot\mathbf{E}_P(f) - \mathbf{z}\cdot\mathbf{E}_P(g) \Big\} \tag{13}$$

that was made in the proof of Lemma 5 (refer to Section 5 for notations).

For stationary Markov chains P and Q, define the information divergence $D(P\|Q)$ by

$$D(P\|Q) \;\triangleq\; \sum_{e\in E_G} P(e)\log\frac{P(e\,|\,\imath(e))}{Q(e\,|\,\imath(e))} \;.$$

Now,

$$D(P\|Q) \;\geq\; 0\;, \qquad \text{with equality if and only if} \quad P=Q\;. \tag{14}$$

This is a well-known application of Jensen's inequality. Now, we define a particular stationary Markov chain $Q = Q_{\mathbf{x},\mathbf{z}}$ based on the matrix $A_{G;\psi}$ as follows. By the Perron-Frobenius Theorem, since G is irreducible, $A_{G;\psi}$ has a positive eigenvector $\mathbf{b} = [b_u]_{u\in V_G}$, unique up to scaling, corresponding to the eigenvalue $\lambda_{G;\psi}$. It is well known that, given strictly positive conditional probability values $P(e\,|\,\imath(e))$ for every $e\in E_G$, the stationary Markov chain P which yields such values is uniquely determined. Let Q be the unique stationary Markov chain whose conditional probabilities are

$$Q(e\,|\,\imath(e)) \;=\; \frac{q^{-\mathbf{x}\cdot f(e)-\mathbf{z}\cdot g(e)}\cdot b_{\tau(e)}}{\lambda_{G;\psi}(\mathbf{x},\mathbf{z})\cdot b_{\imath(e)}}\;,$$

where $\tau(e)$ denotes the terminal state of e. Now, a straightforward computation reveals that for all stationary Markov chains P on G,

$$D(P\|Q) \;=\; \log\lambda_{G;\psi}(\mathbf{x},\mathbf{z}) + \mathbf{x}\cdot\mathbf{E}_P(f) + \mathbf{z}\cdot\mathbf{E}_P(g) - h(P)\;.$$

This, together with (14), yields (13) as desired. In fact, the supremum in the right-hand side of (13) is attained at $P=Q$.

References

[1] G.D. Cohen, S.N. Litsyn, *DC-constrained error-correcting codes with small digital sum, IEEE Trans. Inform. Theory,* 37 (1991), 949–955.

[2] I. Csiszár, T.M. Cover, B.-S. Choi, *Conditional limit theorems under Markov conditioning, IEEE Trans. Inform. Theory,* 33 (1987), 788–801.

[3] H.C. Ferreira, *Lower bounds on the minimum-Hamming distance achievable with runlength constrained or DC-free block codes and the synthesis of a* $(16, 8)$ $d_{\min} = 4$ *DC-free block code, IEEE Trans. Magnetics,* 20 (1984), 881–883.

[4] E.R. Berlekamp, *Algebraic Coding Theory,* McGraw-Hill, New York (1968).

[5] R.G. Gallager, *Information Theory and Reliable Communication,* John Wiley, New York (1968).

[6] J. Justesen, T. Høholdt, *Maxentropic Markov chains, IEEE Trans. Inform. Theory,* 30 (1984), 665–667.

[7] R. Karabed, A. Khayrallah, D. Neuhoff, *Capacity of costly noiseless channels,* preprint.

[8] V.D. Kolesnik, V.Yu. Krachkovsky, *Generating functions and lower bounds on rates for limited error-correcting codes, IEEE Trans. Inform. Theory,* 37 (1991), 778–788.

[9] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming,* Addison-Wesley, Reading, Massachusetts (1973).

[10] B. Marcus, S. Tuncel, *Entropy at a weight-per-symbol and embeddings of Markov chains, Inventiones Mathematicae,* 102 (1990), 235–266.

[11] R. McEliece, L. Popovic, *The capacity-cost function of a noiseless channel with several cost constraints,* preprint.

[12] W. Parry, S. Tuncel, *Classification Problems in Ergodic Theory,* LMS Lecture Note Series No. 67, Cambridge University Press, Cambridge (1982).

[13] T. Rockafellar, *Convex Analysis,* Princeton University Press, Princeton (1970).

[14] Ø. Ytrehus, *On error-controlling* (d, k) *constrained block codes, Proc. Fourth Joint Swedish–Soviet Int'l Workshop Inform. Theory* (1989), 315–317.

[15] E. Zehavi, J.K. Wolf, *On runlength codes, IEEE Trans. Inform. Theory,* 34 (1988), 45–54.