

Capacity of Multidimensional Constrained Channels

Estimates and Exact Computations

by

Erez Loidor

B.Sc., The Technion Inst. of Technology, 1998

M.Sc., The Technion Inst. of Technology, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2010

© Erez Loidor 2010

Abstract

This work considers channels for which the input is constrained to be from a given set of \mathcal{D} -dimensional arrays over a finite alphabet. Such a set is called a *constraint*. An encoder for such a channel transforms arbitrary arrays over the alphabet into constrained arrays in a decipherable manner. The *rate* of the encoder is the ratio of the size of its input to the size of its output. The *capacity* of the channel or constraint is the highest achievable rate of any encoder for the channel. We compute the exact capacity of two families of multidimensional constraints. We also generalize a known method for obtaining lower bounds on the capacity, for a certain class of 2-dimensional constraints, and improve the best known bounds for a few constraints of this class.

Given a binary \mathcal{D} -dimensional constraint, a \mathcal{D} -dimensional array with entries in $\{0, 1, \square\}$ is called “valid”, for the purpose of this abstract, if any “filling” of the ‘ \square ’s in the array with ‘0’s and ‘1’s, independently, results in an array that belongs to the constraint. The density of ‘ \square ’s in the array is called the *insertion rate*. The largest achievable insertion rate in arbitrary large arrays is called the *maximum insertion rate*. An unconstrained encoder for a given insertion rate transforms arbitrary binary arrays into valid arrays having the specified insertion rate. The *tradeoff function* essentially specifies for a given insertion rate the maximum rate of an unconstrained encoder for that insertion rate. We determine the tradeoff function for a certain family of 1-dimensional constraints.

Given a 1-dimensional constraint, one can consider the \mathcal{D} -dimensional constraint formed by collecting all the \mathcal{D} -dimensional arrays for which the original 1-dimensional constraint is satisfied on every “row” in every “direction”. The sequence of capacities of these \mathcal{D} -dimensional generalizations has a limit as \mathcal{D} approaches infinity, sometimes called the *infinite-dimensional capacity*. We partially answer a question of [37], by proving that for a large class of 1-dimensional constraints with maximum insertion rate 0, the infinite dimensional capacity equals 0 as well.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgements	vii
Dedication	viii
Statement of Co-Authorship	ix
1 Overview	1
2 Multidimensional constraints	6
2.1 One-dimensional constraints and labeled directed graphs	6
2.2 Higher dimensional constraints	8
2.3 Capacity	10
2.4 Axial product	13
2.5 Two-dimensional constraints	15
2.5.1 Horizontal and vertical strips	15
2.6 Open questions	17
3 Lower bounds on capacity of 2-dimensional symmetric constraints	18
3.1 Constraints with symmetric edge-constrained strips	18
3.2 Constraints with symmetric vertex-constrained strips	27
3.3 Capacity bounds for axial products of constraints	31
3.4 Heuristics for choosing ϕ	33
3.4.1 Using max-entropic probabilities	33
3.4.2 General optimization	36
3.5 Numerical results for selected constraints	38

Table of Contents

3.5.1	The constraint RWIM	40
3.5.2	The constraint EVEN ^{⊗2}	40
3.5.3	The constraint CHG(<i>b</i>) ^{⊗2}	40
3.6	Open questions	43
4	Exact computation of capacity	44
4.1	The capacity of ODD ^{⊗D}	44
4.2	The capacity of CHG(2) ^{⊗D}	45
5	Multi-choice constraints and independence capacity	49
5.1	Multi-choice constraints	49
5.2	Independence capacity	51
5.3	Independence capacity and axial products	53
5.4	Independence capacity and $\lim_{D \rightarrow \infty} \text{cap}(S^{\otimes D})$	55
5.5	Open questions	61
6	The tradeoff function for binary 1-dimensional constraints	62
6.1	A brief overview of digital recording	62
6.2	Previous work	64
6.3	Background and definitions	64
6.4	Proof of Theorem 12	69
6.5	Proof of Theorem 13	71
6.5.1	Outline of proof	72
6.5.2	Proof of propositions	78
6.6	Open questions	96
7	Bounds on capacity using probability	97
7.1	Some correlation inequalities	97
7.2	Bounds on capacity using probability	99
7.2.1	Proof of Proposition 14	102
7.2.2	Proof of Lemma 11	106
7.3	Open questions	108
	Bibliography	109

List of Tables

3.1	Matrix size in the method of [3, 7] for the NAK constraint.	39
3.2	Best bounds on capacities of certain constraints.	40
3.3	Lower bounds using max-entropic probability heuristic.	41
3.4	Lower bounds using optimization.	42
6.1	Values of $ \Delta(g) $ for $1 \leq g \leq 2d+6$ and $d \geq 2$	90

List of Figures

1.1	Presentations of 1-dimensional constraints.	2
2.1	Presentations of 2-dimensional constraints.	10
3.1	Paths generating an $\ell \alpha \times n$ -array of S , in $\mathcal{G}_n^{(\mathcal{V})}$ and \mathcal{I}	23
4.1	Example of the graph $G_{\mathbf{r}}$ for $\mathfrak{D} = 2, n = 6$	47
5.1	Proof of Theorem 10.	60
6.1	Graphs of the tradeoff function for certain RLL constraints.	68
6.2	The graph $\widehat{\mathcal{G}}_{\mathcal{F}_{\text{RLL}}(d, \infty)}$	69
6.3	The non-trivial component of $\widehat{\mathcal{G}}_{\mathcal{F}_{\text{RLL}}(d, 2d+2)}$	72

Acknowledgements

I am sincerely grateful for the support of my advisor, Prof. Brian Marcus, for being actively involved in my work, being realistically optimistic along the way, and making these abstract concepts a little more tangible by allowing me to discuss them with another. The only downside is that I now have high expectations of the next person I work with.

Dedicated to my parents: Eti & Adam Louidor.

Statement of Co-Authorship

Chapters 2,3 and 4 were jointly authored by Erez Loidor and Brian Marcus. Erez Loidor and Brian Marcus were responsible for the identification and design of the research program. The research was performed by Erez Loidor and Brian Marcus with the majority of the research done by Erez Loidor. The manuscript was prepared by Erez Loidor and Brian Marcus with the majority of the writing done by Erez Loidor. A version of these chapters appears in [28].

Chapter 5 was jointly authored by Erez Loidor, Brian Marcus and Ronnie Pavlov. Panu Chaichanavong, Brian Marcus and Tze-Lei Poo were responsible for the identification and design of the research program. The research was performed by Panu Chaichanavong, Erez Loidor, Brian Marcus, Ronnie Pavlov and Tze-Lei Poo. The manuscript was prepared by Erez Loidor, Brian Marcus and Ronnie Pavlov with the majority of the writing done by Erez Loidor. A version of this chapter appears as part of [29].

Chapter 6 was authored by Erez Loidor. Brian Marcus was responsible for the identification and design of the research program. The research was performed by Erez Loidor. The manuscript was prepared by Erez Loidor. A version of this chapter appears in [27].

Chapter 7 was authored by Erez Loidor. Erez Loidor and Brian Marcus were responsible for the identification and design of the research program. The research was performed by Erez Loidor. The manuscript was prepared by Erez Loidor.

Chapter 1

Overview

Fix an alphabet Σ and let \mathcal{G} be a directed graph whose edges are labeled with symbols in Σ . Each path in \mathcal{G} corresponds to a finite word obtained by reading the labels of the edges of the path in sequence. The path is said to generate the corresponding word, and the set of words generated by all finite paths in the graph is called a *1-dimensional constrained system* or a *1-dimensional constraint*. Such a graph is called a *presentation* of the constraint. We say that a word satisfies the constraint if it belongs to the constrained system. One-dimensional constraints have found widespread applications in digital storage systems, where they are used to model the set of sequences that can be written reliably to a medium. A central example is the binary run-length-limited constraint, denoted $\text{RLL}(d, k)$ for nonnegative integers $0 \leq d \leq k$, consisting of all binary sequences in which the number of ‘0’s between consecutive ‘1’s is at least d , and each “run” of ‘0’s (that is a contiguous sub-sequence of ‘0’s) has length at most k . The value of k is allowed to be ∞ , in which case there is no restriction on the maximum length of a run of ‘0’s. Another 1-dimensional constraint, often used in practice, is the bounded-charge constraint, denoted $\text{CHG}(b)$, for some positive integer b ; it consists of all words $w_1 w_2 \dots w_\ell$, where $\ell = 0, 1, 2, \dots$ and each w_i is either $+1$ or -1 , such that for all $1 \leq i \leq j \leq \ell$, $|\sum_{k=i}^j w_k| \leq b$. This constraint is often used to overcome low frequency noise such as fingerprints on compact discs. Other examples of 1-dimensional constraints are the EVEN and ODD constraints, which contain all finite binary sequences in which the number of ‘0’s between consecutive ‘1’s is even and odd, respectively. Presentations for these constraints are given in Figure 1. See [32] for more examples of 1-dimensional constraints and a more detailed explanation of their use in storage systems.

In this work, we consider multidimensional constraints of dimension \mathcal{D} for some positive integer \mathcal{D} . Such a constraint is a set, specified by \mathcal{D} edge-labeled directed graphs, of finite-size \mathcal{D} -dimensional arrays with entries over some finite alphabet. In Chapter 2 we give a precise definition of what we mean by \mathcal{D} -dimensional constraints. As in the 1-dimensional case, we say that an array satisfies the constraint if it belongs to it. Given a 1-dimensional constraint S , one can construct a \mathcal{D} -dimensional constraint by collecting all the \mathcal{D} -dimensional arrays for which the original 1-dimensional constraint is satisfied on every “row” in every “di-

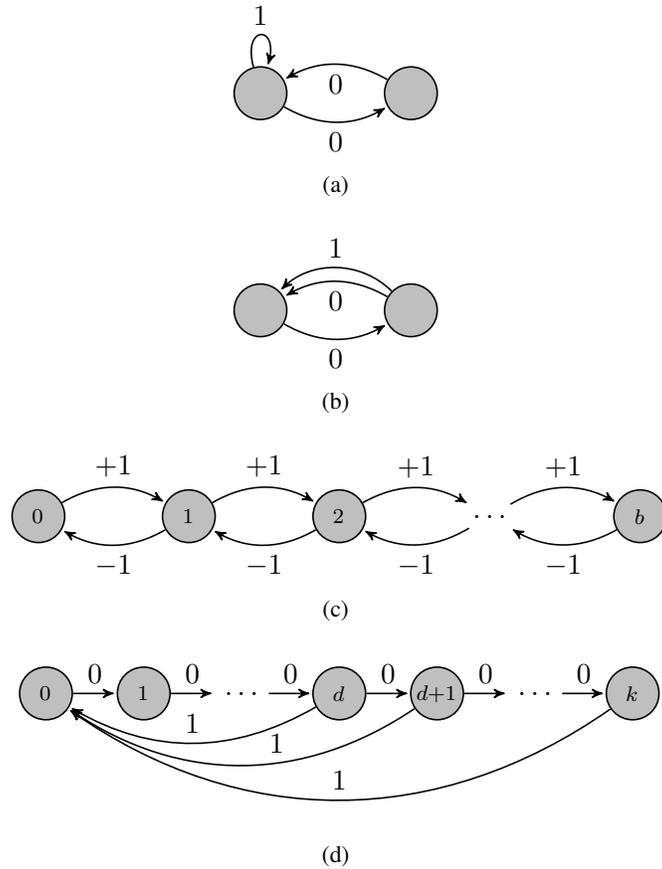


Figure 1.1: Presentations of 1-dimensional constraints: (a) EVEN; (b) ODD; (c) CHG(b); (d) RLL(d, k).

rection” along an “axis” of the array. We denote such a \mathcal{D} -dimensional constraint by $S^{\otimes \mathcal{D}}$. A well-known 2-dimensional constraint studied in statistical mechanics is the so called “hard-square” constraint. It consists of all finite-size (2-dimensional) binary arrays which do not contain 2 adjacent ‘1’s either horizontally or vertically. Two variations of this constraint are the isolated ‘1’s or “non-attacking-kings” constraint, denoted NAK, and the “read-write-isolated-memory” constraint, denoted RWIM. The former consists of all finite-size binary arrays in which there are no two adjacent ‘1’s either horizontally, vertically, or diagonally, and the latter consists of all finite-size binary arrays in which there are no two adjacent ‘1’s either horizontally, or diagonally. Like their 1-dimensional counterparts, 2-dimensional

constraints play a role in storage systems, where with recent developments, information is written in a true 2-dimensional fashion rather than using essentially 1-dimensional tracks. The RWIM constraint is used to model sequences of states of a binary linear memory in which no two adjacent entries may contain a ‘1’, and in every update, no two adjacent entries are both changed. See [4] and [13] for more details.

Let S now be a \mathfrak{D} -dimensional constraint over an alphabet Σ . For a \mathfrak{D} -tuple $\mathbf{m} = (m_1, m_2, \dots, m_{\mathfrak{D}})$ of positive integers, let $S_{\mathbf{m}}$ or $S_{m_1 \times \dots \times m_{\mathfrak{D}}}$ denote the set of all $m_1 \times m_2 \times \dots \times m_{\mathfrak{D}}$ arrays in S , and $\text{vol}(\mathbf{m})$ denote the product of the entries of \mathbf{m} . We say that a sequence $\mathbf{m}_i = (m_1^{(i)}, \dots, m_{\mathfrak{D}}^{(i)})$ diverges to infinity, denoted $\mathbf{m}_i \rightarrow \infty$, if $(m_j^{(i)})_{i=1}^{\infty}$ does for each j . The capacity of S is then defined by

$$\text{cap}(S) = \lim_{i \rightarrow \infty} \frac{\log |S_{\mathbf{m}_i}|}{\text{vol}(\mathbf{m}_i)}, \quad (1.1)$$

where $(\mathbf{m}_i)_{i=1}^{\infty}$ is a sequence of \mathfrak{D} -tuples in $(\mathbb{N})^{\mathfrak{D}}$ diverging to infinity, $|\cdot|$ denotes cardinality, and $\log = \log_2$. We show in Chapter 2 that the limit always exists and is independent of the choice of $(\mathbf{m}_i)_{i=1}^{\infty}$. The capacity is a fundamental quantity that has been studied under different names in several disciplines dealing with constrained systems. In symbolic dynamics it is known as the “topological entropy” and in statistical physics it is derived from the “grand partition function”. In the context of coding for storage systems, capacity has a practical role. As already mentioned, in many such systems due to physical constraints, only a subset of binary sequences can be written to the media reliably. This subset is typically modeled as a 1-dimensional constrained system over the binary alphabet $\{0, 1\}$. In practice, user information, consisting of an arbitrary sequence of binary digits is encoded into a sequence of the constraint before being written to the media. When reading back the data, the constrained sequence is decoded and the original information is recovered. Typically, an encoder divides its input into fixed size blocks of p digits each, emitting, for each block, a block of q digits, such that when all the output blocks are concatenated the result satisfies the constraint. The ratio p/q is called the *rate* of the encoder and, naturally, it is desirable that the rate of the encoder be as large as possible. It turns out, that the capacity of a 1-dimensional constraint, is the largest possible rate of any such encoder for the constraint, and hence, is important for the design and evaluation of efficient encoders for digital storage systems.

While there is a formula for computing the capacity of a general 1-dimensional constrained system (up-to finding the largest root of a polynomial), no such formula is known for 2- and higher-dimensional constraints. There are only a handful of constraints for which the capacity is nonzero and is known exactly [23, 25, 39].

Even for the hard-square constraint, the exact capacity is unknown and the problem goes back more than 40 years [12]. Some rigorous evidence for the hardness of computing the capacity of multidimensional constraints is given in [2], where it is shown that there is no algorithm that accepts a 2-dimensional constraint system S as input and determines whether $|S_{m \times n}| \geq 1$ for all m, n .

In many storage systems, restricting the set of sequences that can be written to the media to be from a constrained system is not enough to ensure the low bit-error-rate required in these systems. Accordingly, a conventional error correcting code, or ECC, is used in addition to further reduce the number of errors. Traditionally, arbitrary user information is encoded twice, first by the ECC encoder and then by the constrained system encoder, before it is written to the media. Immink and Wijngaarden [40] proposed a scheme to embed the ECC directly in the constrained system. In this scheme, the user information sequence is encoded into a binary sequence in which certain preset positions are left blank. These positions are denoted by ‘□’s and are “unconstrained” in the sense that any way of filling them independently with ‘0’s and ‘1’s would result in a sequence that satisfies the given constrained system. The density of ‘□’s in the resulting sequence is called the *insertion rate*. Next, a systematic ECC is used to compute parity check bits, which are stored directly in these unconstrained positions and the resulting sequence is written to the media. In this manner the written sequence is both a constrained sequence and an ECC codeword. As the error correcting capability of an ECC depends on the number of parity-check bits, it is desirable that the number of unconstrained positions emitted by the encoder, or equivalently the insertion rate, be as large as possible. The maximum insertion rate is the largest insertion rate achievable in arbitrary long sequences. On the other hand as the number of constrained sequences of a given length with a given insertion rate is inversely related to the insertion rate, increasing the insertion rate reduces the overall encoding rate (that is the rate of the combined constrained system and ECC encoders). The *tradeoff function* [5, 38] quantifies this and provides for a given insertion rate the highest possible rate of any matching encoder. Accordingly, the tradeoff function of a constraint evaluated at 0 equals its capacity and thus can be regarded as a generalization of capacity.

In this work we generalize some of these concepts to higher dimensional constraints and non-binary alphabets. In particular, we define *independence capacity* of a constraint that, roughly speaking, captures the contribution of independence between symbols in arrays of the constraint to its capacity. For the binary alphabet this coincides with the notion of maximum insertion rate. We denote it by $\text{cap}_{\text{ind}}(S)$ for a constraint S . For a 1-dimensional constraint S , it turns out that $\text{cap}(S^{\otimes 1}) \geq \text{cap}(S^{\otimes 2}) \geq \dots \geq \text{cap}_{\text{ind}}(S)$ and we denote the limit

$\lim_{\mathfrak{D} \rightarrow \infty} \text{cap}(S^{\otimes \mathfrak{D}})$ by $\text{cap}_{\infty}(S)$; hence $\text{cap}_{\infty}(S) \geq \text{cap}_{\text{ind}}(S)$. Chaichanavong and Poo observed the curious fact that for all 1-dimensional constraints S , for which we know $\text{cap}_{\infty}(S)$, it turns out to be equal to $\text{cap}_{\text{ind}}(S)$, and they ask whether this always holds [37]. Here, we give a partial answer by showing that for a large class of constraints, if $\text{cap}_{\text{ind}}(S) = 0$, then so is $\text{cap}_{\infty}(S)$ and the convergence is exponentially fast.

The main contributions of this work are summarized below.

- Calculated the capacity of $\text{CHG}(2)^{\otimes \mathfrak{D}}$ and $\text{ODD}^{\otimes \mathfrak{D}}$, for all $\mathfrak{D} \in \mathbb{N}$ [28].
- Generalized an earlier method for computing lower bounds on a certain class of 2-dimensional constraints, and using the method improved the best bounds on $\text{cap}(\text{NAK})$ and $\text{cap}(\text{RWIM})$, and gave the first published estimates of $\text{cap}(\text{EVEN}^{\otimes 2})$ and $\text{cap}(\text{CHG}(3)^{\otimes 2})$ [28].
- Showed that for a large class of constraints S with zero independence capacity, $\text{cap}(S^{\otimes \mathfrak{D}}) \rightarrow 0$ exponentially fast [29, 30].
- Determined the tradeoff function of $\text{RLL}(d, \infty)$ and $\text{RLL}(d, 2d + 2)$ [27].
- Showed how correlation inequalities can be used to obtain lower bounds on “monotone” 2-dimensional constraints.

This work is organized as follows. In Chapter 2 we define multidimensional constraints and related concepts. In Chapter 3 we show a method for obtaining lower bounds on the capacity, for a class of 2-dimensional constraints. In Chapter 4 we calculate the exact capacity of $\text{CHG}(2)^{\otimes \mathfrak{D}}$ and $\text{ODD}^{\otimes \mathfrak{D}}$. In Chapter 5 we generalize some of the concepts of [37, 38] to dimensions larger than 1 and non-binary alphabets and define multi-choice constraints and independence capacity. We show some of their properties and in particular prove the exponential convergence of $\text{cap}_{\text{ind}}(S)$ to $\text{cap}_{\infty}(S)$ for certain constraints S , discussed above. In Chapter 6 we compute the tradeoff function for $\text{RLL}(d, 2d + 2)$ and $\text{RLL}(d, \infty)$. Finally, in Chapter 7 we show some probabilistic inequalities that hold for certain 2-dimensional binary constraints. From these, we obtain a lower bound on the capacity.

Chapter 2

Multidimensional constraints*

In this chapter we define multidimensional constraints and related concepts that we use in the rest of this thesis.

2.1 One-dimensional constraints and labeled directed graphs

We deal with a finite directed graph $G = (V, E)$, sometimes simply called a graph, with vertices V and edges E . We occasionally refer to the vertices as *states* and to the edges as *transitions*. For $e \in E$ we denote by $\sigma_G(e)$ and $\tau_G(e)$ the initial and terminal vertices of e in G , respectively. We shall omit the subscript G from σ_G and τ_G when the graph is clear from the context. For a sequence $(a_i)_{i=1}^\ell$ and a set A , we abuse notation and write $(a_i) \subseteq A$ to mean that $a_i \in A$ for $i = 1, 2, \dots, \ell$. A *path of length ℓ* in G is a sequence of ℓ edges $(e_i)_{i=1}^\ell \subseteq E$, where for $i = 1, 2, \dots, \ell-1$, $\tau(e_i) = \sigma(e_{i+1})$. The path starts at the vertex $\sigma(e_1)$ and ends at the vertex $\tau(e_\ell)$. A cycle in G is a path that starts and ends at the same vertex. Fix a finite alphabet Σ . A directed *labeled graph* \mathcal{G} with labels in Σ is a pair $\mathcal{G} = (G, \mathcal{L})$, where $G = (V, E)$ is a directed graph, and $\mathcal{L} : E \rightarrow \Sigma$ is a labeling of the edges of G with symbols of Σ . The paths and cycles of \mathcal{G} are inherited from G and we will sometime use $\sigma_{\mathcal{G}}$ and $\tau_{\mathcal{G}}$ to denote σ_G and τ_G respectively. For a path $(e_i)_{i=1}^\ell$ of \mathcal{G} , we say the path *generates* the word $\mathcal{L}(e_1)\mathcal{L}(e_2)\dots\mathcal{L}(e_\ell)$ in Σ^* (Σ^* denote the set of all finite (1-dimensional) words over Σ).

As mentioned in Chapter 1, a *1-dimensional constraint* or *1-dimensional constrained system* over Σ is the set of all words generated from finite paths in some labeled graph with labels in Σ . The graph is called a *presentation* of the constraint. A labeled direct graph is called *lossless* if for any two of its vertices u and v , all paths starting at u and terminating at v generate distinct words. It is called *deterministic* if there are no two distinct edges with the same initial vertex and the same label. Every 1-dimensional constraint S has a deterministic, and therefore lossless,

*A version of this chapter has been published. Loudior, E. and Marcus, B.H. (2010) Improved Lower Bounds on Capacities of Symmetric 2-Dimensional Constraints using Rayleigh Quotients. IEEE Transactions on Information Theory 56:1624–1639.

presentation [32].

A 1-dimensional constraint over an alphabet Σ is said to have memory m , for some positive integer m , if for every word w of more than m letters over Σ , in which every sub-word of $m + 1$ consecutive letters satisfies S , it holds that w satisfies S as well, and m is the smallest integer for which this is true. A 1-dimensional constraint with memory m , for some integer m , is called a *finite-type* constraint. Of the examples introduced in Chapter 1, $\text{RLL}(d, k)$ is a finite-type constraint—with memory k , for $k < \infty$, and memory d , for $k = \infty$ —whereas EVEN , ODD and $\text{CHG}(b)$ for $b \geq 2$ are not finite-type constraints.

We introduce two 1-dimensional constraints defined by general directed graphs. Let $G = (V, E)$ be a directed graph. The *edge constraint defined by G* , denoted $X(G)$, is the 1-dimensional constraint over the alphabet E , presented by $\mathcal{G} = (G, I_E)$ where I_E is the identity map on E . Equivalently, an edge constraint is a constraint that can be presented by a labeled graph in which all the edges have distinct labels. For a graph $G = (V, E)$ with no parallel edges, *the vertex-constraint defined by G* , denoted $\check{X}(G)$, is the set

$$\left\{ (v_i)_{i=1}^{\ell} \subseteq V : \ell = 0, 1, 2, \dots, \text{ and for } 1 \leq i < \ell, \exists e_i \in E \text{ s.t. } \right. \\ \left. \sigma(e_i) = v_i, \tau(e_i) = v_{i+1} \right\}.$$

It is not hard to verify that vertex-constraints and edge-constraints are 1-dimensional constraints with memory (at most) 1. In fact, the vertex constraints are precisely the finite-type constraints with memory (at most) 1, and it can be shown that edge constraints are characterized as follows. The *follower set* of a symbol a in a constraint S is defined to be $\{b : ab \in S\}$; edge constraints are precisely the constraints with memory 1 such that any two follower sets are either disjoint or identical [26, exercise 2.3.4].

A graph $G = (V, E)$ is *irreducible* if for any pair of vertices $u, v \in V$ there is a path in G starting at u and terminating at v ; otherwise it is *reducible*. A graph G is *primitive* if it is irreducible and the gcd of the lengths of all cycles of G is 1. These concepts naturally extend to labeled graphs as well. We denote by $A(G)$ the adjacency matrix of G : namely the $|V| \times |V|$ matrix where $(A(G))_{i,j}$ is the number of edges in G from i to j , where $|\cdot|$ denotes cardinality. We use $\mathbf{1}$ in this work to denote a real vector in which each entry is 1 and for two real matrices (or vectors) M, N of the same size we write $M \leq N$ and $M < N$ if the corresponding inequality holds entry-wise. We say that a graph G is *symmetric* if $A(G)$ is symmetric. We say that a vertex of a graph is *isolated* if it has neither outgoing nor incoming edges. We say that a vertex-constraint (resp. edge-constraint) is *symmetric* if it is defined by a symmetric graph. For a vertex-constraint, this definition is equivalent to requiring that the constraint is closed under reversal of the order of symbols

in words. Note, that in a symmetric edge-constraint, up to removal of isolated vertices, the (unlabeled) graph defining the constraint is unique.

2.2 Higher dimensional constraints

We consider multidimensional arrays of dimension \mathfrak{D} —a positive integer. We use \mathbb{Z}^+ to denote the set of nonnegative integers. For a \mathfrak{D} -tuple $\mathbf{m} = (m_1, \dots, m_{\mathfrak{D}}) \in (\mathbb{Z}^+)^{\mathfrak{D}}$ we denote by $[\mathbf{m}]$ the Cartesian product $\prod_i 0, \dots, m_i - 1$, and for a finite set \mathcal{A} , we call an $m_1 \times m_2 \times \dots \times m_{\mathfrak{D}}$ \mathfrak{D} -dimensional array with entries in \mathcal{A} , a \mathfrak{D} -dimensional array of size \mathbf{m} over \mathcal{A} . We shall index the entries of such an array by $[\mathbf{m}]$. We use $\mathcal{A}^{\mathbf{m}}$ and $\mathcal{A}^{m_1 \times \dots \times m_{\mathfrak{D}}}$ to denote the set of all \mathfrak{D} -dimensional arrays of size \mathbf{m} over \mathcal{A} . We define $\mathcal{A}^{*\dots*} = \mathcal{A}^{*\mathfrak{D}}$, where the number of ‘*’s in the superscript is \mathfrak{D} , by

$$\mathcal{A}^{*\mathfrak{D}} = \bigcup_{\mathbf{m}} \mathcal{A}^{\mathbf{m}},$$

as the set of all finite-size \mathfrak{D} -dimensional arrays with entries in \mathcal{A} . Let $\Gamma \in \Sigma^{*\mathfrak{D}}$ be such an array. Given an integer $1 \leq i \leq \mathfrak{D}$, a *row in direction i* of Γ is a sequence of entries of Γ of the form $(\Gamma_{(k_1, k_2, \dots, k_{i-1}, j, k_{i+1}, \dots, k_{\mathfrak{D}})})_{j=0}^{m_i-1}$ for some integers $k_l \in [m_l]$; $1 \leq l \leq \mathfrak{D}$, $l \neq i$. In this work, for $\mathfrak{D} = 2$, we use the convention that direction 1 is the vertical direction and direction 2 is the horizontal; thus the columns of a 2-dimensional array are its rows in direction 1, and its “traditional rows” are its rows in direction 2. Let \mathcal{A}, \mathcal{B} be finite sets and $\mathcal{L} : \mathcal{A} \rightarrow \mathcal{B}$ be a mapping. We extend \mathcal{L} to a mapping $\mathcal{L} : \mathcal{A}^{*\mathfrak{D}} \rightarrow \mathcal{B}^{*\mathfrak{D}}$ as follows. For a \mathfrak{D} -dimensional array $\Gamma \in \mathcal{A}^{\mathbf{m}}$, $\mathcal{L}(\Gamma)$ is the array in $\mathcal{B}^{\mathbf{m}}$ obtained by applying \mathcal{L} to each entry of Γ , that is

$$(\mathcal{L}(\Gamma))_{\mathbf{j}} = \mathcal{L}((\Gamma)_{\mathbf{j}}) \quad , \quad \mathbf{j} \in [\mathbf{m}].$$

Additionally, for a subset $S \subseteq \mathcal{A}^{*\mathfrak{D}}$ we define $\mathcal{L}(S) = \{\mathcal{L}(\Gamma) : \Gamma \in S\}$.

We generalize the definition of a constrained system to \mathfrak{D} dimensions. Let $\bar{\mathcal{G}} = (\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{\mathfrak{D}})$, be a \mathfrak{D} -tuple of labeled graphs with the same set of edges E and the same labeling $\mathcal{L} : E \rightarrow \Sigma$. The edge e has \mathfrak{D} pairs of initial and terminal vertices $(\sigma_{\mathcal{G}_i}(e), \tau_{\mathcal{G}_i}(e))$ —one for each graph \mathcal{G}_i in $\bar{\mathcal{G}}$. We say that an array $\Gamma \in \Sigma^{*\mathfrak{D}}$ of size \mathbf{m} is *generated* by $\bar{\mathcal{G}}$ if there exists an array $\Gamma' \in E^{*\mathfrak{D}}$ of size \mathbf{m} , such that for $i = 1, 2, \dots, \mathfrak{D}$, every row in direction i of Γ' is a path in \mathcal{G}_i , and $\mathcal{L}(\Gamma') = \Gamma$. We call the set of all arrays $\Gamma \in \Sigma^{*\mathfrak{D}}$ generated by $\bar{\mathcal{G}}$, the *\mathfrak{D} -dimensional constrained system* or the *\mathfrak{D} -dimensional constraint* presented by $\bar{\mathcal{G}}$, and denote it by $\mathcal{X}(\bar{\mathcal{G}})$ (note the difference from the notation used for edge-constraints where

2.2. Higher dimensional constraints

the argument inside the parentheses is an unlabeled graph). We say that $\bar{\mathcal{G}}$ is a *presentation* for $X(\bar{\mathcal{G}})$.

In [14], 2-dimensional constrained systems are defined by vertex-labeled graphs, with a common set of vertices and a common labeling on the vertices. It can be shown that their definition (generalized to higher dimensions) is equivalent to ours. We find it more convenient to use our definition, since, just as in one dimension, it permits use of parallel edges and often enables a smaller presentation of a given constraint.

Figure 2.2 shows presentations for the NAK and RWIM constraints defined in Chapter 1. In these presentations \mathcal{G}_1 and \mathcal{G}_2 describe the vertical and horizontal constraints on the edges, respectively. Each edge $e = (e)_{i,j}$ is a 2×2 binary matrix of the form

$$e = \begin{pmatrix} (e)_{(0,0)} & (e)_{(0,1)} \\ (e)_{(1,0)} & (e)_{(1,1)} \end{pmatrix},$$

and it is labeled by $(e)_{(1,1)}$, i.e., the labeling of an edge simply picks out the entry in the lower-right corner. For NAK, the edges $E = E_{\text{NAK}}$ are the 2×2 matrices which satisfy NAK, that is, with at most one 1. Similarly, for RWIM, the edges $E = E_{\text{RWIM}}$ are the 2×2 matrices which satisfy RWIM, namely, the elements of E_{NAK} together with

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$

In the figures, each edge is drawn twice—once in \mathcal{G}_1 and once in \mathcal{G}_2 —and the matrix identifying it is written next to it. The states are 1×2 blocks for \mathcal{G}_1 and 2×1 blocks for \mathcal{G}_2 ; for an edge e ,

$$\sigma_{\mathcal{G}_1}(e) = (e)_{(0,0)}(e)_{(0,1)}, \text{ and } \tau_{\mathcal{G}_1}(e) = (e)_{(1,0)}(e)_{(1,1)},$$

and

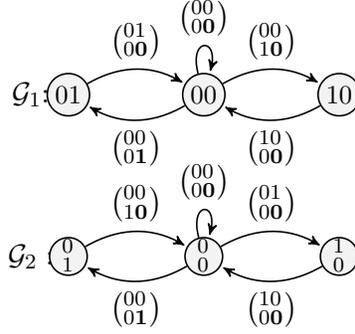
$$\sigma_{\mathcal{G}_2}(e) = \begin{pmatrix} (e)_{(0,0)} \\ (e)_{(1,0)} \end{pmatrix}, \text{ and } \tau_{\mathcal{G}_2}(e) = \begin{pmatrix} (e)_{(0,1)} \\ (e)_{(1,1)} \end{pmatrix}.$$

It follows that, for both constraints, and any rectangular array $\Gamma' \in E^{m \times n}$ with each of its rows a path in \mathcal{G}_2 and each of its columns a path in \mathcal{G}_1 , it holds that

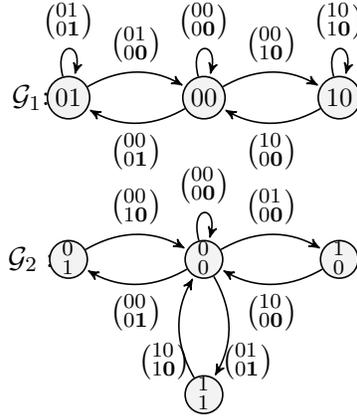
$$\Gamma'_{(i,j)} = \begin{pmatrix} \mathcal{L}(\Gamma'_{(i-1,j-1)}) & \mathcal{L}(\Gamma'_{(i-1,j)}) \\ \mathcal{L}(\Gamma'_{(i,j-1)}) & \mathcal{L}(\Gamma'_{(i,j)}) \end{pmatrix}$$

for $i = 1, 2, \dots, m-1$, $j = 1, 2, \dots, n-1$. Therefore, the only 2×2 sub-arrays appearing in the array $\mathcal{L}(\Gamma')$ are elements of E , and it follows that $\mathcal{L}(\Gamma')$ satisfies the corresponding constraint. Similarly, it can be shown that any array satisfying the constraint can be generated by the presentation.

2.3. Capacity



(a)



(b)

Figure 2.1: Presentations of 2-dimensional constraints: (a) NAK constraint; (b) RWIM constraint.

2.3 Capacity

In Chapter 1, we introduced the notion of capacity of a \mathfrak{D} -dimensional constraint. Here we expand on the definition; in particular we show that capacity is well-defined. We first need a generalization of Fekete's Subadditivity Lemma which will prove useful in subsequent chapters as well. We denote by \mathbb{N} the set of positive integers and by $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ the extended real numbers. We say that a function $f : \mathbb{N}^{\mathfrak{D}} \rightarrow \bar{\mathbb{R}}$ has a limit $L \in \bar{\mathbb{R}}$ at infinity, denoted $\lim_{\mathbf{m} \rightarrow \infty} f(\mathbf{m}) = L$, if for every sequence $(\mathbf{m}_i)_{i=1}^{\infty} \subseteq \mathbb{N}^{\mathfrak{D}}$ with $\mathbf{m}_i \rightarrow \infty$, we have $\lim_{i \rightarrow \infty} f(\mathbf{m}_i) = L$. This is equivalent to requiring that for any real $\epsilon > 0$ there exists $N_{\epsilon} \in \mathbb{N}$ such that for all $\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}$ with every entry at least N_{ϵ} , $|f(\mathbf{m}) - L| < \epsilon$. We call a function

2.3. Capacity

$f : \mathbb{N}^{\mathfrak{D}} \rightarrow \bar{\mathbb{R}}$ *entry-wise subadditive* if for all $\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}$, $f(\mathbf{m}) < \infty$, and for any $(m_1, \dots, m_{\mathfrak{D}}) \in \mathbb{N}^{\mathfrak{D}}$, $i \in \{1, 2, \dots, \mathfrak{D}\}$ and $n \in \mathbb{N}$, it holds that

$$f(m_1, \dots, m_{i-1}, m_i+n, m_{i+1}, \dots, m_{\mathfrak{D}}) \leq f(m_1, \dots, m_{\mathfrak{D}}) + f(m_1, \dots, m_{i-1}, n, m_{i+1}, \dots, m_{\mathfrak{D}}). \quad (2.1)$$

We call a function $f : \mathbb{N}^{\mathfrak{D}} \rightarrow \bar{\mathbb{R}}$ *entry-wise superadditive* if $-f$ is entry-wise subadditive.

Lemma 1. *Let $f : \mathbb{N}^{\mathfrak{D}} \rightarrow \bar{\mathbb{R}}$ be a function, then the following statements hold.*

1. *If f is entry-wise subadditive then*

$$\lim_{\mathbf{m} \rightarrow \infty} \frac{f(\mathbf{m})}{|\mathbf{m}|} = \inf_{\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}} \frac{f(\mathbf{m})}{|\mathbf{m}|}. \quad (2.2)$$

2. *If f is entry-wise superadditive then*

$$\lim_{\mathbf{m} \rightarrow \infty} \frac{f(\mathbf{m})}{|\mathbf{m}|} = \sup_{\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}} \frac{f(\mathbf{m})}{|\mathbf{m}|}. \quad (2.3)$$

Remark. Note that if f is entry-wise subadditive then for $i = 1, 2, \dots, \mathfrak{D}$, for any $m_{i+1}, \dots, m_{\mathfrak{D}} \in \mathbb{N}$ the mapping

$$m_i \rightarrow \lim_{m_{i-1} \rightarrow \infty} \lim_{m_{i-2} \rightarrow \infty} \dots \lim_{m_1 \rightarrow \infty} \frac{f(m_1, \dots, m_{\mathfrak{D}})}{m_1 \dots m_{i-1}}$$

is subadditive; and hence by Part 1 we have the limit

$$\begin{aligned} & \lim_{m_i \rightarrow \infty} \left(\frac{1}{m_i} \lim_{m_{i-1} \rightarrow \infty} \dots \lim_{m_1 \rightarrow \infty} \frac{f(m_1, \dots, m_{\mathfrak{D}})}{m_1 \dots m_{i-1}} \right) = \\ & \inf_{m_i \in \mathbb{N}} \left(\frac{1}{m_i} \lim_{m_{i-1} \rightarrow \infty} \dots \lim_{m_1 \rightarrow \infty} \frac{f(m_1, \dots, m_{\mathfrak{D}})}{m_1 \dots m_{i-1}} \right). \end{aligned}$$

By repeating this argument \mathfrak{D} times, one has,

$$\begin{aligned} \lim_{m_{\mathfrak{D}} \rightarrow \infty} \dots \lim_{m_1 \rightarrow \infty} \frac{f(m_1, \dots, m_{\mathfrak{D}})}{m_1 \dots m_{\mathfrak{D}}} &= \inf_{m_{\mathfrak{D}} \in \mathbb{N}} \dots \inf_{m_1 \in \mathbb{N}} \frac{f(m_1, \dots, m_{\mathfrak{D}})}{m_1 \dots m_{\mathfrak{D}}} \\ &= \inf_{\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}} \frac{f(\mathbf{m})}{|\mathbf{m}|} = \lim_{\mathbf{m} \rightarrow \infty} \frac{f(\mathbf{m})}{|\mathbf{m}|}. \end{aligned}$$

An analogous result holds for the superadditive case.

2.3. Capacity

Proof. Part 1. Let f be entry-wise subadditive. It's easy to check that (2.1) implies that for all $(a_1^{(0)}, \dots, a_{\mathfrak{D}}^{(0)}), (a_1^{(1)}, \dots, a_{\mathfrak{D}}^{(1)}) \in \mathbb{N}^{\mathfrak{D}}, k \in \mathbb{N}$, and $i \in \{1, \dots, \mathfrak{D}\}$, we have

$$f(a_1^{(0)}, \dots, a_{i-1}^{(0)}, ka_i^{(0)}, a_{i+1}^{(0)}, \dots, a_{\mathfrak{D}}^{(0)}) \leq kf(a_1^{(0)}, \dots, a_{\mathfrak{D}}^{(0)}) \quad (2.4)$$

$$f(a_1^{(0)} + a_1^{(1)}, \dots, a_{\mathfrak{D}}^{(0)} + a_{\mathfrak{D}}^{(1)}) \leq \sum_{\mathbf{x} \in \{0,1\}^{\mathfrak{D}}} f(a_1^{(x_1)}, \dots, a_{\mathfrak{D}}^{(x_{\mathfrak{D}})}), \quad (2.5)$$

where $\mathbf{x} = (x_1, \dots, x_{\mathfrak{D}})$ in the sum in the RHS of (2.5).

Now, let $(\mathbf{m}^{(i)})_{i=1}^{\infty} \subseteq \mathbb{N}^{\mathfrak{D}}$ be a sequence satisfying $\mathbf{m}^{(i)} \rightarrow \infty$. Write $\mathbf{m}^{(i)} = (m_1^{(i)}, \dots, m_{\mathfrak{D}}^{(i)})$, and let $\mathbf{n} = (n_1, \dots, n_{\mathfrak{D}})$ be a vector in $\mathbb{N}^{\mathfrak{D}}$. Since $\mathbf{m}^{(i)} \rightarrow \infty$ (and \mathfrak{D} is finite), there is an $i_0 = i_0(\mathbf{n})$ such that for all $i \geq i_0$, $\mathbf{m}^{(i)} \geq 2\mathbf{n}$. Set $M = \max(\{0\} \cup \{f(\mathbf{x}) : \mathbf{x} \in \prod_j \{1, \dots, n_j\}\})$ and let $i \geq i_0$. Note, that by our assumption on f , $M \neq \pm \infty$. For each $j = 1, \dots, \mathfrak{D}$, there are (unique) $q_j^{(i)}, r_j^{(i)} \in \mathbb{N}$, such that $m_j^{(i)} = q_j^{(i)} n_j + r_j^{(i)}$, and $1 \leq r_j^{(i)} \leq n_j$. We define $\alpha_j^{(0)} = \beta_j^{(0)} = r_j^{(i)}$, $\alpha_j^{(1)} = q_j^{(i)} n_j$, and $\beta_j^{(1)} = n_j$. Let $T = \{0, 1\}^{\mathfrak{D}} \setminus \{\mathbf{1}\}$ (where $\mathbf{1}$ denotes the vector in $\mathbb{N}^{\mathfrak{D}}$ with every entry equal to 1). Then using (2.5) and (2.4) we have

$$\begin{aligned} \frac{f(\mathbf{m}^{(i)})}{|\mathbf{m}^{(i)}|} &= \frac{f(q_1^{(i)} n_1 + r_1^{(i)}, \dots, q_{\mathfrak{D}}^{(i)} n_{\mathfrak{D}} + r_{\mathfrak{D}}^{(i)})}{|\mathbf{m}^{(i)}|} \\ &\leq \frac{\sum_{\mathbf{x} \in \{0,1\}^{\mathfrak{D}}} f(\alpha_1^{(x_1)}, \dots, \alpha_{\mathfrak{D}}^{(x_{\mathfrak{D}})})}{|\mathbf{m}^{(i)}|} \\ &\leq \frac{\sum_{\mathbf{x} \in \{0,1\}^{\mathfrak{D}}} \left(f(\beta_1^{(x_1)}, \dots, \beta_{\mathfrak{D}}^{(x_{\mathfrak{D}})}) \prod_j (q_j^{(i)})^{x_j} \right)}{|\mathbf{m}^{(i)}|} \\ &= \frac{f(\mathbf{n}) \prod_j q_j^{(i)} + \sum_{\mathbf{x} \in T} \left(f(\beta_1^{(x_1)}, \dots, \beta_{\mathfrak{D}}^{(x_{\mathfrak{D}})}) \prod_j (q_j^{(i)})^{x_j} \right)}{|\mathbf{m}^{(i)}|} \\ &\leq f(\mathbf{n}) \prod_j \frac{q_j^{(i)}}{q_j^{(i)} n_j + r_j^{(i)}} + \sum_{\mathbf{x} \in T} \left(M \prod_j \frac{(q_j^{(i)})^{x_j}}{q_j^{(i)} n_j + r_j^{(i)}} \right), \quad (2.6) \end{aligned}$$

where, again, in each sum above $\mathbf{x} = (x_1, \dots, x_{\mathfrak{D}})$. Since $\mathbf{m}^{(i)} \rightarrow \infty$, it follows that $\lim_{i \rightarrow \infty} q_j^{(i)} = \infty$ for each $j=1, \dots, \mathfrak{D}$. Observe that, as i goes to infinity, the first summand of the RHS of (2.6) converges to $f(\mathbf{n})/|\mathbf{n}|$, and each of the other summands converges to 0. Hence, taking the lim sup of both sides of the last inequality, we obtain $\limsup_{i \rightarrow \infty} (f(\mathbf{m}^{(i)})/|\mathbf{m}^{(i)}|) \leq f(\mathbf{n})/|\mathbf{n}|$. Since \mathbf{n} is arbitrary, this implies

$$\limsup_{i \rightarrow \infty} \frac{f(\mathbf{m}^{(i)})}{|\mathbf{m}^{(i)}|} \leq \inf_{\mathbf{n} \in \mathbb{N}^{\mathfrak{D}}} \frac{f(\mathbf{n})}{|\mathbf{n}|}. \quad (2.7)$$

2.4. Axial product

On the other hand, clearly,

$$\inf_{\mathbf{n} \in \mathbb{N}^{\mathfrak{D}}} \frac{f(\mathbf{n})}{|\mathbf{n}|} \leq \liminf_{i \rightarrow \infty} \frac{f(\mathbf{m}^{(i)})}{|\mathbf{m}^{(i)}|} \leq \limsup_{i \rightarrow \infty} \frac{f(\mathbf{m}^{(i)})}{|\mathbf{m}^{(i)}|}.$$

Combining this with (2.7) we get $\lim_{i \rightarrow \infty} (f(\mathbf{m}^{(i)})/|\mathbf{m}^{(i)}|) = \inf_{\mathbf{n}} \{f(\mathbf{n})/|\mathbf{n}|\}$, and the result follows.

Part 2. Easily follows by applying Part 1 to $-f$. ■

Observe that \mathfrak{D} -dimensional constraints are closed under taking contiguous sub-arrays, meaning that if an array belongs to the constraint then any of its \mathfrak{D} -dimensional contiguous sub-arrays also belongs to the constraint. This implies that the mapping $\mathbf{m} \rightarrow \log |S_{\mathbf{m}}|$ for $\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}$, where we define $\log 0 = -\infty$, is entry-wise subadditive. Thus by Lemma 1, the limit in (1.1) always exists, is independent of the choice of $(\mathbf{m}_i)_{i=1}^{\infty}$, and satisfies

$$\text{cap}(S) = \inf_{\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}} \frac{\log |S_{\mathbf{m}}|}{|\mathbf{m}|} \quad (2.8)$$

For a nonnegative matrix A denote by $\lambda(A)$ its Perron eigenvalue, that is, its largest real eigenvalue. It is well-known that for a 1-dimensional constraint S presented by a lossless labeled graph $\mathcal{G} = (G, \mathcal{L})$, the capacity of S is $\log \lambda(A(G))$. In particular, for a graph G , it holds that $\text{cap}(X(G)) = \log \lambda(A(G))$.

2.4 Axial product

The *axial product* of \mathfrak{D} sets $L_1, \dots, L_{\mathfrak{D}} \subseteq \Sigma^*$, denoted $L_1 \otimes L_2 \otimes \dots \otimes L_{\mathfrak{D}} \subseteq \Sigma^{*\mathfrak{D}}$, is the set of all arrays $\Gamma \in \Sigma^{*\mathfrak{D}}$ such that for $i = 1, 2, \dots, \mathfrak{D}$ every row of Γ in direction i belongs to L_i . If $L_1 = L_2 = \dots = L_{\mathfrak{D}} = L$ we say that the axial-product is *isotropic* and denote it by $L^{\otimes \mathfrak{D}}$. Given a presentation $\bar{\mathcal{G}} = ((G_1, \mathcal{L}), \dots, (G_{\mathfrak{D}}, \mathcal{L}))$ for a \mathfrak{D} -dimensional constraint S with a common set of edges E , the set $X(G_1) \otimes \dots \otimes X(G_{\mathfrak{D}}) \subseteq E^{*\mathfrak{D}}$ is a \mathfrak{D} -dimensional constraint presented by $((G_1, I_E), \dots, (G_{\mathfrak{D}}, I_E))$, where I_E is the identity map on E .

If $\text{cap}(X(G_1) \otimes \dots \otimes X(G_{\mathfrak{D}})) = \text{cap}(S)$, we say that $\bar{\mathcal{G}}$ is *capacity-preserving*. For $\mathfrak{D}=1$ any lossless, and therefore deterministic, presentation of S is capacity-preserving, since in a lossless graph with $|V|$ vertices there are at most $|V|^2$ paths generating any given word; for $\mathfrak{D} > 1$, the question whether every \mathfrak{D} -dimensional constraint has a capacity-preserving presentation is open. This is a major open problem in symbolic dynamics, although it is usually formulated in a slightly different manner; see [6], where it is shown that for every \mathfrak{D} -dimensional constraint S and $\epsilon > 0$, there is a presentation $\bar{\mathcal{G}} = ((G_1, \mathcal{L}), \dots, (G_{\mathfrak{D}}, \mathcal{L}))$ such

2.4. Axial product

that $\text{cap}(S) \leq \text{cap}(X(G_1) \otimes \dots \otimes X(G_{\mathfrak{D}})) < \text{cap}(S) + \epsilon$. We show in the next proposition that the answer to this question is positive, if S is an axial product of \mathfrak{D} one-dimensional constraints.

Proposition 1. *Let $S^{(1)}, S^{(2)}, \dots, S^{(\mathfrak{D})} \subseteq \Sigma^*$ be \mathfrak{D} one-dimensional constraints over Σ , and let $S = S^{(1)} \otimes \dots \otimes S^{(\mathfrak{D})}$, then S is a \mathfrak{D} -dimensional constraint over Σ . Moreover, S has a capacity-preserving presentation.*

Remark . There are \mathfrak{D} -dimensional constraints which are not axial products of \mathfrak{D} one-dimensional constraints. For example for $\mathfrak{D}=2$, the NAK constraint defined in Chapter 1 is not an axial product.

Proof. Let $\mathcal{G}_{S^{(1)}}, \dots, \mathcal{G}_{S^{(\mathfrak{D})}}$ be presentations of $S^{(1)}, \dots, S^{(\mathfrak{D})}$, where, for $i = 1, 2, \dots, \mathfrak{D}$, $\mathcal{G}_{S^{(i)}} = ((V_i, E_i), \mathcal{L}_i)$. Define the \mathfrak{D} -tuple of labeled graphs $\bar{\mathcal{G}} = (\mathcal{G}_1, \dots, \mathcal{G}_{\mathfrak{D}})$, as follows. Let

$$E = \left\{ (e_1, \dots, e_{\mathfrak{D}}) \in \prod_{i=1}^{\mathfrak{D}} E_i : \mathcal{L}_1(e_1) = \mathcal{L}_2(e_2) = \dots = \mathcal{L}_{\mathfrak{D}}(e_{\mathfrak{D}}) \right\},$$

and let $\mathcal{L} : E \rightarrow \Sigma$ be given by

$$\mathcal{L}(e_1, \dots, e_{\mathfrak{D}}) = \mathcal{L}_1(e_1) \ ; \ (e_1, \dots, e_{\mathfrak{D}}) \in E.$$

For $i = 1, 2, \dots, \mathfrak{D}$, the graph \mathcal{G}_i is defined by $\mathcal{G}_i = (G_i, \mathcal{L})$ with $G_i = (V_i, E)$, where for $\mathbf{e} = (e_1, \dots, e_{\mathfrak{D}}) \in E$, $\sigma_{\mathcal{G}_i}(\mathbf{e}) = \sigma_{\mathcal{G}_{S^{(i)}}}(e_i)$ and $\tau_{\mathcal{G}_i}(\mathbf{e}) = \tau_{\mathcal{G}_{S^{(i)}}}(e_i)$. It's easy to verify that $\bar{\mathcal{G}}$ is a presentation of $S^{(1)} \otimes S^{(2)} \otimes \dots \otimes S^{(\mathfrak{D})}$.

Assume now that every $\mathcal{G}_{S^{(i)}}$ is lossless. We show that in this case $\bar{\mathcal{G}}$ is capacity preserving. Let $X = X(G_1) \otimes \dots \otimes X(G_{\mathfrak{D}})$, n be a positive integer, and let \mathbf{n} be the \mathfrak{D} -tuple with every entry equal to n . We extend the mapping \mathcal{L} to $\mathcal{L} : X_{\mathbf{n}} \rightarrow S_{\mathbf{n}}$ as described above. Now, fix an array $\Gamma \in S_{\mathbf{n}}$, and for $i = 1, 2, \dots, \mathfrak{D}$ let $\Gamma'^{(i)} \in (E_i)^{\mathbf{n}}$ be an array such that every row in direction i , $(\Gamma'_{\mathbf{j}k})_{k=1}^n$ is a path in $\mathcal{G}_{S^{(i)}}$ generating the corresponding row $(\Gamma_{\mathbf{j}k})_{k=1}^n$ in Γ . Let $\Gamma' \in E^{\mathbf{n}}$ be the array with entries given by $\Gamma'_{\mathbf{j}} = (\Gamma'_{\mathbf{j}}^{(1)}, \dots, \Gamma'_{\mathbf{j}}^{(\mathfrak{D})})$, $\mathbf{j} \in [n]^{\mathfrak{D}}$. It follows from the construction of $\bar{\mathcal{G}}$ that $\Gamma' \in X_{\mathbf{n}}$, and that $\mathcal{L}(\Gamma') = \Gamma$. Moreover, any array $\Delta \in X_{\mathbf{n}}$ such that $\mathcal{L}(\Delta) = \Gamma$ can be constructed in this manner. Now, as each $\mathcal{G}_{S^{(i)}}$ is lossless, there are at most $|V_i|^2$ possibilities of choosing each row in direction i of $\Gamma'^{(i)}$, and as there are $n^{\mathfrak{D}-1}$ such rows, there are at most $|V_i|^{2n^{\mathfrak{D}-1}}$ possibilities of choosing each $\Gamma'^{(i)}$. It follows that

$$|\mathcal{L}^{-1}(\{\Gamma\})| \leq \prod_{i=1}^{\mathfrak{D}} |V_i|^{2n^{\mathfrak{D}-1}},$$

2.5. Two-dimensional constraints

where $\mathcal{L}^{-1}(\{\Gamma\}) = \{\Gamma' \in X_{\mathbf{n}} : \mathcal{L}(\Gamma') = \Gamma\}$. Summing the latter inequality over all $\Gamma \in \mathcal{S}_{\mathbf{n}}$, we obtain

$$|X_{\mathbf{n}}| \leq |S_{\mathbf{n}}| \prod_{i=1}^{\mathfrak{D}} |V_i|^{2n^{\mathfrak{D}-1}}.$$

Taking the log, dividing through by $n^{\mathfrak{D}}$, and taking the limit as n approaches infinity, we have $\text{cap}(X) \leq \text{cap}(S)$. Clearly, $\text{cap}(X) \geq \text{cap}(S)$, since X is a presentation of S . The result follows. \blacksquare

Let S be a 1-dimensional constraint. Since for any $\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}$ $|(S^{\otimes \mathfrak{D}})_{\mathbf{m}}| = |(S^{\otimes (\mathfrak{D}+1)})_{(\mathbf{m},1)}|$ it follows from (2.8) that $\text{cap}(S^{\otimes \mathfrak{D}})$ is non-increasing in \mathfrak{D} .

2.5 Two-dimensional constraints

We introduce some specialized definitions for 2-dimensional constraints. For a 2-dimensional array $\Gamma \in \Sigma^{m_1 \times m_2}$, for nonnegative integers m_1, m_2 , we denote by $\Gamma^{\mathfrak{t}}$ its transpose, namely $(\Gamma^{\mathfrak{t}})_{(i,j)} = (\Gamma)_{(j,i)}$ for all $(i,j) \in [m_1] \times [m_2]$. For a 2-dimensional constraint S over Σ , we use $S^{\mathfrak{t}}$ to denote the set

$$S^{\mathfrak{t}} = \{\Gamma \in \Sigma^{**} : \Gamma^{\mathfrak{t}} \in S\},$$

Clearly $S^{\mathfrak{t}}$ is a 2-dimensional constraint with $\text{cap}(S^{\mathfrak{t}}) = \text{cap}(S)$.

We shall use the following (MATLAB-like) notation for 2-dimensional arrays. Let \mathcal{A} be a set and $\Gamma \in \mathcal{A}^{s \times t}$. For integers $0 \leq s_1 \leq s_2 < s$ and $0 \leq t_1 \leq t_2 < t$, we denote by $\Gamma_{s_1:s_2, t_1:t_2}$ the sub-array:

$$(\Gamma_{s_1:s_2, t_1:t_2})_{i,j} = \Gamma_{s_1+i, t_1+j}; \quad (i,j) \in [s_2-s_1+1] \times [t_2-t_1+1],$$

and by $\Gamma_{s_1:s_2, *}$ (resp. $\Gamma_{*, t_1:t_2}$) the sub-array $\Gamma_{s_1:s_2, 0:t-1}$ (resp. $\Gamma_{0:s-1, t_1:t_2}$). We also abbreviate $x:x$ in the subscript by x . We shall use the same notation for one-dimensional vectors: for a vector $\mathbf{v} \in \mathcal{A}^s$, $\mathbf{v}_{s:t}$ denotes the sub-vector

$$(\mathbf{v}_{s:t})_i = \mathbf{v}_{s+i}; \quad i \in [t-s+1].$$

2.5.1 Horizontal and vertical strips

Let S be a 2-dimensional constraint over Σ , and m be a positive integer. The *horizontal* (resp. *vertical*) *strip of height* (resp. *width*) m of S , denoted $\mathcal{H}_m(S)$ (resp. $\mathcal{V}_m(S)$) is the subset of S given by

$$\mathcal{H}_m(S) = \bigcup_n S_{m \times n} \quad (\text{resp. } \mathcal{V}_m(S) = \bigcup_n S_{n \times m}).$$

2.5. Two-dimensional constraints

Let S be a 2-dimensional constraint over an alphabet Σ , and consider a horizontal (resp. vertical) strip $\mathcal{H}_m(S)$ (resp. $\mathcal{V}_m(S)$) of S for some positive integer m . We regard such a strip as a set of 1-dimensional words over Σ^m where each $m \times n$ (resp. $n \times m$) array in the strip is considered a word of length n over Σ^m . Below we show that the horizontal and vertical strips of S are 1-dimensional constraints over Σ^m . For this, we need the following definition. Let $G = (V, E)$ be a graph, and let m be a positive integer. Let $G^{\times m}$ be the graph given by $G^{\times m} = (V^m, E^m)$, where for each $\mathbf{e} = (e_1, \dots, e_m) \in E^m$, $\sigma_{G^{\times m}}(\mathbf{e}) = (\sigma_G(e_1), \dots, \sigma_G(e_m))$ and $\tau_{G^{\times m}}(\mathbf{e}) = (\tau_G(e_1), \dots, \tau_G(e_m))$. For a labeled graph $\mathcal{G} = (G, \mathcal{L})$ with $G = (V, E)$ and $\mathcal{L} : E \rightarrow \Sigma$, let $\mathcal{G}^{\times m}$ be the labeled graph defined by $\mathcal{G}^{\times m} = (G^{\times m}, \mathcal{L}^{\times m})$, where $\mathcal{L}^{\times m} : E^m \rightarrow \Sigma^m$ is given by

$$\mathcal{L}^{\times m}(e_1, \dots, e_m) = (\mathcal{L}(e_1), \dots, \mathcal{L}(e_m)); (e_1, \dots, e_m) \in E^m$$

We call $G^{\times m}$ (resp., $\mathcal{G}^{\times m}$) the m th tensor-power of G (resp., \mathcal{G}). We can now state the following proposition.

Proposition 2. *Let S be a 2-dimensional constraint over Σ and let m be a positive integer. Then*

1. $\mathcal{H}_m(S)$ (resp. $\mathcal{V}_m(S)$) is a 1-dimensional constraint over Σ^m .
2. Let $S = T^{(\mathcal{V})} \otimes T^{(\mathcal{H})}$ for 1-dimensional constraints $T^{(\mathcal{V})}, T^{(\mathcal{H})}$ over Σ , presented by labeled graphs $\mathcal{G}^{(\mathcal{V})}, \mathcal{G}^{(\mathcal{H})}$, respectively. Then the 1-dimensional constraint $\mathcal{H}_m(S)$ is presented by the labeled graph $\mathcal{G}_m^{(\mathcal{H})}$ defined as the subgraph of the labeled graph $(\mathcal{G}^{(\mathcal{H})})^{\times m}$ consisting of only those edges whose label (an m -letter word over Σ) satisfies $T^{(\mathcal{V})}$. An analogous statement holds for $\mathcal{V}_m(S)$, with respect to the graph $\mathcal{G}_m^{(\mathcal{V})}$ formed in a similar way from $(\mathcal{G}^{(\mathcal{V})})^{\times m}$.

Proof. It suffices to prove this only for horizontal strips $\mathcal{H}_m(S)$ —the argument for the vertical strip being analogous. We first prove part 2. It's easy to verify that the labeled graph $(\mathcal{G}^{(\mathcal{H})})^{\times m}$ presents the constraint over Σ^m , consisting of all $m \times n$ arrays of Σ^{**} such that every row satisfies $T^{(\mathcal{H})}$. It follows that the subgraph $\mathcal{G}_m^{(\mathcal{H})}$, formed by removing all the edges of $(\mathcal{G}^{(\mathcal{H})})^{\times m}$ that are labeled with a word that does not satisfy $T^{(\mathcal{V})}$, presents the 1-dimensional constraint consisting of all $m \times n$ arrays with every row satisfying $T^{(\mathcal{H})}$ and every column satisfying $T^{(\mathcal{V})}$. This is precisely the constraint $\mathcal{H}_m(S)$.

We proceed to prove part 1. Let the pair of labeled graphs $(\mathcal{G}^{(\mathcal{V})}, \mathcal{G}^{(\mathcal{H})})$ be a presentation of S , where $\mathcal{G}^{(\mathcal{V})} = ((V^{(\mathcal{V})}, E), \mathcal{L})$ and $\mathcal{G}^{(\mathcal{H})} = ((V^{(\mathcal{H})}, E), \mathcal{L})$. Define the edge-constraints $\mathcal{E}^{(\mathcal{V})} = \mathbf{X}(V^{(\mathcal{V})}, E)$ and $\mathcal{E}^{(\mathcal{H})} = \mathbf{X}(V^{(\mathcal{H})}, E)$. Since $(\mathcal{G}^{(\mathcal{V})}, \mathcal{G}^{(\mathcal{H})})$ is a presentation of S , we have $S = \mathcal{L}(\mathcal{E}^{(\mathcal{V})} \otimes \mathcal{E}^{(\mathcal{H})})$, and therefore

$\mathcal{H}_m(S) = \mathcal{L}(\mathcal{H}_m(\mathcal{E}^{(\mathcal{V})} \otimes \mathcal{E}^{(\mathcal{H})}))$. By part 2, $\mathcal{H}_m(\mathcal{E}^{(\mathcal{V})} \otimes \mathcal{E}^{(\mathcal{H})})$ is a 1-dimensional constraint, presented by a labeled graph $\mathcal{G}_m^{(\mathcal{H})}$ with edges labeled by words in E^m . Replacing each such label $\mathbf{e} \in E^m$ in that graph with $\mathcal{L}(\mathbf{e})$, we clearly obtain a presentation of $\mathcal{L}(\mathcal{H}_m(\mathcal{E}^{(\mathcal{V})} \otimes \mathcal{E}^{(\mathcal{H})})) = \mathcal{H}_m(S)$. ■

2.6 Open questions

A major problem in the theory of multidimensional constrained systems is the computation of their capacity. As already mentioned before, the problem is essentially solved for 1-dimensional constraints, however, in higher dimensions, only a few methods for estimating the capacity of a general \mathfrak{D} -dimensional constraint exist. In the light of [2], no “computable” formula for the capacity of such constraints exists; yet, it still may be true that for some specialized sub-classes of constraints, one can compute the capacity exactly. For example, the hard square constraint is a relatively old open problem for which no “closed-form” formula for computing the capacity is currently known. However, such a formula is known for the hexagonal-lattice version of this constraint (the “hard-hexagon constraint”) [1], which perhaps suggests that a similar formula exists for the capacity of the hard-square constraint.

Chapter 3

Lower bounds on capacity of 2-dimensional symmetric constraints*

A method for computing very good lower-bounds on the capacity of the hard-square constraint is given in [7] (see also [34, 41]). [3] generalizes the method slightly and also presents a method for obtaining good upper-bounds on the capacity of this constraint. Both the method for obtaining the lower- and the method for obtaining the upper-bounds can be shown to work on any 2-dimensional constraint for which every horizontal or every vertical strip is a symmetric vertex-constraint. In this chapter we show a generalization of the method for obtaining the lower-bounds that gives improved bounds on capacities of such constraints. Moreover, we show how this generalization as well as the method for obtaining the upper-bounds may be applied to a larger class of 2-dimensional constraints that includes constraints in which the vertical and horizontal strips are not necessarily finite-type. We illustrate this by computing lower and upper bounds on the capacities of the $\text{EVEN}^{\otimes 2}$ and $\text{CHG}(3)^{\otimes 2}$ constraints, and show that

$$\begin{aligned} 0.4402086447 &\leq \text{cap}(\text{EVEN}^{\otimes 2}) \leq 0.4452873312, & \text{and} \\ 0.4222689819 &\leq \text{cap}(\text{CHG}(3)^{\otimes 2}) \leq 0.5328488954. \end{aligned}$$

3.1 Constraints with symmetric edge-constrained strips

In this section we generalize the method presented in [3, 7] to provide improved lower bounds on capacities of 2-dimensional constraints whose horizontal strips are symmetric edge-constraints.

Fix an alphabet Σ , and let S be a 2-dimensional constraint over Σ . We say that S has *horizontal edge-constrained-strips* if for every positive integer m , the

*A version of this chapter has been published. Louidor, E. and Marcus, B.H. (2010) Improved Lower Bounds on Capacities of Symmetric 2-Dimensional Constraints using Rayleigh Quotients. IEEE Transactions on Information Theory 56:1624–1639.

3.1. Constraints with symmetric edge-constrained strips

constraint $\mathcal{H}_m(S)$ is an edge-constraint. If, in addition, every horizontal strip is symmetric, we say that S has *symmetric horizontal edge-constrained strips*. Analogously, using $\mathcal{V}_m(S)$, we have the notions of a 2-dimensional constraint with vertical edge-constrained-strips and symmetric vertical edge-constrained-strips

Here, we consider constraints of the form $S = T \otimes \mathcal{E}$, where $\mathcal{E} = \mathcal{X}(G_{\mathcal{E}})$ is an edge-constraint defined by the graph $G_{\mathcal{E}} = (V_{\mathcal{E}}, E_{\mathcal{E}})$ and T is an arbitrary 1-dimensional constraint over Σ . Then \mathcal{E} is presented by $\mathcal{G}_{\mathcal{E}} = (G_{\mathcal{E}}, I_E)$ where I_E is the identity map on $E_{\mathcal{E}}$. Let m be a positive integer. By Proposition 2, part 2, $\mathcal{H}_m(S)$ is a 1-dimensional constraint presented by a subgraph $\mathcal{G}_m^{(\mathcal{H})} = (G_m^{(\mathcal{H})}, I_E^{\times m})$ of $\mathcal{G}_{\mathcal{E}}^{\times m}$. It follows that $\mathcal{H}_m(S) = \mathcal{X}(G_m^{(\mathcal{H})})$, and so S has horizontal edge-constrained strips. Henceforth, we further assume that it has symmetric horizontal edge-constrained strips; note that symmetry of the graph $G_{\mathcal{E}}$ is necessary but not sufficient for this assumption (see Proposition 3 below).

For a positive integer m , let $F_m = |(V_{\mathcal{E}})^m|$, and let H_m denote the $F_m \times F_m$ adjacency matrix of $G_m^{(\mathcal{H})}$. Since $\lim_{n \rightarrow \infty} (\log(|S_{m \times n}|)/n) = \text{cap}(\mathcal{H}_m(S))$ for every positive integer m , we have

$$\begin{aligned} \text{cap}(S) &= \lim_{m, n \rightarrow \infty} \frac{\log |S_{m \times n}|}{mn} \\ &= \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\log |S_{m \times n}|}{mn} \\ &= \lim_{m \rightarrow \infty} \frac{\text{cap}(\mathcal{H}_m(S))}{m} \\ &= \lim_{m \rightarrow \infty} \frac{\log \lambda(H_m)}{m}. \end{aligned} \tag{3.1}$$

For a matrix M , let M^{t} denote its transpose. Fix a positive integer m . Following [3, 7], since H_m is real and symmetric, we obtain by the min-max principle [16]

$$\lambda(H_m^p) \geq \frac{\mathbf{y}_m^{\text{t}} H_m^p \mathbf{y}_m}{\mathbf{y}_m^{\text{t}} \mathbf{y}_m},$$

for any $F_m \times 1$ real vector $\mathbf{y}_m \neq \mathbf{0}$ and positive integer p . The RHS of the last inequality is known as a *Rayleigh quotient*. Choosing \mathbf{y}_m to be the vector $H_m^q \mathbf{x}_m$, for some positive integer q and $F_m \times 1$ real vector \mathbf{x}_m such that $\mathbf{y}_m \neq \mathbf{0}$, we have

$$\lambda(H_m^p) \geq \frac{\mathbf{x}_m^{\text{t}} H_m^{2q+p} \mathbf{x}_m}{\mathbf{x}_m^{\text{t}} H_m^{2q} \mathbf{x}_m}. \tag{3.2}$$

Thus by (3.1), it follows that

$$\text{cap}(S) \geq \frac{1}{p} \limsup_{m \rightarrow \infty} \frac{1}{m} \log \frac{\mathbf{x}_m^{\text{t}} H_m^{2q+p} \mathbf{x}_m}{\mathbf{x}_m^{\text{t}} H_m^{2q} \mathbf{x}_m}. \tag{3.3}$$

3.1. Constraints with symmetric edge-constrained strips

In [3, 7], each \mathbf{x}_m is chosen to be the vector $\mathbf{1}$ with F_m entries. We obtain improved lower bounds in many cases by choosing other sequences of vectors, $(\mathbf{x}_m)_{m=1}^\infty$, as follows. We fix integers $\mu \geq 0$ and $\alpha \geq 1$, and let $\phi : (V_{\mathcal{E}})^{\mu+\alpha} \rightarrow [0, \infty)$ be a nonnegative function. Our method works for sequences $(\mathbf{x}_{m_k})_{k=1}^\infty$, where $m_k = \mu + k\alpha$ for positive integers k , and the entries of each \mathbf{x}_{m_k} are indexed by $(V_{\mathcal{E}})^{m_k}$ and given by

$$(\mathbf{x}_{m_k})_{\mathbf{v}} = \prod_{i=0}^{k-1} \phi(\mathbf{v}_{i\alpha:i\alpha+\mu+\alpha-1}) \quad ; \quad \mathbf{v} \in (V_{\mathcal{E}})^{m_k}. \quad (3.4)$$

For such sequences and a fixed positive integer n , we will show that one can compute L_n , the growth rate of $\mathbf{x}_{m_k}^t H_{m_k}^n \mathbf{x}_{m_k}$:

$$L_n = \lim_{k \rightarrow \infty} \frac{\log \mathbf{x}_{m_k}^t H_{m_k}^n \mathbf{x}_{m_k}}{m_k},$$

and from (3.3) we obtain the lower bound $\text{cap}(S) \geq (L_{2q+p} - L_{2q})/p$.

Before doing this for general μ and α , it is instructive to look at the special case: $\mu = 0$ and $\alpha = 1$. In this case $m_k = k$, and using (3.4) to define \mathbf{x}_k we obtain

$$(\mathbf{x}_k)_{\mathbf{v}} = \prod_{i=0}^{k-1} \phi((\mathbf{v})_i) \quad ; \quad \mathbf{v} \in (V_{\mathcal{E}})^k.$$

Let n be a positive integer. For a word $w = w_1 \dots w_n \in \mathcal{E}$ define its weight, $\mathcal{W}_\phi(w)$, by $\mathcal{W}_\phi(w) = \phi(\sigma(w_1))\phi(\tau(w_n))$, where w_1, w_n are regarded as edges in $G_{\mathcal{E}}$ and extend this to arrays $\Gamma \in S_{m \times n}$ by

$$\mathcal{W}_\phi(\Gamma) = \prod_{i=0}^{m-1} \mathcal{W}_\phi(\Gamma_{i,*}).$$

Observe that for an array $\Gamma \in S_{m \times n}$ that is a path in $G_m^{(\mathcal{H})}$ of length n starting at $\mathbf{v} \in (V_{\mathcal{E}})^m$ and ending at $\mathbf{u} \in (V_{\mathcal{E}})^m$, it holds that $\mathcal{W}_\phi(\Gamma) = (\mathbf{x}_m)_{\mathbf{v}} (\mathbf{x}_m)_{\mathbf{u}}$. It follows that

$$\mathbf{x}_m^t H_m^n \mathbf{x}_m = \sum_{\Gamma \in S_{m \times n}} \mathcal{W}_\phi(\Gamma). \quad (3.5)$$

Now, pick a deterministic presentation, $\mathcal{G}_n^{(\mathcal{V})}$, of $\mathcal{V}_n(S)$, where $\mathcal{G}_n^{(\mathcal{V})} = ((V_n^{(\mathcal{V})}, E_n^{(\mathcal{V})}), \mathcal{L}_n^{(\mathcal{V})})$, and let $\mathcal{W}_\phi : E_n^{(\mathcal{V})} \rightarrow [0, \infty)$ be the edge weighting defined by $\mathcal{W}_\phi(e) = \mathcal{W}_\phi(\mathcal{L}_n^{(\mathcal{V})}(e))$, for $e \in E_n^{(\mathcal{V})}$. Let $A(\mathcal{G}_n^{(\mathcal{V})}, \mathcal{W}_\phi)$ be the $|V_n^{(\mathcal{V})}| \times |V_n^{(\mathcal{V})}|$

3.1. Constraints with symmetric edge-constrained strips

weighted adjacency matrix of $\mathcal{G}_n^{(\mathcal{V})}$ with entries indexed by $V_n^{(\mathcal{V})} \times V_n^{(\mathcal{V})}$ and given by

$$\left(A(\mathcal{G}_n^{(\mathcal{V})}, \mathcal{W}_\phi) \right)_{i,j} = \sum_{\substack{e \in E_n^{(\mathcal{V})}: \\ \sigma(e)=i, \tau(e)=j}} \mathcal{W}_\phi(e) \quad ; \quad i, j \in V_n^{(\mathcal{V})}.$$

Then

$$\mathbf{1}^\dagger A(\mathcal{G}_n^{(\mathcal{V})}, \mathcal{W}_\phi)^m \mathbf{1} = \sum_{\gamma} \mathcal{W}_\phi(\mathcal{L}_n^{(\mathcal{V})}(\gamma)),$$

where the sum is taken over all paths γ in $\mathcal{G}_n^{(\mathcal{V})}$ of length m and $\mathcal{L}_n^{(\mathcal{V})}(\gamma)$ denotes the array in $S_{m \times n}$ generated by γ . Since $\mathcal{G}_n^{(\mathcal{V})}$ is deterministic it follows that

$$\lim_{m \rightarrow \infty} \frac{\log \mathbf{1}^\dagger A(\mathcal{G}_n^{(\mathcal{V})}, \mathcal{W}_\phi)^m \mathbf{1}}{m} = \lim_{m \rightarrow \infty} \frac{\log \sum_{\Gamma \in S_{m \times n}} \mathcal{W}_\phi(\Gamma)}{m}.$$

By (3.5) the RHS is L_n and by Perron-Frobenius theory the LHS is $\log \lambda(A(\mathcal{G}_n^{(\mathcal{V})}, \mathcal{W}_\phi))$, and thus $L_n = \log \lambda(A(\mathcal{G}_n^{(\mathcal{V})}, \mathcal{W}_\phi))$.

For general μ, α , we proceed similarly. We pick a deterministic presentation, $\mathcal{G}_n^{(\mathcal{V})}$, of the vertical strip $\mathcal{V}_n(S)$, with $\mathcal{G}_n^{(\mathcal{V})} = ((V_n^{(\mathcal{V})}, E_n^{(\mathcal{V})}), \mathcal{L}_n^{(\mathcal{V})})$, and construct a labeled directed graph $\mathcal{I} = \mathcal{I}(\mu, \alpha, n, \mathcal{G}_n^{(\mathcal{V})}, G_\mathcal{E}) = ((V_\mathcal{I}, E_\mathcal{I}), \mathcal{L}_\mathcal{I})$, with nonnegative real weights on its edges given by $\mathcal{W}_\phi : E_\mathcal{I} \rightarrow [0, \infty)$. The graph \mathcal{I} and weight function \mathcal{W}_ϕ are defined as follows. The set of vertices $V_\mathcal{I}$ is given by

$$V_\mathcal{I} = \left\{ (\mathbf{f}, v, \mathbf{l}) : v \in V_n^{(\mathcal{V})}, \mathbf{f}, \mathbf{l} \in (V_\mathcal{E})^\mu \right\},$$

and the function $\mathcal{L}_\mathcal{I} : E_\mathcal{I} \rightarrow \Sigma^{\alpha \times n}$ labels each edge with an $\alpha \times n$ array over Σ . We specify the edges of \mathcal{I} by describing the outgoing edges of each of its vertices along with their weights. Let $\mathbf{v} = (\mathbf{f}, v, \mathbf{l}) \in V_\mathcal{I}$ be a vertex of \mathcal{I} . The set of outgoing edges of \mathbf{v} consists of exactly one edge for every path of length α in $\mathcal{G}_n^{(\mathcal{V})}$ starting at v . Let $\gamma = (e_i)_{i=0}^{\alpha-1} \subseteq E_n^{(\mathcal{V})}$ be such a path and let u be its terminating vertex. We regard the word generated by γ in $\mathcal{G}_n^{(\mathcal{V})}$ as an array $\Gamma \in \Sigma^{\alpha \times n}$ with entries given by $(\Gamma)_{i,j} = (\mathcal{L}_n^{(\mathcal{V})}(e_i))_j$. Let $\mathbf{f} = (f_0, \dots, f_{\mu-1})$ and $\mathbf{l} = (l_0, \dots, l_{\mu-1})$ and for $i = \mu, \mu+1, \dots, \mu+\alpha-1$, define f_i to be $\sigma(\Gamma_{i-\mu,0})$ and l_i to be $\tau(\Gamma_{i-\mu, n-1})$, where $\Gamma_{i-\mu,0}$ and $\Gamma_{i-\mu, n-1}$ are regarded as edges in the graph $G_\mathcal{E}$. For such a path γ the corresponding outgoing edge $e \in E_\mathcal{I}$ of \mathbf{v} satisfies $\sigma(e) = \mathbf{v}$, $\mathcal{L}_\mathcal{I}(e) = \Gamma$, $\tau(e) = ((f_\alpha, f_{\alpha+1}, \dots, f_{\mu+\alpha-1}), u, (l_\alpha, l_{\alpha+1}, \dots, l_{\mu+\alpha-1}))$. The weight of e , $\mathcal{W}_\phi(e)$, is given by $\mathcal{W}_\phi(e) = \phi(f_0, \dots, f_{\mu+\alpha-1}) \phi(l_0, \dots, l_{\mu+\alpha-1})$. We shall regard the label of a path $(e_i)_{i=0}^{\ell-1}$ in \mathcal{I} as the $\ell \alpha \times n$ array Γ over Σ resulting

3.1. Constraints with symmetric edge-constrained strips

from concatenating the labels of the edges of γ in order in the vertical direction, namely $\Gamma_{i\alpha+k,j} = (\mathcal{L}_{\mathcal{I}}(e_i))_{k,j}$, for all $i \in [\ell]$, $k \in [\alpha]$ and $j \in [n]$. Finally, we define the weighted adjacency matrix of the labeled directed graph \mathcal{I} with weights given by \mathcal{W}_ϕ as the $|V_{\mathcal{I}}| \times |V_{\mathcal{I}}|$ nonnegative real matrix $A(\mathcal{I}, \mathcal{W}_\phi)$ with entries indexed by $V_{\mathcal{I}} \times V_{\mathcal{I}}$ and given by

$$(A(\mathcal{I}, \mathcal{W}_\phi))_{i,j} = \sum_{\substack{e \in E_{\mathcal{I}}: \\ \sigma(e)=i, \tau(e)=j}} \mathcal{W}_\phi(e) \quad ; \quad i, j \in V_{\mathcal{I}}.$$

Figure 3.1 shows paths generating an $\ell\alpha \times n$ array $\Gamma \in S$: the left part of the figure shows such a path in $\mathcal{G}_n^{(\mathcal{V})}$ and the right part of the figure shows the “corresponding” path in \mathcal{I} . The label of each edge in $\mathcal{G}_n^{(\mathcal{V})}$ and \mathcal{I} is “overlaid” on top of it. Each row of Γ —a path in $G_{\mathcal{E}}$ of length n —is depicted in the figure as a grey “snake-like” curve. In the figure, we denote by s_i, t_i the states $\sigma(\Gamma_{i,0}), \tau(\Gamma_{i,n-1}) \in G_{\mathcal{E}}$ respectively, for $i = 0, 1, \dots, \ell\alpha - 1$.

The following lemma generalizes ideas in [3, 7] and uses the weighted labeled graph \mathcal{I} to compute L_n , when $(\mathbf{x}_{m_k})_{k=1}^\infty$ is the sequence defined by (3.4).

Lemma 2. *For a sequence $(\mathbf{x}_{m_k})_{k=1}^\infty$ with each \mathbf{x}_{m_k} given by (3.4), and $\mathcal{I} = \mathcal{I}(\mu, \alpha, n, \mathcal{G}_n^{(\mathcal{V})}, G_{\mathcal{E}})$,*

$$\lim_{k \rightarrow \infty} \frac{\log \mathbf{x}_{m_k}^\top H_{m_k}^n \mathbf{x}_{m_k}}{m_k} = \frac{\log \lambda(A(\mathcal{I}, \mathcal{W}_\phi))}{\alpha}.$$

Proof. Let $(\mathbf{x}_{m_k})_{k=1}^\infty$ be a sequence of vectors with each \mathbf{x}_{m_k} given by (3.4). We shall show that there are positive real constants c, d (depending on \mathcal{I} and ϕ) such that for all positive integers k ,

$$\begin{aligned} c \cdot \mathbf{1}^\top (A(\mathcal{I}, \mathcal{W}_\phi))^{k + \lceil \mu/\alpha \rceil} \mathbf{1} &\leq \mathbf{x}_{m_k}^\top H_{m_k}^n \mathbf{x}_{m_k} \\ &\leq d \cdot \mathbf{1}^\top (A(\mathcal{I}, \mathcal{W}_\phi))^k \mathbf{1}. \end{aligned} \quad (3.6)$$

For a positive integer s and vector $\mathbf{e} = (e_0, \dots, e_{s-1})$ in $(E_{\mathcal{E}})^s$ denote by $\sigma(\mathbf{e}), \tau(\mathbf{e}) \in (V_{\mathcal{E}})^s$ the vectors with entries given by

$$\begin{aligned} (\sigma(\mathbf{e}))_i &= \sigma(e_i) \\ (\tau(\mathbf{e}))_j &= \tau(e_j) \quad ; \quad i \in \{0, 1, \dots, s-1\}. \end{aligned}$$

Now, fix a positive integer k . Let Γ be an array in $S_{m_k \times n}$. Recall that each entry of Γ is an edge in $G_{\mathcal{E}}$ and define the weight of Γ , denoted $\mathcal{W}_\phi(\Gamma)$, by

$$\mathcal{W}_\phi(\Gamma) = \prod_{i=0}^{k-1} \phi(\sigma(\Gamma_{i\alpha:i\alpha+\mu+\alpha-1,0})) \phi(\tau(\Gamma_{i\alpha:i\alpha+\mu+\alpha-1,n-1})).$$

3.1. Constraints with symmetric edge-constrained strips

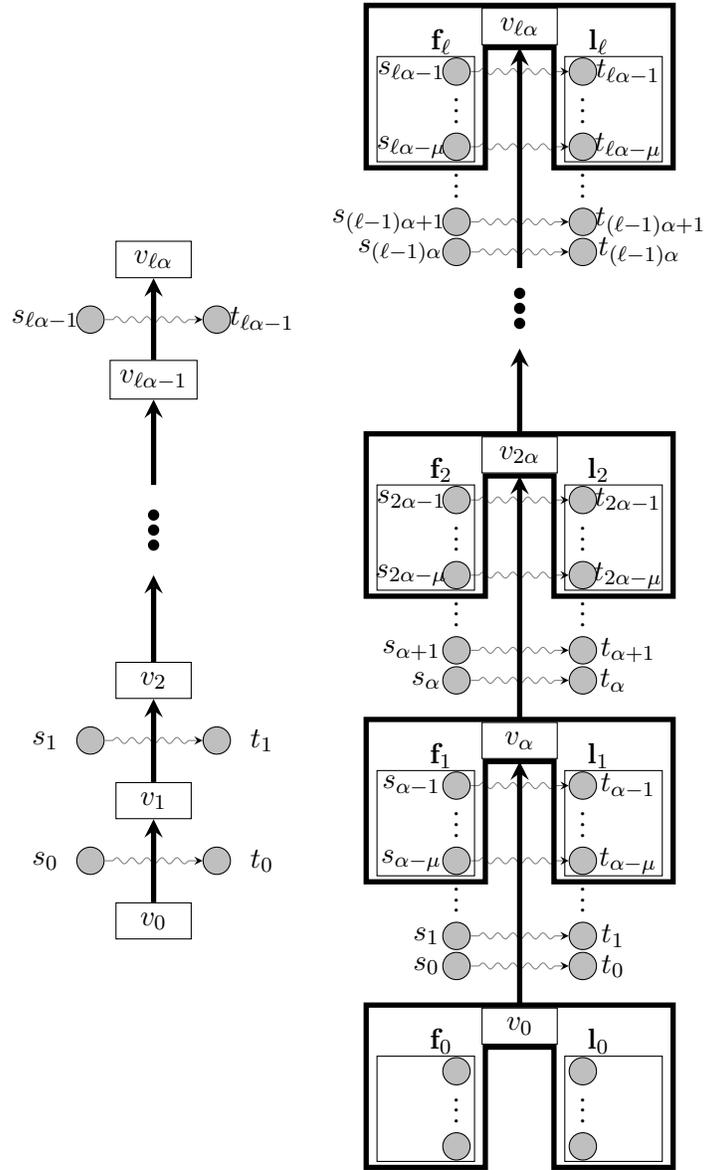


Figure 3.1: Paths generating an $l\alpha \times n$ -array of S , in $\mathcal{G}_n^{(v)}$ (left) and \mathcal{I} (right).

3.1. Constraints with symmetric edge-constrained strips

For a positive integer ℓ , let \mathcal{P}_ℓ denote the set of paths of length ℓ in \mathcal{I} . We denote the label of a path $\gamma \in \mathcal{P}_\ell$ by $\mathcal{L}_\mathcal{I}(\gamma)$. It is easily verified that there exists a path in \mathcal{P}_ℓ with label $\Gamma \in \Sigma^{\ell\alpha \times n}$ if and only if there exists a path in $\mathcal{G}_n^{(\mathcal{V})}$ of length $\ell\alpha$ that generates Γ . As $\mathcal{G}_n^{(\mathcal{V})}$ is a presentation of \mathcal{V}_n , the set of labels of paths in \mathcal{P}_ℓ is $S_{\ell\alpha \times n}$.

For a finite path γ in \mathcal{I} , define its weight, denoted $\mathcal{W}_\phi(\gamma)$, as the product of the weights of the edges in the path. Recalling that the entries of \mathbf{x}_{m_k} are indexed by $(V_\mathcal{E})^{m_k}$, we observe that

$$\begin{aligned} \mathbf{x}_{m_k}^\dagger H_{m_k}^n \mathbf{x}_{m_k} &= \sum_{\Gamma \in S_{m_k \times n}} (\mathbf{x}_{m_k})_{\sigma(\Gamma_{*,0})} (\mathbf{x}_{m_k})_{\tau(\Gamma_{*,n-1})} \\ &= \sum_{\Gamma \in S_{m_k \times n}} \mathcal{W}_\phi(\Gamma). \end{aligned}$$

For an array $\Gamma \in S_{m_k \times n}$, we say that a path $\gamma \in \mathcal{P}_k$ *matches* Γ if it is labeled by the sub-array $\Gamma_{\mu:m_k-1,*}$ and starts at a vertex $(\mathbf{f}, v, \mathbf{l}) \in V_\mathcal{I}$ with $\mathbf{f} = \sigma(\Gamma_{0:\mu-1,0})$ and $\mathbf{l} = \tau(\Gamma_{0:\mu-1,n-1})$. It can be verified from the construction of \mathcal{I} that if γ matches Γ then $\mathcal{W}_\phi(\gamma) = \mathcal{W}_\phi(\Gamma)$.

Now, since $\mathcal{G}_n^{(\mathcal{V})}$ is a presentation of $\mathcal{V}_n(S)$, it follows from the construction of \mathcal{I} that every $\Gamma \in S_{m_k \times n}$ has a path in \mathcal{P}_k matching it. Conversely, since for a path $\gamma \in \mathcal{P}_k$ all arrays $\Gamma \in S_{m_k \times n}$ that it matches have the same sub-array $\Gamma_{\mu:m_k-1,*}$, it follows that there are at most $|\Sigma|^{\mu n}$ arrays in $S_{m_k \times n}$ that γ matches. Therefore,

$$\begin{aligned} \mathbf{x}_{m_k}^\dagger H_{m_k}^n \mathbf{x}_{m_k} &= \sum_{\Gamma \in S_{m_k \times n}} \mathcal{W}_\phi(\Gamma) \\ &\leq |\Sigma|^{\mu n} \sum_{\gamma \in \mathcal{P}_k} \mathcal{W}_\phi(\gamma) \\ &= |\Sigma|^{\mu n} \mathbf{1}^\dagger (A(\mathcal{I}, \mathcal{W}_\phi))^k \mathbf{1}. \end{aligned}$$

This shows the right inequality of (3.6), we now turn to the left. Set $k' = k + \lceil \mu/\alpha \rceil$, $s = \lceil \mu/\alpha \rceil \alpha - \mu$, and let $\psi : \mathcal{P}_{k'} \rightarrow S_{m_k \times n}$ be given by

$$\psi(\gamma) = (\mathcal{L}_\mathcal{I}(\gamma))_{s:k'\alpha-1,*} ; \gamma \in \mathcal{P}_{k'}.$$

3.1. Constraints with symmetric edge-constrained strips

For a path $\gamma \in \mathcal{P}_{k'}$, with $\gamma = (e_i)_{i=0}^{k'-1} \subseteq E_{\mathcal{I}}$, its weight satisfies

$$\begin{aligned} \mathcal{W}_\phi(\gamma) &= \prod_{i=0}^{k'-1} \mathcal{W}_\phi(e_i) \\ &= \left(\prod_{i=0}^{\lceil \mu/\alpha \rceil - 1} \mathcal{W}_\phi(e_i) \right) \mathcal{W}_\phi(\psi(\gamma)) \\ &\leq \Phi^{2\lceil \mu/\alpha \rceil} \mathcal{W}_\phi(\psi(\gamma)), \end{aligned}$$

where we take Φ to be a positive constant satisfying $\Phi \geq \max \{\phi(\mathbf{v}) : \mathbf{v} \in (V_{\mathcal{E}})^{\mu+\alpha}\}$. Now let Γ be an array in $S_{m_k \times n}$. Since $\mathcal{G}_n^{(\mathcal{V})}$ is deterministic, so is \mathcal{I} , and thus for every vertex $v \in V_{\mathcal{I}}$, the paths in \mathcal{P}'_k starting at v are labeled distinctly. As all paths γ that map to Γ under ψ have the same sub-array $(\mathcal{L}_{\mathcal{I}}(\gamma))_{s:k'\alpha-1,*}$, it follows that there are at most $|\Sigma|^{sn}$ paths $\gamma \in \mathcal{P}'_k$ starting at v such that $\psi(\gamma) = \Gamma$. Consequently, there are at most $|V_{\mathcal{I}}| |\Sigma|^{sn}$ paths in $\mathcal{P}_{k'}$ that map to Γ under ψ . Therefore,

$$\begin{aligned} \mathbf{1}^\dagger (A(\mathcal{I}, \mathcal{W}_\phi))^{k+\lceil \mu/\alpha \rceil} \mathbf{1} &= \sum_{\gamma \in \mathcal{P}_{k'}} \mathcal{W}_\phi(\gamma) \\ &\leq \Phi^{2\lceil \mu/\alpha \rceil} \sum_{\gamma \in \mathcal{P}_{k'}} \mathcal{W}_\phi(\psi(\gamma)) \\ &\leq \Phi^{2\lceil \mu/\alpha \rceil} |V_{\mathcal{I}}| |\Sigma|^{sn} \sum_{\Gamma \in S_{m_k \times n}} \mathcal{W}_\phi(\Gamma) \\ &= \Phi^{2\lceil \mu/\alpha \rceil} |V_{\mathcal{I}}| |\Sigma|^{sn} \\ &\quad \cdot \mathbf{x}_{m_k}^\dagger H_{m_k}^n \mathbf{x}_{m_k}. \end{aligned}$$

Dividing both sides by $\Phi^{2\lceil \mu/\alpha \rceil} |V_{\mathcal{I}}| |\Sigma|^{sn}$ we obtain the left inequality of (3.6). The claim of the lemma now follows from Perron-Frobenius theory by taking the log of (3.6), dividing it by m_k and taking the limit as k approaches infinity. \blacksquare

We thus obtain the following lower bound on the capacity of a 2-dimensional constraint.

Theorem 1. *Let T, \mathcal{E} be 1-dimensional constraints over an alphabet Σ , with \mathcal{E} an edge constraint defined by a graph $G_{\mathcal{E}} = (V_{\mathcal{E}}, E_{\mathcal{E}})$. Set $S = T \otimes \mathcal{E}$ and suppose that S has symmetric horizontal edge-constrained strips. Let $\mu \geq 0$ and $\alpha, p, q > 0$ be integers and $\phi : (V_{\mathcal{E}})^{\mu+\alpha} \rightarrow [0, \infty)$ be a nonnegative real function. For a positive integer n , let \mathcal{G}_n be a deterministic presentation of $\mathcal{V}_n(S)$, and set $A_{n,\phi} =$*

3.1. Constraints with symmetric edge-constrained strips

$A(\mathcal{I}(\mu, \alpha, n, \mathcal{G}_n, G_{\mathcal{E}}), \mathcal{W}_{\phi})$. Then

$$\text{cap}(S) \geq \frac{\log \lambda(A_{2q+p, \phi}) - \log \lambda(A_{2q, \phi})}{p\alpha}. \quad (3.7)$$

Remark 1. In addition to computing lower-bounds, [3] gives a method for computing upper bounds on the capacity of the hard-square constraint. It can be shown that this method can also be applied to all constraints of the form $T \otimes \mathcal{E}$, with \mathcal{E} an edge constraint, having symmetric horizontal edge-constrained strips.

Remark 2. Theorem 1 can be generalized to apply to 2-dimensional constraints having symmetric horizontal edge-constrained strips, which are not necessarily axial-products. Let S be such a constraint, and for every positive integer m , let $G_m^{(\mathcal{H})} = (V_m^{(\mathcal{H})}, E_m^{(\mathcal{H})})$ be the symmetric graph, with no isolated vertices, defining $\mathcal{H}_m(S)$. Set $G_{\mathcal{E}} = G_1^{(\mathcal{H})}$, $V_{\mathcal{E}} = V_1^{(\mathcal{H})}$ and $E_{\mathcal{E}} = E_1^{(\mathcal{H})}$. We claim there exists a mapping $f_m : V_m^{(\mathcal{H})} \rightarrow (V_{\mathcal{E}})^m$ such that for every edge $\mathbf{e} = e_0 e_1 \dots e_{m-1} \in E_m^{(\mathcal{H})}$, with each $e_i \in E_{\mathcal{E}}$,

$$\begin{aligned} f_m(\sigma(\mathbf{e})) &= (\sigma(e_0), \dots, \sigma(e_{m-1})) \text{ and} \\ f_m(\tau(\mathbf{e})) &= (\tau(e_0), \dots, \tau(e_{m-1})). \end{aligned} \quad (3.8)$$

This mapping is defined as follows. For a vertex $v \in V_m^{(\mathcal{H})}$, pick an incoming edge $\mathbf{e} = e_0 e_1 \dots e_{m-1} \in E_m^{(\mathcal{H})}$ and define $f_m(v)$ as $(\tau(e_0), \dots, \tau(e_{m-1}))$. This mapping is uniquely-defined: indeed if $\mathbf{e}' = e'_0 \dots e'_{m-1} \in E_m^{(\mathcal{H})}$ is another incoming edge of v , and $\mathbf{g} = g_0 \dots g_{m-1} \in E_m^{(\mathcal{H})}$ is an outgoing edge of v , then clearly, for every $i \in [m]$, both $e_i g_i$ and $e'_i g_i$ are paths in $G_{\mathcal{E}}$; consequently $\tau(e_i) = \tau(e'_i)$. It is easy to check that f_m satisfies the conditions in (3.8). Now, replace the definition of $(\mathbf{x}_{m_k})_{\mathbf{v}}$ in (3.4) with

$$(\mathbf{x}_{m_k})_{\mathbf{v}} = \prod_{i=0}^{k-1} \phi(\mathbf{u}_{i\alpha:i\alpha+\mu+\alpha-1}) \quad ; \quad \mathbf{u} = f_{m_k}(\mathbf{v}), \mathbf{v} \in V_{m_k}^{(\mathcal{H})}.$$

With this new definition and the aid of (3.8), it can be verified that Lemma 2 and consequently Theorem 1 still hold.

Remark 3. Clearly, it is sufficient, for the theorem to hold, that $\mathcal{H}_m(S)$ is symmetric for large enough m .

We now give a sufficient condition for the constraint $S = T \otimes \mathcal{E}$ to have symmetric horizontal edge-constrained strips. For this to happen, we (generally) must have that $G_{\mathcal{E}}$ is symmetric. This means that there exists a ‘‘matching’’ between edges, where each edge is matched with an edge in the ‘‘reverse’’ direction. More precisely there is a bijection $R : E_{\mathcal{E}} \rightarrow E_{\mathcal{E}}$ such that for all $e \in E_{\mathcal{E}}$,

$(\sigma(e), \tau(e)) = (\tau(R(e)), \sigma(R(e)))$ and $R(R(e)) = e$. We call such a bijection an *edge-reversing matching*, and we denote by $\mathcal{R}(G_{\mathcal{E}})$ the set of all edge-reversing matchings of $G_{\mathcal{E}}$. Clearly a graph G is symmetric iff it has an edge-reversing matching. Thus $T \otimes \mathcal{E}$ has symmetric horizontal edge-constrained strips iff for every m , $G_m^{(\mathcal{H})}$ has an edge-reversing matching. We present a sufficient condition for this to hold.

Proposition 3. *Let T, \mathcal{E} be 1-dimensional constraints over an alphabet Σ , with \mathcal{E} an edge constraint defined by a graph $G_{\mathcal{E}} = (V_{\mathcal{E}}, E_{\mathcal{E}})$ with $R \in \mathcal{R}(G_{\mathcal{E}})$ an edge-reversing matching. If for every word $e_1 \dots e_m \in T$ one has $R(e_1) \dots R(e_m) \in T$ as well (for all m), then $T \otimes \mathcal{E}$ has symmetric horizontal edge-constrained strips.*

Proof. Let $G_m^{(\mathcal{H})} = (V_{\mathcal{E}}^m, E_m^{(\mathcal{H})})$ be the subgraph of $G_{\mathcal{E}}^{\times m}$ that defines $\mathcal{H}_m(S)$. We show that $G_m^{(\mathcal{H})}$ is symmetric. Let $R^{\times m} : E_{\mathcal{E}}^m \rightarrow E_{\mathcal{E}}^m$ be defined by

$$R^{\times m}(e_1, \dots, e_m) = (R(e_1), \dots, R(e_m)).$$

Clearly, $R^{\times m}$ is an edge-reversing matching of $G_{\mathcal{E}}^{\times m}$. Recall that $E_m^{(\mathcal{H})}$ consists of all the edges in $E_{\mathcal{E}}^m$ that, when regarded as m -letter words over Σ , satisfy T . Therefore, by the assumption, it follows that for all $\mathbf{e} \in E_m^{(\mathcal{H})}$, $R^{\times m}(\mathbf{e}) \in E_m^{(\mathcal{H})}$ as well. Consequently, $R^{\times m}$ restricted to $E_m^{(\mathcal{H})}$, is an edge-reversing matching of $G_m^{(\mathcal{H})}$ and hence it is symmetric. ■

If \mathcal{G}_T is a presentation of T and $R \in \mathcal{R}(G_{\mathcal{E}})$, a sufficient condition for the hypothesis of Proposition 3 to hold, which may be easier to check, is the existence of a function $f : E_T \rightarrow E_T$ satisfying: 1) $\mathcal{L}_T(f(e)) = R(\mathcal{L}_T(e))$ for all $e \in E_T$ and 2) for any path $e_1 e_2$ of length 2 in \mathcal{G}_T , the sequence $f(e_1) f(e_2)$ is also a path in \mathcal{G}_T . Indeed, if such a function exists then any path $e_1 e_2 \dots e_m$ in \mathcal{G}_T generating a word $e_1 e_2 \dots e_m$ has a corresponding path $f(e_1) f(e_2) \dots f(e_m)$ generating the word $R(e_1) R(e_2) \dots R(e_m)$, and thus the hypothesis of Proposition 3 is fulfilled. In fact, it can be shown that when \mathcal{G}_T is irreducible, deterministic and has the minimum number of vertices among all deterministic presentations of T , this condition is also necessary for the hypothesis of Proposition 3 to hold (see [26, Section 3.3]).

In Section 3.3 we use Proposition 3 to show that the method described in this section can be used to compute lower bounds on $\text{CHG}(b_1) \otimes \text{CHG}(b_2)$ for any positive integers b_1 and b_2 .

3.2 Constraints with symmetric vertex-constrained strips

In this section we present an analog to Theorem 1 that gives lower bounds on the capacities of constraints for which every horizontal or every vertical strip is a sym-

3.2. Constraints with symmetric vertex-constrained strips

metric vertex-constraint. We do this by transforming a 2-dimensional constraint with symmetric vertex-constrained strips to a 2-dimensional constraint with symmetric edge-constrained strips, having the same capacity.

Fix an alphabet Σ , and let S be a 2-dimensional constraint over Σ . We say that S has *horizontal vertex-constrained strips* if for every positive integer m , the constraint $\mathcal{H}_m(S)$ is a vertex-constraint. If, in addition, every horizontal strip is symmetric, we say that S has *symmetric horizontal vertex-constrained strips*. The notions of a 2-dimensional constraint with vertical vertex-constrained strips and symmetric vertical vertex-constrained strips are defined analogously.

It turns out that RWIM and NAK do not have horizontal or vertical edge-constrained strips, and so the method in section 3.1 does not apply directly. We illustrate this only for horizontal strips for $S = \text{RWIM}$. Recall from Chapter 2 that an edge constraint is a constraint of memory 1 such that any two follower sets are either disjoint or identical. We claim that this condition does not hold for $\mathcal{H}_m(S)$. To see this, given any m , let $w = w_0 \dots w_{m-1}$ be the all-zeros word of length m and $u = u_0 \dots u_{m-1}$ be any other word of length m . Now, the $m \times 2$ arrays

$$\begin{array}{cccccc} w_0 & w_0 & w_0 & u_0 & u_0 & w_0 \\ w_1 & w_1 & w_1 & u_1 & u_1 & w_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{m-1} & w_{m-1} & w_{m-1} & u_{m-1} & u_{m-1} & w_{m-1} \end{array},$$

belong to S , yet the $m \times 2$ array

$$\begin{array}{cc} u_0 & u_0 \\ u_1 & u_1 \\ \vdots & \vdots \\ u_{m-1} & u_{m-1} \end{array},$$

does not. Thus, w and u , regarded as $m \times 1$ columns, have different but non-disjoint follower sets. Consequently, RWIM does not have horizontal edge-constrained strips.

However, it is not hard to show that RWIM and NAK have both symmetric horizontal vertex-constrained strips and symmetric vertical vertex-constrained strips. For instance, for $S = \text{RWIM}$, $\mathcal{H}_m(S)$ is the vertex constraint defined by the graph $G = (V, E)$, where V consists of all binary vectors $u_0 \dots u_{m-1}$ of length m and E consists of a single edge from $u \in V$ to $v \in V$ iff for all i , whenever $u_i = 1$, then $v_{i+1} = v_i = v_{i-1} = 0$ (with the obvious modification when $i = 0$ or $m - 1$). And $\mathcal{V}_m(S)$ is the vertex constraint defined by the graph of $G' = (V', E')$, where V' consists of all binary vectors $u_0 \dots u_{m-1}$ of length m which do not contain two

3.2. Constraints with symmetric vertex-constrained strips

adjacent '1's and E' consists of a single edge from $u \in V$ to $v \in V$ iff for all i , whenever $u_i = 1$, then $v_{i+1} = v_{i-1} = 0$ (again, with the obvious modification when $i = 0$ or $m - 1$). Clearly, both G and G' are symmetric.

Now, let S be a 2-dimensional constraint over Σ . For a finite $m \times n$ array Γ with $m \geq 1$ and $n \geq 2$ over Σ its $[1 \times 2]$ -higher block recoding or $[1 \times 2]$ -recoding is an $m \times (n - 1)$ array $\tilde{\Gamma}$ over $\Sigma^{1 \times 2}$ with entries given by

$$\tilde{\Gamma}_{i,j} = (\Gamma_{i,j} \Gamma_{i,j+1}) ; i = 0, \dots, m - 1, j = 0, \dots, n - 2.$$

We denote by $S^{[1 \times 2]}$ the set of all $[1 \times 2]$ -recodings of arrays in S and refer to it as the $[1 \times 2]$ -higher block recoding of S . The $[1 \times 2]$ -higher block recoding of a constraint is a constraint. This is stated in the following proposition.

Proposition 4. *Let S be a 2-dimensional constraint over Σ . Then $S^{[1 \times 2]}$ is a 2-dimensional constraint over $\Sigma^{1 \times 2}$.*

Remark . We may, of course, define, in a similar manner, the $[s \times t]$ -higher block recoding of S , for any positive integers s and t , and the $[s \times t]$ -higher block recoding of a 2-dimensional constraint S is a 2-dimensional constraint.

Proof. Set $S' = S^{[1 \times 2]}$ and let $(\mathcal{G}_V, \mathcal{G}_H)$ be a presentation of S with $\mathcal{G}_V = (V_V, E, \mathcal{L})$ and $\mathcal{G}_H = (V_H, E, \mathcal{L})$. We construct labeled graphs $\mathcal{G}'_V = (V_V \times V_V, E', \mathcal{L}')$ and $\mathcal{G}'_H = (E, E', \mathcal{L}')$ as follows. The set of edges E' is defined as

$$E' = \{(e_0 e_1) \in E^{1 \times 2} : e_0, e_1 \text{ is a path in } \mathcal{G}_H\},$$

and the labeling function $\mathcal{L}' : E' \rightarrow \Sigma^{1 \times 2}$ is given by

$$\mathcal{L}'(e_0 e_1) = (\mathcal{L}(e_0) \mathcal{L}(e_1)) ; (e_0 e_1) \in E'.$$

For every edge $(e_0 e_1) \in E'$ we define

$$\begin{aligned} \sigma_{\mathcal{G}'_V}(e_0 e_1) &= (\sigma_{\mathcal{G}_V}(e_0), \sigma_{\mathcal{G}_V}(e_1)) \\ \tau_{\mathcal{G}'_V}(e_0 e_1) &= (\tau_{\mathcal{G}_V}(e_0), \tau_{\mathcal{G}_V}(e_1)) \\ \sigma_{\mathcal{G}'_H}(e_0 e_1) &= e_0 \\ \tau_{\mathcal{G}'_H}(e_0 e_1) &= e_1 \end{aligned} .$$

It's easy to verify that $(\mathcal{G}'_V, \mathcal{G}'_H)$ is a presentation of S' . ■

Clearly, recoding is an injective mapping, thus $|S_{m \times n}| = |S^{[1 \times 2]}_{m \times (n-1)}|$ for all positive integers $m \geq 1, n \geq 2$. It follows that $\text{cap}(S) = \text{cap}(S^{[1 \times 2]})$. The next proposition shows that the $[1 \times 2]$ -higher block recoding of a constraint with symmetric horizontal vertex-constrained strips has symmetric horizontal edge-constrained strips.

3.2. Constraints with symmetric vertex-constrained strips

Proposition 5. *Let S be a 2-dimensional constraint with horizontal vertex-constrained strips.*

1. $S^{[1 \times 2]}$ has horizontal edge-constrained strips. Moreover, $S^{[1 \times 2]}$ has symmetric horizontal edge-constrained strips iff S has symmetric horizontal vertex-constrained strips.
2. $S^{[1 \times 2]} = \mathcal{V}_1(S^{[1 \times 2]}) \otimes \mathcal{H}_1(S^{[1 \times 2]})$.

Proof. (1). Let m be a positive integer, $G_m^{(\mathcal{H})} = (V_m^{(\mathcal{H})}, E_m)$ be the graph defining the vertex-constraint $\mathcal{H}_m(S)$ and set $\Delta = \Sigma^{1 \times 2}$, where Σ is the alphabet of S . We define a labeling $\mathcal{L} : E_m \rightarrow \Delta^{m \times 1}$ of the edges of $G_m^{(\mathcal{H})}$. For $\mathbf{e} \in E_m$, we regard $\sigma(\mathbf{e})$ and $\tau(\mathbf{e})$ as $m \times 1$ arrays over Σ , and define $\mathcal{L}(\mathbf{e})$ to be the array in $\Delta^{m \times 1}$ with entries given by

$$\mathcal{L}(\mathbf{e})_{(i,0)} = (\sigma(\mathbf{e})_{i,0} \tau(\mathbf{e})_{i,0}) ; \quad i \in [m].$$

It's easily verified that the word generated by every path in the labeled graph $(G_m^{(\mathcal{H})}, \mathcal{L})$ is the $[1 \times 2]$ -higher block recoding of the array formed by concatenating the vertices along the path horizontally in sequence. It follows that $(G_m^{(\mathcal{H})}, \mathcal{L})$ is a presentation of $\mathcal{H}_m(S^{[1 \times 2]})$. Since the labels of the edges in $(G_m^{(\mathcal{H})}, \mathcal{L})$ are distinct, we may identify each edge with its label, and it follows that $\mathcal{H}_m(S^{[1 \times 2]})$ is an edge-constraint. Since the same graph defines both $\mathcal{H}_m(S)$ and $\mathcal{H}_m(S^{[1 \times 2]})$ (the former as a vertex-constraint and the latter as an edge-constraint), it follows that $\mathcal{H}_m(S)$ is symmetric iff $\mathcal{H}_m(S^{[1 \times 2]})$ is. This completes the proof.

(2). Clearly, $S^{[1 \times 2]} \subseteq \mathcal{V}_1(S^{[1 \times 2]}) \otimes \mathcal{H}_1(S^{[1 \times 2]})$. As for the reverse inclusion, let $\tilde{\Gamma} \in \mathcal{V}_1(S^{[1 \times 2]}) \otimes \mathcal{H}_1(S^{[1 \times 2]})$ be an $m \times n$ array over $\Sigma^{1 \times 2}$. Since every row of $\tilde{\Gamma}$ is in $\mathcal{H}_1(S^{[1 \times 2]})$, every row has a unique $1 \times (n + 1)$ pre-image under the recoding map. Let Γ be the $m \times (n + 1)$ array over Σ , whose i th row is the pre-image under the recoding map of the i th row of $\tilde{\Gamma}$, for $i = 0, \dots, m - 1$. Clearly, $\tilde{\Gamma}$ is the $[1 \times 2]$ -higher block recoding of Γ . Thus, it suffices to show that $\Gamma \in S$. For $i = 0, 1, \dots, n - 1$ clearly, the $m \times 2$ array $\Gamma_{*,i:i+1}$ over Σ recodes to the column $\tilde{\Gamma}_{*,i}$. By our assumption this column is in $\mathcal{V}_1(S^{[1 \times 2]})$. Since recoding is injective, $\Gamma_{*,i:i+1}$ must be in S . Since this holds for all $i = 0, 1, \dots, n - 1$ and since $\mathcal{H}_m(S)$ has memory 1, it follows that $\Gamma \in S$ and therefore $\tilde{\Gamma} \in S^{[1 \times 2]}$. ■

We can now use the method described in Section 3.1 to get lower bounds on 2-dimensional constraints with symmetric horizontal vertex-constrained strips. This is stated in the following theorem.

3.3. Capacity bounds for axial products of constraints

Theorem 2. *Let S be a 2-dimensional constraint over an alphabet Σ with symmetric horizontal vertex-constrained strips. Let $\mu \geq 0$, and $\alpha, p, q > 0$ be integers, $G_{\mathcal{E}} = (V_{\mathcal{E}}, E_{\mathcal{E}})$ be the graph defining the vertex-constraint $\mathcal{H}_1(S)$ (hence $V_{\mathcal{E}} \subseteq \Sigma$), and $\phi : (V_{\mathcal{E}})^{\mu+\alpha} \rightarrow [0, \infty)$ be a nonnegative function. For an integer $n \geq 2$, let \mathcal{G}_n be a labeled graph obtained from a deterministic presentation for $\mathcal{V}_n(S)$ by replacing each edge-label with its $[1 \times 2]$ -higher block recoding. Set $\tilde{A}_{n,\phi} = A(\mathcal{I}(\mu, \alpha, n-1, \mathcal{G}_n, G_{\mathcal{E}}), \mathcal{W}_{\phi})$, where \mathcal{I} , \mathcal{W}_{ϕ} , and $A(\mathcal{I}, \mathcal{W}_{\phi})$ are as defined in Section 3.1. Then*

$$\text{cap}(S) \geq \frac{\log \lambda(\tilde{A}_{p+2q+1,\phi}) - \log \lambda(\tilde{A}_{2q+1,\phi})}{p\alpha}.$$

Proof. Let $S' = S^{[1 \times 2]}$. By Proposition 5, $S' = \mathcal{V}_1(S') \otimes \mathcal{H}_1(S')$, and S' has horizontal symmetric edge-constrained strips. Since $G_{\mathcal{E}}$ has no parallel edges, we may identify each edge $e \in E_{\mathcal{E}}$ with the pair $(\sigma(e), \tau(e))$; then, with this identification, $\mathcal{H}_1(S') = \mathbb{X}(G_{\mathcal{E}})$. Also, note that \mathcal{G}_{2q+p+1} and \mathcal{G}_{2q+1} are deterministic presentations for $\mathcal{V}_{2q+p}(S')$ and $\mathcal{V}_{2q}(S')$, respectively. The result follows from Theorem 1 applied to S' . \blacksquare

3.3 Capacity bounds for axial products of constraints

In this section we show how the method described in Section 3.1 can be applied to axial products of certain 1-dimensional constraints. Let S and T be two 1-dimensional constraints over an alphabet Σ . We wish to lower bound the capacity of the 2-dimensional constraint $T \otimes S$. To this end, we pick a lossless presentation $\mathcal{G}_S = (G_S, \mathcal{L}_S)$, with $G_S = (V_S, E_S)$, for S . We extend the function \mathcal{L}_S to multidimensional arrays over E_S in the manner described in Chapter 2, and for a set $A \subseteq \Sigma^*$, we denote by $\mathcal{L}_S^{-1}(A) \subseteq E_S^*$ the inverse image of A under this map, namely

$$\mathcal{L}_S^{-1}(A) = \{w \in E_S^* : \mathcal{L}_S(w) \in A\}.$$

The following proposition shows that we can reduce the problem of calculating the capacity of $T \otimes S$ to that of calculating the capacity of $\mathcal{L}_S^{-1}(T) \otimes \mathbb{X}(G_S)$.

Proposition 6. *Let S, T be two 1-dimensional constraints and let $\mathbb{X}(G_S)$ and $\mathcal{L}_S^{-1}(T)$ be as defined above. Then*

1. $\mathcal{L}_S^{-1}(T)$ is a 1-dimensional constraint.
2. $\text{cap}(T \otimes S) = \text{cap}(\mathcal{L}_S^{-1}(T) \otimes \mathbb{X}(G_S))$.

3.3. Capacity bounds for axial products of constraints

Proof. 1. Let $\mathcal{G}_T = (V_T, E_T, \mathcal{L}_T)$ be a presentation of T . We shall construct a presentation $\mathcal{F} = (V_T, E_{\mathcal{F}}, \mathcal{L}_{\mathcal{F}})$ of $\mathcal{L}_S^{-1}(T)$. The set of edges is given by $E_{\mathcal{F}} = \{(e_T, e_S) \in E_T \times E_S : \mathcal{L}_T(e_T) = \mathcal{L}_S(e_S)\}$, and for an edge $(e_T, e_S) \in E_{\mathcal{F}}$, $\sigma_{\mathcal{F}}(e_T, e_S) = \sigma_{\mathcal{G}_T}(e_T)$, $\tau_{\mathcal{F}}(e_T, e_S) = \tau_{\mathcal{G}_T}(e_T)$ and $\mathcal{L}_{\mathcal{F}}(e_T, e_S) = e_S$. It is easily verified that $\mathcal{L}_S^{-1}(T)$ is presented by \mathcal{F} , and therefore it is a 1-dimensional constraint.

2. We set $R = T \otimes S$, $U = \mathcal{L}_S^{-1}(T) \otimes X(G_S)$. For an array $\Delta \in R_{m \times n}$, define $P_{\Delta} = \{\Gamma \in U_{m \times n} : \mathcal{L}_S(\Gamma) = \Delta\}$, we claim that

$$1 \leq |P_{\Delta}| \leq |V_S|^{2m}. \quad (3.9)$$

Indeed, it's easily verified that an array $\Gamma \in E_S^{m \times n}$ is in P_{Δ} iff for all $i \in [m]$ the row $(\Gamma_{i,j})_{j=0}^{n-1}$ is a path in \mathcal{G}_S that generates $(\Delta_{i,j})_{j=0}^{n-1}$. Since \mathcal{G}_S is a lossless presentation of S , for every $i \in [m]$, there is at least one path in \mathcal{G}_S generating $(\Delta_{i,j})_{j=0}^{n-1}$ and at most $|V_S|^2$ such paths; the claim follows. Now, clearly for any $\Gamma \in U_{m \times n}$ the array $\mathcal{L}_S(\Gamma)$ is in $R_{m \times n}$. It follows that the sets P_{Δ} , for $\Delta \in R_{m \times n}$ form a partition of $U_{m \times n}$, and we have

$$|U_{m \times n}| = \sum_{\Delta \in R_{m \times n}} |P_{\Delta}|.$$

Therefore, by (3.9), we get

$$|R_{m \times n}| \leq |U_{m \times n}| \leq |R_{m \times n}| |V_S|^{2m},$$

and it follows from (1.1) that $\text{cap}(R) = \text{cap}(U)$. ■

Therefore if $\mathcal{L}_S^{-1}(T) \otimes X(G_S)$ has symmetric horizontal edge-constrained strips, we can apply the method of Section 3.1 to obtain lower bounds on $\text{cap}(T \otimes S)$. In this case, it also follows from Remark 1 of Theorem 1, that the method of [3] for obtaining upper bounds on the capacity of the hard-square constraint, can be used to obtain upper bounds on $\text{cap}(T \otimes S)$. Proposition 3 present a sufficient condition for $\mathcal{L}_S^{-1}(T) \otimes X(G_S)$ to have symmetric horizontal edge-constrained strips. Here we give another stronger sufficient condition involving only the presentation \mathcal{G}_S . We say that a labeled graph (G, \mathcal{L}) , with $G = (V, E)$, is *symmetric as a labeled graph*, if there exists an edge-reversing matching $R \in \mathcal{R}(G)$ which preserves \mathcal{L} , that is $\mathcal{L}(R(e)) = \mathcal{L}(e)$ for all $e \in E$. We assume now that \mathcal{G}_S is symmetric as a labeled graph, and that $R \in \mathcal{R}(G_S)$ is an edge-reversing matching which preserves \mathcal{L}_S . Since for any positive integer m and $e_1 \dots e_m \in E_S^m$, the label $\mathcal{L}(e_1) \dots \mathcal{L}(e_m) = \mathcal{L}(R(e_1)) \dots \mathcal{L}(R(e_m))$, it follows that $e_1 \dots e_m \in \mathcal{L}_S^{-1}(T)$ iff $R(e_1) \dots R(e_m) \in \mathcal{L}_S^{-1}(T)$. Consequently, the hypothesis of Proposition 3 holds and we have the following corollary.

3.4. Heuristics for choosing ϕ

Corollary 1. *If \mathcal{G}_S is symmetric as a labeled graph then $\mathcal{L}_S^{-1}(T) \otimes \mathsf{X}(G_S)$ has symmetric horizontal edge-constrained strips.*

Since the presentation in Figure 1.1a is symmetric as a labeled graph, we can apply the method of Section 3.1 to get lower bounds on the capacity of all constraints $T \otimes \text{EVEN}$ for any 1-dimensional constraint T .

Let $S = \text{CHG}(b_1)$ and let $T = \text{CHG}(b_2)$ for integers $b_1, b_2 \geq 2$. Let $\mathcal{G}_S = (G_S, \mathcal{L}_S)$, with $G_S = (V_S, E_S)$, be the presentation given in Figure 1.1c for $b = b_1$. Evidently, \mathcal{G}_S is symmetric with exactly one edge-reversing matching, $R : E_S \rightarrow E_S$. Fix a positive integer m and let $\mathbf{e} = e_1 e_2 \dots e_m \in E_S^m$. Obviously, T is closed under negation of words (i.e., negating each symbol), and we have

$$\begin{aligned}
 & e_1 e_2 \dots e_m \in \mathcal{L}_S^{-1}(T) \\
 \iff & \mathcal{L}_S(e_1) \mathcal{L}_S(e_2) \dots \mathcal{L}_S(e_m) \in T \\
 \iff & (-\mathcal{L}_S(e_1)) (-\mathcal{L}_S(e_2)) \dots (-\mathcal{L}_S(e_m)) \in T \\
 \iff & \mathcal{L}_S(R(e_1)) \mathcal{L}_S(R(e_2)) \dots \mathcal{L}_S(R(e_m)) \in T \\
 \iff & R(e_1) R(e_2) \dots R(e_m) \in \mathcal{L}_S^{-1}(T).
 \end{aligned}$$

Consequently, it follows by Proposition 3 that $\mathcal{L}_S^{-1}(T) \otimes \mathsf{X}(G_S)$ has symmetric horizontal edge-constrained strips and we can apply the method of Section 3.1 to obtain lower bounds on the capacity of $\text{CHG}(b_2) \otimes \text{CHG}(b_1)$.

The reader will note a similarity in the constructions in proofs of Propositions 1 and 6. Indeed, as an alternative approach, one may be able to use the construction in Proposition 1 to obtain bounds on $\text{cap}(S \otimes T)$: namely, if G_1 and G_2 are the underlying graphs of a capacity-preserving presentation $(\mathcal{G}_1, \mathcal{G}_2)$ of $S \otimes T$ and $\mathsf{X}(G_1) \otimes \mathsf{X}(G_2)$ has symmetric horizontal edge-constrained strips. However, the approach given by Proposition 6 seems to be more direct and simpler than the alternative approach.

3.4 Heuristics for choosing ϕ

In this section, we use the notation defined in Section 3.1, and assume that $S = T \otimes \mathcal{E}$ is a 2-dimensional constraint with symmetric horizontal edge-constrained strips, where \mathcal{E} is an edge constraint. We describe heuristics for choosing the function ϕ to obtain “good” lower bounds on the capacity of S .

3.4.1 Using max-entropic probabilities

Recall that a vertex of a directed graph is isolated if no edges in the graph are connected to it. Note, that since $G_m^{(\mathcal{H})}$ is symmetric, every vertex is either isolated

3.4. Heuristics for choosing ϕ

or has both incoming and outgoing edges. We assume here that for every positive integer m , ignoring isolated vertices, $G_m^{(\mathcal{H})}$ is a primitive graph. In this case, the Perron eigenvector of H_m is unique up to multiplication by a scalar. Let \mathbf{r}_m be the right Perron eigenvector of H_m normalized to be a unit vector in the L_2 -norm. Observe, that substituting \mathbf{r}_m for \mathbf{x}_m satisfies (3.2) with an equality. This motivates us to choose ϕ so that the resulting vector \mathbf{x}_m approximates \mathbf{r}_m . Since $G_m^{(\mathcal{H})}$ (without its isolated vertices) is irreducible, there is a unique stationary probability measure having maximum entropy on arrays of \mathcal{H}_m , namely the max-entropic probability measure on \mathcal{H}_m . We denote it here by $\text{Pr}^{*,m}$. It is given by

$$\text{Pr}^{*,m}(\Gamma) = \frac{(\mathbf{r}_m)_{\sigma(\Gamma)} (\mathbf{r}_m)_{\tau(\Gamma)}}{\lambda(H_m)^\ell}.$$

For $\Gamma \in S_{m \times \ell}$, for some positive integer ℓ , and where $\sigma(\Gamma), \tau(\Gamma) \in V_{\mathcal{E}}^m$ are given by

$$\begin{aligned} (\sigma(\Gamma))_i &= \sigma(\Gamma_{i,0}) \\ (\tau(\Gamma))_i &= \tau(\Gamma_{i,\ell-1}) \end{aligned} \quad ; \quad i = 0, 1, 2, \dots, m-1.$$

Let $\mathbf{V}^{(m)}$ be a random variable taking values in $V_{\mathcal{E}}^m$ with distribution given by

$$\text{Pr}(\mathbf{V}^{(m)} = \mathbf{v}) = \text{Pr}^{*,m} \{ \Gamma \in S_{m \times 1} : \sigma(\Gamma) = \mathbf{v} \} \quad ; \quad \mathbf{v} \in V_{\mathcal{E}}^m.$$

It's easily verified that

$$\text{Pr}(\mathbf{V}^{(m)} = \mathbf{v}) = ((\mathbf{r}_m)_{\mathbf{v}})^2. \quad (3.10)$$

Thus approximating $\text{Pr}(\mathbf{V}^{(m)} = \mathbf{v})$ and taking a square root will give us an approximation for $(\mathbf{r}_m)_{\mathbf{v}}$. Roughly speaking, $\text{Pr}(\mathbf{V}^{(m)} = \mathbf{v})$ is the probability of seeing the column of vertices \mathbf{v} in the “middle” of an $m \times \ell$ array chosen uniformly at random from $S_{m \times \ell}$, for large ℓ . Fix integers $\mu \geq 0, \alpha \geq 1$ as in Section 3.1, and assume now that $m = m_k = \mu + k\alpha$, for a positive integer k . For an integer $0 \leq s < m$ and vectors $\mathbf{u} \in V_{\mathcal{E}}^\ell, \mathbf{w} \in V_{\mathcal{E}}^r$, with lengths satisfying $\ell \leq m - s, r \leq s$, denote by $p_s^{(m)}(\mathbf{u})$ and $p_s^{(m)}(\mathbf{u}|\mathbf{w})$ the probabilities given by

$$\begin{aligned} p_s^{(m)}(\mathbf{u}) &= \text{Pr}(\mathbf{V}_{s:s+\ell-1}^{(m)} = \mathbf{u}) \\ p_s^{(m)}(\mathbf{u}|\mathbf{w}) &= \text{Pr}(\mathbf{V}_{s:s+\ell-1}^{(m)} = \mathbf{u} | \mathbf{V}_{s-r:s-1}^{(m)} = \mathbf{w}). \end{aligned}$$

Then by the chain rule for conditional probability we have, for any vector $\mathbf{v} \in V_{\mathcal{E}}^m$,

$$\begin{aligned} \text{Pr}(\mathbf{V}^{(m)} = \mathbf{v}) &= p_0^{(m)}(\mathbf{v}_{0:\mu-1}) \\ &\quad \cdot \prod_{i=0}^{k-1} p_{\mu+i\alpha}^{(m)}(\mathbf{v}_{\mu+i\alpha:\mu+(i+1)\alpha-1} | \mathbf{v}_{0:\mu+i\alpha-1}). \end{aligned}$$

3.4. Heuristics for choosing ϕ

A plausible way to approximate $\Pr(\mathbf{V} = \mathbf{v})$, is by treating \mathbf{V} as the outcome of a Markov process. Here we use a Markov process with memory μ , and assume that $p_s(\mathbf{u}|\mathbf{w})$ can be “well” approximated by $p_s(\mathbf{u}|\mathbf{w}_{r-\mu:r-1})$, for vectors $\mathbf{u} \in V_{\mathcal{E}}^\ell$, $\mathbf{w} \in V_{\mathcal{E}}^r$, with r, ℓ as above, and $r \geq \mu$. Using this approximation we get

$$\begin{aligned} \Pr(\mathbf{V}^{(m)} = \mathbf{v}) &\approx p_0^{(m)}(\mathbf{v}_{0:\mu-1}) \\ &\cdot \prod_{i=0}^{k-1} p_{\mu+i\alpha}^{(m)}(\mathbf{v}_{\mu+i\alpha:\mu+(i+1)\alpha-1} | \mathbf{v}_{i\alpha:\mu+i\alpha-1}). \end{aligned}$$

We hypothesize that for fixed vectors $\mathbf{u} \in V_{\mathcal{E}}^\alpha$, $\mathbf{w} \in V_{\mathcal{E}}^\mu$, as m gets large, the conditional probabilities $p_s^{(m)}(\mathbf{u}|\mathbf{w})$, for $0 \ll s \ll m-1$, are “approximately equal” to the value when s is in the “middle” of the interval $[0, m-1]$. We hypothesize that this holds for “most” of the integers s in that interval and moreover that this middle value converges as m gets large. Accordingly, we try to approximate the conditional probability $p_s^{(m)}(\mathbf{u}|\mathbf{w})$ by the conditional probability found in the “middle” of a “tall” horizontal strip. More precisely, we fix an integer $\delta \geq 0$, set $\omega = 2\delta + \mu + \alpha$, and approximate $p_s^{(m)}(\mathbf{u}|\mathbf{w})$ by $p_{\delta+\mu}^{(\omega)}(\mathbf{u}|\mathbf{w})$. We also approximate $p_0^{(m)}(\mathbf{w})$ by $p_0^{(\omega)}(\mathbf{w})$. This gives us

$$\begin{aligned} \Pr(\mathbf{V}^{(m)} = \mathbf{v}) &\approx p_0^{(\omega)}(\mathbf{v}_{0:\mu-1}) \\ &\cdot \prod_{i=0}^{k-1} p_{\delta+\mu}^{(\omega)}(\mathbf{v}_{\mu+i\alpha:\mu+(i+1)\alpha-1} | \mathbf{v}_{i\alpha:\mu+i\alpha-1}), \end{aligned}$$

which, by (3.10), implies that

$$\begin{aligned} (\mathbf{r}_m)_{\mathbf{v}} &\approx \sqrt{p_0^{(\omega)}(\mathbf{v}_{0:\mu-1})} \\ &\cdot \prod_{i=0}^{k-1} \sqrt{p_{\delta+\mu}^{(\omega)}(\mathbf{v}_{\mu+i\alpha:\mu+(i+1)\alpha-1} | \mathbf{v}_{i\alpha:\mu+i\alpha-1})}. \end{aligned} \tag{3.11}$$

Set $F_m = |V_{\mathcal{E}}|^{m_k}$, and denote by $\widetilde{\mathbf{r}}_{m_k} \in \mathbb{R}^{F_m}$ the nonnegative real vector with entries indexed by $V_{\mathcal{E}}^{m_k}$ and given by the RHS of equation (3.11). Let $\phi : (V_{\mathcal{E}})^{\mu+\alpha} \rightarrow [0, \infty)$ be given by

$$\phi(\mathbf{u}) = \sqrt{p_{\delta+\mu}^{(\omega)}(\mathbf{u}_{\mu:\mu+\alpha-1} | \mathbf{u}_{0:\mu-1})} ; \quad \mathbf{u} \in (V_{\mathcal{E}})^{\mu+\alpha}, \tag{3.12}$$

and let $\mathbf{x}_{m_k} \in \mathbb{R}^{F_m}$ be the vector with entries indexed by $V_{\mathcal{E}}^{m_k}$ and defined by (3.4). We obtain

$$(\mathbf{r}_{m_k})_{\mathbf{v}} \approx (\widetilde{\mathbf{r}}_{m_k})_{\mathbf{v}} = (\mathbf{x}_{m_k})_{\mathbf{v}} \sqrt{p_0^{(\omega)}(\mathbf{v}_{0:\mu-1})} ; \quad \mathbf{v} \in (V_{\mathcal{E}})^{m_k}.$$

3.4. Heuristics for choosing ϕ

Now for $m_k \geq \omega$, if $\mathbf{v} \in V_{\mathcal{E}}^{m_k}$ is not an isolated vertex in G_{m_k} , then clearly, $\mathbf{v}_{0:\omega-1}$ is not an isolated vertex in G_{ω} as well. Therefore $(\mathbf{r}_{\omega})_{\mathbf{v}_{0:\omega-1}} > 0$, which implies that $p_0^{(\omega)}(\mathbf{v}_{0:\omega-1}) > 0$ and thus $p_0^{(\omega)}(\mathbf{v}_{0:\mu-1}) > 0$. Let $p_{\min} = p_{\min}^{(\omega)} = \min\{p_0^{(\omega)}(\mathbf{w}) : \mathbf{w} \in V_{\mathcal{E}}^{\mu} \text{ and } p_0^{(\omega)}(\mathbf{w}) > 0\}$. It follows that for all vertices $\mathbf{v} \in (V_{\mathcal{E}})^{m_k}$ of G_{m_k} that are not isolated, we have

$$(\widetilde{\mathbf{r}}_{m_k})_{\mathbf{v}} \geq \sqrt{p_{\min}} (\mathbf{x}_{m_k})_{\mathbf{v}}.$$

Now, for any positive integer ℓ and $(F_{m_k} \times 1)$ -real vector \mathbf{y} , the product $\mathbf{y}^{\mathsf{t}} H_{m_k}^{\ell} \mathbf{y}$ depends only on the values of the entries of \mathbf{y} indexed by non-isolated vertices of G_{m_k} . Consequently, we may write

$$p_{\min} \mathbf{x}_{m_k}^{\mathsf{t}} H_{m_k}^{\ell} \mathbf{x}_{m_k} \leq \widetilde{\mathbf{r}}_{m_k}^{\mathsf{t}} H_{m_k}^{\ell} \widetilde{\mathbf{r}}_{m_k} \leq \mathbf{x}_{m_k}^{\mathsf{t}} H_{m_k}^{\ell} \mathbf{x}_{m_k},$$

for all positive integers ℓ . Taking the log, dividing by m_k , and taking the limit as k approaches infinity, we obtain

$$\lim_{k \rightarrow \infty} \frac{\log \widetilde{\mathbf{r}}_{m_k}^{\mathsf{t}} H_{m_k}^{\ell} \widetilde{\mathbf{r}}_{m_k}}{m_k} = \lim_{k \rightarrow \infty} \frac{\log \mathbf{x}_{m_k}^{\mathsf{t}} H_{m_k}^{\ell} \mathbf{x}_{m_k}}{m_k},$$

where by Lemma 2, the limit in the RHS exists. Thus, choosing ϕ as given by (3.12) and computing the lower bound by the method described in Section 3.1 is equivalent to computing the limit of the lower bound in (3.2), with $\widetilde{\mathbf{r}}_m$ substituted for \mathbf{x}_m , as m approaches infinity. If $\widetilde{\mathbf{r}}_m$ approximates \mathbf{r}_m well enough, we expect to get good bounds. Note that we may use the heuristic described here even for constraints for which the graphs $G_m^{(\mathcal{H})}$ are not always irreducible. In this case, the geometric multiplicity of the Perron eigenvalue may be larger than 1, and there may be more than one choice of the vector \mathbf{r}_{ω} in the computation of $p_{\delta+\mu}^{(\omega)}(\cdot|\cdot)$. Regardless of our choice, we will get a nonnegative function ϕ and a lower bound on the capacity. In Section 3.5 we show numerical results obtained using the heuristic described here for several constraints.

3.4.2 General optimization

We may also use general optimization techniques to find functions ϕ which maximize the lower bound on the capacity. Fix integers $\mu \geq 0$ and $p, q, \alpha > 0$, and for a positive integer ℓ , set $\mathcal{D}_{\ell} = (V_{\mathcal{E}})^{\ell}$. In this subsection, we identify a function $\phi: \mathcal{D}_{\mu+\alpha} \rightarrow \mathbb{R}$ with a real vector $\phi \in \mathbb{R}^{|\mathcal{D}_{\mu+\alpha}|}$ indexed by $\mathcal{D}_{\mu+\alpha}$; for each $\mathbf{j} \in \mathcal{D}_{\mu+\alpha}$ we identify $\phi(\mathbf{j})$ with the entry $\phi_{\mathbf{j}}$. For a positive integer n , let \mathcal{G}_n be a deterministic presentation for $\mathcal{V}_n(S)$, $\mathcal{I}_n = \mathcal{I}(\mu, \alpha, n, \mathcal{G}_n, G_{\mathcal{E}})$, and for a function $\phi: \mathcal{D}_{\mu+\alpha} \rightarrow [0, \infty)$, set $A_{n,\phi} = A(\mathcal{I}_n, \mathcal{W}_{\phi})$. Observe that for a scalar $c \in [0, \infty)$,

3.4. Heuristics for choosing ϕ

$A_{n,c\phi} = c^2 A_{n,\phi}$. It follows that using $c\phi$ in place of ϕ in equation (3.7) of Theorem 1 does not change the lower bound. Consequently, (as ϕ cannot be the constant 0 function), it's enough to consider functions ϕ whose images (of all vectors in $(V_{\mathcal{E}})^{\mu+\alpha}$) sum to 1. We thus have the following optimization problem.

$$\begin{aligned} & \text{maximize} && (\log \lambda(A_{2q+p,\phi}) - \log \lambda(A_{2q,\phi})) / (p\alpha) \\ & \text{subject to} && \phi \geq \mathbf{0}, \\ & && \phi \cdot \mathbf{1} = 1, \end{aligned} \tag{3.13}$$

where $\mathbf{0}$ and $\mathbf{1}$ denote the real vectors of size $|\mathcal{D}_{\mu+\alpha}|$ with every entry equal to 0 and 1 respectively, and for two real vectors of the same size, \mathbf{t}, \mathbf{r} we write $\mathbf{t} \geq \mathbf{r}$ or $\mathbf{t} > \mathbf{r}$ if the corresponding inequality holds entry-wise.

Finding a global solution for a general optimization problem can be hard. We proceed to show that if we replace the constraint $\phi \geq \mathbf{0}$ with $\phi > \mathbf{0}$ in (3.13), thereby changing the feasible set and possibly decreasing the optimal solution, it can be formulated as an instance of a particular class of optimization problems known as ‘‘DC optimization’’ which may be easier to solve. Let d be a positive integer. A real-valued function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is called a *DC* (difference of convex) function, if it can be written as the difference of two real-valued convex functions on \mathbb{R}^d . An optimization problem of the form

$$\begin{aligned} & \text{maximize} && f(x) \\ & \text{subject to} && x \in X, \\ & && h_i(x) \leq 0 ; \quad i = 0, 1, \dots, \ell, \end{aligned}$$

where $X \subseteq \mathbb{R}^d$ is a convex closed subset of \mathbb{R}^d and the functions f, h_0, \dots, h_ℓ are DC functions, is called a *DC optimization* or *DC programming* problem. See [17] and the references within for an overview of the theory of DC optimization.

A nonnegative function $f : \mathbb{R}^d \rightarrow [0, \infty)$ is called *log-convex* or *superconvex*, if either $f(\mathbf{t}) > 0$ for all $\mathbf{t} \in \mathbb{R}^d$ and $\log f$ is convex in \mathbb{R}^d , or $f \equiv 0$. A log-convex function is convex, and in [24], it is shown that the class of log-convex functions is closed under addition, multiplication, raising to positive real powers, taking limits, and additionally that for a square matrix $A(\mathbf{t}) = (a_{i,j}(\mathbf{t}))$ whose entries are log-convex functions $a_{i,j} : \mathbb{R}^d \rightarrow [0, \infty)$, the function $\mathbf{t} \rightarrow \lambda(A(\mathbf{t}))$ is log-convex as well.

Now, observe, that for a positive integer n , every entry of $A_{n,\phi}$ is a quadratic form in the entries $\phi(\mathbf{j})$, $\mathbf{j} \in \mathcal{D}_{\mu+\alpha}$, with nonnegative integer coefficients. Such a function is generally not log-convex. To fix this, we perform the change of variables $\phi = e^\psi$, where ψ is a real-valued function $\psi : \mathcal{D}_{\mu+\alpha} \rightarrow \mathbb{R}$. Note that by doing so, we added the constraint $\phi > \mathbf{0}$. Since every entry of ϕ is now positive,

3.5. Numerical results for selected constraints

we may replace the constraint $\phi \cdot \mathbf{1} = 1$ by the constraint $\phi(\mathbf{v}_0) = 1$ or equivalently $\psi(\mathbf{v}_0) = 0$, for some fixed $v_0 \in \mathcal{D}_{\mu+\alpha}$. Problem (3.13) with the additional constraint $\phi > \mathbf{0}$, can now be rewritten as

$$\begin{aligned} & \text{maximize} && (\log \lambda(A_{2q+p, e^\psi}) - \log \lambda(A_{2q, e^\psi})) / (p\alpha) \\ & \text{subject to} && \psi(\mathbf{v}_0) = 0. \end{aligned} \tag{3.14}$$

Obviously, we may substitute the maximization problem constraint, $\psi(\mathbf{v}_0) = 0$, above into the objective function, thereby reducing the number of variables by 1; however, this is not relevant for the discussion, so, for simplicity, we do not do so here. Now, for a positive integer n , the entries of the matrix A_{n, e^ψ} are of the form

$$\sum_{k=1}^{q_{i,j}} e^{\psi(\mathbf{w}_{k,i,j}) + \psi(\mathbf{u}_{k,i,j})},$$

where $q_{i,j}$ are nonnegative integers, and $\mathbf{w}_{k,i,j}$ and $\mathbf{u}_{k,i,j}$ are vectors in $\mathcal{D}_{\mu+\alpha}$, for all $i, j \in V_{\mathcal{I}_n}$ and integers $1 \leq k \leq q_{i,j}$. It can be verified that a function of such a form is log-convex in ψ . It follows that the function $\psi \rightarrow \lambda(A_{n, e^\psi})$ for $\psi \in \mathbb{R}^{|\mathcal{D}_{\mu+\alpha}|}$ is log-convex as well. Therefore either $\lambda(A_{n, e^\psi}) \equiv 0$, or $\lambda(A_{n, e^\psi}) > 0$ for all $\psi \in \mathbb{R}^{|\mathcal{D}_{\mu+\alpha}|}$. In particular, for $\psi \equiv 0$ the matrix $A_{n, \mathbf{1}}$ is the adjacency matrix of the graph \mathcal{I}_n . Since \mathcal{I}_n is deterministic, and for every nonnegative integer ℓ , the set of labels of its paths of length ℓ is $S_{\ell\alpha \times n}$, it follows that $(1/\alpha) \log \lambda(A_{n, \mathbf{1}}) = \text{cap}(\mathcal{V}_n(S)) \geq \text{cap}(S)$ (the latter inequality follows from (2.8)). Hence, if $\text{cap}(S) > -\infty$ (or equivalently $\text{cap}(S) \geq 0$), then $\lambda(A_{n, e^\psi}) > 0$ for all $\psi \in \mathbb{R}^{|\mathcal{D}_{\mu+\alpha}|}$ and $\log \lambda(A_{n, e^\psi})$ is a convex function of ψ . Clearly $\text{cap}(S) > -\infty$ iff. $S_{m \times n} \neq \emptyset$, for all positive integers m, n . We thus obtain the following theorem.

Theorem 3. *Let S be a constraint such that for all positive integers m, n , $S_{m \times n} \neq \emptyset$ then Problem (3.14) is a DC optimization problem.*

3.5 Numerical results for selected constraints

In this section we give numerical lower bounds on the capacity of some 2-dimensional constraints obtained using the method presented in the sections above. The constraints considered are NAK, RWIM, EVEN^{⊗2}, and CHG(3)^{⊗2}. Table 3.2 summarizes the best lower bounds obtained using our method. For comparison, we provide the best lower bounds that we could obtain using other methods. We also give upper bounds on the capacity of these constraints obtained using the method of [3]. Table 3.3 shows the lower bounds obtained using our max-entropic probability heuristic for choosing ϕ , described in Section 3.4.1. Table 3.4 shows the

3.5. Numerical results for selected constraints

lower bounds obtained with our method by trying to solve the optimization problem described in Section 3.4.2. In this, we did not make use of the DC property of the optimization problem; instead, we used a generic sub-optimal optimization algorithm whose results are not guaranteed to be global solutions. Utilizing special algorithms for solving DC optimization problems may give better lower bounds. The rightmost column of each of these tables shows the lower bound calculated for the same values of p and q using the method of [3, 7]. The largest lower-bound obtained for each constraint is marked with a ‘*’. In the next subsections we give remarks specific to some of these constraints.

As can be seen from the results in Table 3.4, using the optimization heuristic with our method gives better lower bounds on the capacity than the method of [3, 7]. On the other hand, using optimization typically requires many evaluations of the objective function which results in longer running times compared to the method of [3, 7], for the same values of p and q . Increasing the value of q in the method of [3, 7] usually gives better lower bounds but requires more time and memory to run. It is therefore relevant to check if their method can be used to attain the lower bounds, obtained using the technique described here, with the same computational resources. Our experiments show that this is typically not the case. For example, for the NAK constraint, when $\mu = 1$, $\alpha = 1$, $p = 1$ and $q = 4$, an implementation of the method described here with the optimization heuristic took roughly 70 seconds to run, required operating on matrices of size 144×144 and resulted with the lower bound 0.4250767727. In contrast, our implementation of the method of [3, 7] with $p = 1$, and $q = 8$, took roughly 1000 seconds to run, required operating on matrices of size 6765×6765 and resulted with the lower bound 0.4250725619. Table 3.1 shows some of the lower bounds obtained using the method of [3, 7] along with the size of the largest square matrix involved in the computation.

The numerical results were computed using the eigenvalue routines in Matlab and rounded (down for lower bounds and up for upper-bounds) to 10 decimal places. Given accuracy problems with possibly defective matrices, we verified the results using the technique described in [35, Section IV].

Table 3.1: Matrix size in the method of [3, 7] for the NAK constraint.

p	q	Lower bound using [3, 7]	Largest matrix size
1	5	0.4248771038	377×377
2	4	0.4249055702	233×233
1	6	0.4250215987	987×987
1	7	0.4250615286	2584×2584
2	6	0.4250636891	1597×1597
1	8	0.4250725619	6765×6765

3.5.1 The constraint RWIM

Observe that this constraint has both symmetric horizontal and symmetric vertical vertex-constrained strips. Thus, we can apply our method in the vertical as well as the horizontal direction to get lower bounds. Clearly, $\text{cap}(\text{RWIM}^\dagger) = \text{cap}(\text{RWIM})$, so we can obtain additional lower bounds on $\text{cap}(\text{RWIM})$ by using our method to get lower bounds on $\text{cap}(\text{RWIM}^\dagger)$. Some of these bounds are given in Tables 3.3 and 3.4.

3.5.2 The constraint $\text{EVEN}^{\otimes 2}$

We used the reduction described in Section 3.3 with $\mathcal{G}_{\text{EVEN}}$ being the presentation of EVEN given in Figure 1.1a, to get lower bounds on the capacity of $\text{EVEN}^{\otimes 2}$. Table 3.4 gives the results obtained with our method using the optimization described in Section 3.4.2. We also used the method with the max-entropic probability heuristic of Section 3.4.1 and the results are given in Table 3.3.

3.5.3 The constraint $\text{CHG}(b)^{\otimes 2}$

For this constraint, the case $b=1$ is degenerate. Indeed, there are exactly two $m \times n$ arrays in $\text{CHG}(1)^{\otimes 2}$ for all positive integers m and n , and consequently, $\text{cap}(\text{CHG}(1)^{\otimes 2})=0$. For $b=2$, we show in Theorem 5 in Chapter 4 that the capacity is exactly $1/4$. For $b=3$, we used the reduction of Section 3.3 with $\mathcal{G}_{\text{CHG}(3)}$ being the presentation of $\text{CHG}(3)$ given in Figure 1.1c, to get lower bounds on the capacity of $\text{CHG}(3)^{\otimes 2}$. Table 3.4 gives the results obtained with our method using the optimization described in Section 3.4.2. We also used the method with the max-entropic probability heuristic of Section 3.4.1 and the results are given in Table 3.3.

Table 3.2: Best bounds on capacities of certain constraints.

Constraint	Previous best lower bound	New lower bound	Upper bound
NAK	0.4250725619**	0.4250767745	0.4250767997*
RWIM	0.5350150***	0.5350151497	0.5350428519*
$\text{EVEN}^{\otimes 2}$	0.4385027973**	0.4402086447	0.4452873312*
$\text{CHG}(3)^{\otimes 2}$	0.4210209862**	0.4222689819	0.5328488954*

*Calculated using the method of [3].

**Calculated using the method of [3, 7].

***Appears in [42].

3.5. Numerical results for selected constraints

Table 3.3: Lower bounds using max-entropic probability heuristic (Section 3.4.1).

Constraint	δ	μ	α	p	q	Lower bound	Using [3, 7]
NAK	3	1	1	1	5	0.4250766244	0.4248771038
	3	1	1	2	4	0.4250766446	0.4249055702
	6	1	1	1	5	0.4250767227	0.4248771038
	3	3	4	1	5	0.4250767590	0.4248771038
	7	1	1	1	5	0.4250767617	0.4248771038
	3	1	1	2	6	0.4250767647	0.4250636891
	3	1	4	1	5	0.4250767733	0.4248771038
	5	1	1	1	5	0.4250767744	0.4248771038
	3	1	4	2	6	0.4250767745	0.4250636891
	3	1	2	2	6	0.4250767745	0.4250636891
RWIM	3	1	3	1	6	0.5350147968	0.5235145644
	1	1	1	3	6	0.5350148753	0.5318753627
	3	2	2	1	5	0.5350148814	0.5160533001
	3	1	2	2	6	0.5350149069	0.5337927416
	2	1	2	2	6	0.5350149071	0.5337927416
	0	1	2	2	6	0.5350149136	0.5337927416
	2	2	2	1	5	0.5350149271	0.5160533001
	1	1	2	2	6	0.5350149462	0.5337927416
	1	1	3	1	6	0.5350149525	0.5235145644
	1	1	1	1	7	0.5350149707	0.5280406048
RWIM ^t	4	1	3	2	4	0.5350145937	0.5350144722
	1	1	1	1	5	0.5350146612	0.5350149478
	4	2	1	1	4	0.5350147212	0.5350142142
	3	1	1	1	5	0.5350147328	0.5350149478
	5	1	1	1	5	0.5350147619	0.5350149478
	2	2	1	1	4	0.5350147969	0.5350142142
	4	1	1	1	5	0.5350148255	0.5350149478
	2	1	1	1	5	0.5350148449	0.5350149478
	0	1	1	1	5	0.5350148814	0.5350149478
	0	2	1	1	4	0.5350148980	0.5350142142
EVEN ^{⊗2}	3	2	1	1	3	0.4383238232	0.4347423815
	3	1	1	1	4	0.4383243738	0.4367818624
	3	1	3	2	3	0.4383632350	0.4356897662
	3	1	2	4	3	0.4383838005	0.4364303826
	3	1	1	2	4	0.4384647082	0.4371709990
	3	1	3	3	3	0.4384906740	0.4360537982
	3	1	2	1	4	0.4385448358	0.4367818624
	3	1	2	2	4	0.4386655840	0.4371709990
	3	1	3	1	4	0.4387455520	0.4367818624
	CHG(3) ^{⊗2}	0	0	1	1	2	0.4188210386
0		0	1	1	4	0.4222689819*	0.4197053158

*Best lower bound.

3.5. Numerical results for selected constraints

Table 3.4: Lower bounds using optimization (Section 3.4.2).

Constraint	μ	α	p	q	Lower bound	Using [3, 7]
NAK	2	1	2	4	0.4250767692	0.4249055702
	1	2	1	5	0.4250767736	0.4248771038
	1	1	3	4	0.4250767737	0.4248960814
	1	2	1	3	0.4250767739	0.4224650194
	1	1	4	4	0.4250767739	0.4249674993
	1	1	5	4	0.4250767740	0.4249783192
	1	1	6	4	0.4250767741	0.4249995626
	1	2	3	3	0.4250767743	0.4244240822
	1	2	6	3	0.4250767744	0.4247979797
	1	2	2	5	0.4250767745*	0.4250294285
RWIM	1	1	1	3	0.5350147515	0.4832292495
	1	1	2	3	0.5350148675	0.5300373650
	1	1	3	3	0.5350149371	0.5212673183
	1	1	1	4	0.5350150805	0.5037272248
	1	1	2	4	0.5350151001	0.5318663054
	1	1	3	4	0.5350151123	0.5265953036
	1	1	1	5	0.5350151372	0.5160533001
	1	1	2	5	0.5350151410	0.5330440001
	1	1	2	6	0.5350151491	0.5337927416
	1	2	1	5	0.5350151497*	0.5160533001
RWIM ^t	1	2	4	3	0.5350151364	0.5350130576
	1	2	3	4	0.5350151377	0.5350146307
	1	2	5	3	0.5350151392	0.5350134356
	2	1	1	4	0.5350151405	0.5350142142
	1	1	1	5	0.5350151442	0.5350149478
	1	2	1	4	0.5350151465	0.5350142142
	1	2	1	5	0.5350151481	0.5350149478
	1	2	2	4	0.5350151482	0.5350144722
	1	3	1	4	0.5350151483	0.5350142142
EVEN ^{⊗2}	1	1	1	3	0.4395381520	0.4347423815
	1	1	2	3	0.4397347451	0.4356897662
	1	1	1	4	0.4402086447*	0.4367818624
CHG(3) ^{⊗2}	0	1	1	2	0.4189237100	0.4101473707
	0	1	2	2	0.4197037681	0.4182017399
	0	1	3	2	0.4201450063	0.4176642274
	0	1	1	3	0.4210954837	0.4165892023
	0	1	2	3	0.4214748454	0.4210209862
				1	4	

*Best lower bound.

3.6 Open questions

The authors of [9, 10, 35] show how the methods of [7] and [3] can be extended to get lower and upper bounds on the capacity of constraints with dimension larger than 2 which are “symmetric in all but one direction”. Similarly, it should be a relatively easy exercise to extend our generalization of the method to obtain better lower bounds to the capacities of such constraints.

For a \mathcal{D} -dimensional constraint S over Σ , one can define a more general notion than capacity called “weighted-capacity” or “pressure” [10, 11]. First, one assigns a positive weight to each symbol in Σ . Then, the weight of a \mathcal{D} -dimensional array Γ over Σ , denoted $\mathcal{W}(\Gamma)$ is defined to be the product of the weights of its entries. The *weighted-capacity* of S is now given by

$$\lim_{\mathbf{m} \rightarrow \infty} \frac{\sum_{\Gamma \in S_{\mathbf{m}}} \mathcal{W}(\Gamma)}{|\mathbf{m}|}.$$

As before, the limit exists since the numerator is an entry-wise subadditive function. It would be beneficial to extend the method of this chapter to obtain lower bounds on the weighted-capacity of symmetric constraints.

Finally, consider the method of [3] for obtaining upper-bounds on the capacity of 2-dimensional symmetric constraints. While the method gives better upper-bounds than those obtained by computing the normalized capacity of a horizontal or vertical strip, empirical results suggest that for many constraints the upper-bounds approach the capacity slower than their lower-bound counterparts. Hence, a better method for upper bounding the capacity of symmetric (and general) constraints would be useful.

Chapter 4

Exact computation of capacity*

As already stated, finding the exact capacity of \mathfrak{D} -dimensional constraints, for $\mathfrak{D} > 1$, is hard, and the list of constraints for which the capacity is known precisely is quite small. In this chapter we add two families of isotropic multidimensional constraints to this list, namely $\text{ODD}^{\otimes \mathfrak{D}}$ and $\text{CHG}(2)^{\otimes \mathfrak{D}}$. Some of the results presented here have been extended to other similar constraints in [22].

4.1 The capacity of $\text{ODD}^{\otimes \mathfrak{D}}$

Theorem 4. *For all positive integers \mathfrak{D} ,*

$$\text{cap}\left(\text{ODD}^{\otimes \mathfrak{D}}\right) = \frac{1}{2}.$$

Proof. Let S be the \mathfrak{D} -dimensional constraint $\text{ODD}^{\otimes \mathfrak{D}}$. We first show $\text{cap}(S) \geq 1/2$. For an integer n , let $Y_n \subseteq [2n]^{\mathfrak{D}}$ be the set of all vectors in $[2n]^{\mathfrak{D}}$ whose entries sum to an even number, and let X_n be the set of all binary \mathfrak{D} -dimensional arrays Γ of size $2n \times 2n \times \dots \times 2n$, with entries satisfying $(\Gamma)_j = 0$ for all $j \in Y_n$. Then the number of zeros between consecutive ‘1’s, in any row of an array in X_n is odd since it must be of the form $i - j - 1$ for some integers i, j —either both odd, or both even. Thus, all such arrays satisfy the constraint S , and since $|X_n| = 2^{(2n)^{\mathfrak{D}} - |Y_n|} = 2^{(2n)^{\mathfrak{D}}/2}$, we have $|S_{2n \times 2n \times \dots \times 2n}| \geq 2^{(2n)^{\mathfrak{D}}/2}$ for all positive integers n , which implies $\text{cap}(S) \geq 1/2$.

It remains to show that $\text{cap}(S) \leq 1/2$. Since $\text{cap}(\text{ODD}^{\otimes \mathfrak{D}})$ is non-increasing in \mathfrak{D} , it’s enough to show $\text{cap}(S) \leq 1/2$ for $\mathfrak{D} = 1$. Let n be a positive integer. It can be easily verified that any 1-dimensional array $\Gamma \in \text{ODD}_n$ with entries indexed by $[n]$, satisfies either $\Gamma_j = 0$ for all even integers $j \in [n]$, or $\Gamma_j = 0$ for all odd integers $j \in [n]$. It follows that $|\text{ODD}_n| \leq 2^{\lceil n/2 \rceil} + 2^{\lfloor n/2 \rfloor}$ which implies the desired inequality ■

*A version of this chapter has been published. Loudior, E. and Marcus, B.H. (2010) Improved Lower Bounds on Capacities of Symmetric 2-Dimensional Constraints using Rayleigh Quotients. IEEE Transactions on Information Theory 56:1624–1639.

4.2 The capacity of $\text{CHG}(2)^{\otimes \mathfrak{D}}$

Theorem 5. For all positive integers \mathfrak{D} ,

$$\text{cap}\left(\text{CHG}(2)^{\otimes \mathfrak{D}}\right) = \frac{1}{2^{\mathfrak{D}}}.$$

Proof. Let $S = \text{CHG}(2)^{\otimes \mathfrak{D}}$. We first show that $\text{cap}(S) \geq 1/2^{\mathfrak{D}}$. Let $\Gamma^{(0)}, \Gamma^{(1)}$ be the \mathfrak{D} -dimensional arrays of size $2 \times 2 \times \dots \times 2$ with entries indexed by $\{0, 1\}^{\mathfrak{D}}$ and given by

$$\left(\Gamma^{(i)}\right)_{\mathbf{j}} = (-1)^{i+\mathbf{j} \cdot \mathbf{1}} \quad ; \quad \mathbf{j} \in \{0, 1\}^{\mathfrak{D}},$$

where as usual $\mathbf{1}$ denotes the \mathfrak{D} -dimensional vector with every entry equal to 1. Observe that the sum of every row of both of these arrays is zero. Now, let n be a positive integer. For any \mathfrak{D} -dimensional array of size $n \times n \times \dots \times n$ with entries in $\{0, 1\}$, it can be easily verified that replacing every entry equal to 0 with $\Gamma^{(0)}$ and every entry equal to 1 with $\Gamma^{(1)}$ results in a \mathfrak{D} -dimensional array of size $2n \times 2n \times \dots \times 2n$ that satisfies S . It follows that $|S_{2n \times 2n \times \dots \times 2n}| \geq 2^{n^{\mathfrak{D}}}$ for all positive integers n , which implies $\text{cap}(S) \geq 1/2^{\mathfrak{D}}$.

We now show that $\text{cap}(S) \leq 1/2^{\mathfrak{D}}$. For a positive integer $n \geq 2$, denote by $\mathcal{N}_n^{(0)}$ the set of all even integers in $\{0, 1, \dots, n-2\}$ and by $\mathcal{N}_n^{(1)}$ the set of all odd integers in $\{0, 1, \dots, n-2\}$. We shall make use of the following lemma.

Lemma 3. Fix a positive integer $n \geq 2$, and let $(a_i)_{i=0}^{n-1} \subseteq \{+1, -1\}$ be a sequence of length n . Then $a_0 \dots a_{n-1} \in \text{CHG}(2)$ if and only if at least one of the following statements hold.

1. For all $i \in \mathcal{N}_n^{(0)}$, $a_i = -a_{i+1}$.
2. For all $i \in \mathcal{N}_n^{(1)}$, $a_i = -a_{i+1}$.

Proof. We first show the “if” direction. Let $(a_i)_{i=0}^{n-1} \subseteq \{+1, -1\}$ be a sequence for which at least one of statements 1,2 of the lemma holds. Then clearly, for any integers $0 \leq i < j < n$, all the terms in the sum $\sum_{k=i}^j a_k$, with the possible exception of the first and last terms, cancel. Therefore $|\sum_{k=i}^j a_k| \leq |a_i| + |a_j| = 2$ and $a_0 \dots a_{n-1} \in \text{CHG}(2)$.

As for the “only if” direction, let $(a_i)_{i=0}^{n-1} \subseteq \{+1, -1\}$ such that $a_0 \dots a_{n-1} \in \text{CHG}(2)$, and consider the presentation of the CHG constraint given in Figure 1.1c for $b = 2$ (and vertices $\{0, 1, 2\}$). Let $(e_i)_{i=0}^{n-1}$ be a path in this presentation generating $a_0 \dots a_{n-1}$. It’s easily verified that if $\sigma(e_i) = 1$, for some $i \in [n-1]$, then $a_j = -a_{j+1}$ for all integers $i \leq j \leq n-2$ such that $j \equiv i \pmod{2}$. Evidently, either $\sigma(e_0) = 1$ and so statement 1 holds, or $\sigma(e_1) = 1$ implying statement 2. ■

4.2. The capacity of $\text{CHG}(2)^{\otimes \mathfrak{D}}$

We now return to the claim that $\text{cap}(S) \leq 1/2^{\mathfrak{D}}$. Fix a positive integer $n \geq 2$. For an integer $1 \leq i \leq \mathfrak{D}$, let $\mathbf{e}^{(i)} \in \{0, 1\}^{\mathfrak{D}}$, be the vector, indexed by $\{1, 2, \dots, \mathfrak{D}\}$, containing 1 in its i th entry and 0 everywhere else and let $\mathcal{J}_i \subseteq [n]^{\mathfrak{D}}$ denote the subset of all the vectors indexed by $\{1, 2, \dots, \mathfrak{D}\}$ with a 0 in the i th entry. For a vector $\mathbf{j} \in \mathcal{J}_i$, the sequence $(\mathbf{j} + k\mathbf{e}^{(i)})_{k=0}^{n-1}$, is a sequence of indices of entries of a row in direction i of a \mathfrak{D} -dimensional $n \times n \times \dots \times n$ array, and we shall say that it is a sequence in direction i . Let $\mathbf{R}(n, \mathfrak{D})$ be the set of all such sequences for all integers $1 \leq i \leq \mathfrak{D}$ and vectors $\mathbf{j} \in \mathcal{J}_i$, and let $\mathbf{r} \in \{0, 1\}^{|\mathbf{R}(n, \mathfrak{D})|}$ be a binary vector indexed by $\mathbf{R}(n, \mathfrak{D})$. For the purpose of this proof, let us refer to a sequence $(a_i)_{i=0}^{n-1} \subseteq \{+1, -1\}$ as a phase-0 sequence if statement 1 of Lemma 3 holds, and as a phase-1 sequence if statement 2 holds (note that a sequence may be both a phase 0 and a phase 1 sequence). Also, we denote by $X(\mathbf{r}) \subseteq \{+1, -1\}^{*\mathfrak{D}}$, the set of all \mathfrak{D} -dimensional arrays Γ of size $n \times n \times \dots \times n$ for which the row $\Gamma_{\bar{\varrho}}$ is a phase- $\mathbf{r}_{\bar{\varrho}}$ sequence, for all $\bar{\varrho} = (\mathbf{j} + k\mathbf{e}^{(i)})_{k=0}^{n-1} \in \mathbf{R}(n, \mathfrak{D})$. Then by Lemma 3, we have

$$S_{n \times \dots \times n} = \bigcup_{\mathbf{r}} X(\mathbf{r}). \quad (4.1)$$

We shall give an upper bound on the size of $X(\mathbf{r})$. For a vector $\mathbf{v} \in [n]^{\mathfrak{D}}$, denote by $\rho(i, \mathbf{v})$ the unique sequence in direction i in $\mathbf{R}(n, \mathfrak{D})$ that has \mathbf{v} as one of its elements. Let $T_{\mathbf{r}, i} : [n]^{\mathfrak{D}} \rightarrow \mathbb{Z}^{\mathfrak{D}}$ be given by:

$$T_{\mathbf{r}, i}(\mathbf{v}) = \begin{cases} \mathbf{v} + \mathbf{e}^{(i)} & \text{if } v_i \equiv \mathbf{r}_{\rho(i, \mathbf{v})} \pmod{2} \\ \mathbf{v} - \mathbf{e}^{(i)} & \text{otherwise} \end{cases},$$

$$\mathbf{v} \in [n]^{\mathfrak{D}}, \mathbf{v} = (v_1, \dots, v_{\mathfrak{D}}).$$

Next, we define the undirected graph $G_{\mathbf{r}} = (V, E_{\mathbf{r}})$ (without parallel edges), with vertices given by

$$V = [n]^{\mathfrak{D}},$$

and edges given by

$$E_{\mathbf{r}} = \left\{ \mathbf{u} - \mathbf{v} : \begin{array}{l} \mathbf{u}, \mathbf{v} \in V \text{ and } \mathbf{v} = T_{\mathbf{r}, i}(\mathbf{u}) \\ \text{for some integer } 1 \leq i \leq \mathfrak{D} \end{array} \right\},$$

where $\mathbf{u} - \mathbf{v}$ denotes the undirected edge connecting vertices \mathbf{u}, \mathbf{v} . It's easy to verify that an array $\Gamma \in \{+1, -1\}^{*\mathfrak{D}}$ of size $n \times n \times \dots \times n$ is in $X(\mathbf{r})$ iff for every edge $\mathbf{u} - \mathbf{v} \in E_{\mathbf{r}}$, it holds that $\Gamma_{\mathbf{u}} = -\Gamma_{\mathbf{v}}$. Figure 4.1 shows an example of $G_{\mathbf{r}}$ for $\mathfrak{D} = 2$.

Let C_1, \dots, C_{ℓ} be the connected components of $G_{\mathbf{r}}$, and let $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(\ell)}$ be arbitrary vertices such that $\mathbf{v}^{(i)} \in C_i$, for $i = 1, 2, \dots, \ell$. It follows that for every

4.2. The capacity of $\text{CHG}(2)^{\otimes \mathfrak{D}}$

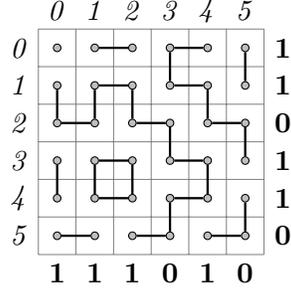


Figure 4.1: Example of the graph $G_{\mathbf{r}}$ for $\mathfrak{D} = 2$, $n = 6$. Each entry of \mathbf{r} corresponding to a row (column) is written to the right of it (below it). The index of each row (column) is written to its left (above it).

vector $\mathbf{b} = (b_1, \dots, b_\ell) \in \{+1, -1\}^\ell$, there exists at most one array $\Gamma \in X(\mathbf{r})$ satisfying $\Gamma_{\mathbf{v}^{(i)}} = b_i$ for all $i \in \{1, 2, \dots, \ell\}$, and consequently, $|X(\mathbf{r})| \leq 2^\ell$ (in fact, while this is not needed for the proof, such an array $\Gamma \in X(\mathbf{r})$ does exist, for any choice of \mathbf{b} , since each C_i is bipartite; thus $|X(\mathbf{r})| = 2^\ell$).

Now, let $\mathbf{u} = (u_1, \dots, u_{\mathfrak{D}}) \in ([n] \setminus \{0, n-1\})^{\mathfrak{D}}$ be a vertex in the “interior” of $G_{\mathbf{r}}$. We show that the connected component of $G_{\mathbf{r}}$ containing \mathbf{u} has at least $2^{\mathfrak{D}}$ vertices. To this end, we match each word $\mathbf{w} = w_1 w_2 \dots w_{\mathfrak{D}} \in \{0, 1\}^{\mathfrak{D}}$, with a sequence of vertices $(\pi^{(\mathbf{w}, j)})_{j=0}^{\mathfrak{D}} \subseteq V$, defined recursively by

$$\pi^{(\mathbf{w}, j)} = \begin{cases} \mathbf{u} & \text{if } j=0 \\ \pi^{(\mathbf{w}, j-1)} & \text{if } j > 0 \text{ and } w_j = 0 \\ T_{\mathbf{r}, j}(\pi^{(\mathbf{w}, j-1)}) & \text{if } j > 0 \text{ and } w_j = 1 \end{cases} .$$

It’s easy to verify that since every $1 \leq u_i \leq n-2$, the sequence is well-defined and indeed $(\pi^{(\mathbf{w}, j)})_{j=0}^{\mathfrak{D}} \subseteq [n]^{\mathfrak{D}}$. Clearly, the sequence is contained entirely in the connected component containing \mathbf{u} , and so this component contains the vertex $\pi^{(\mathbf{w}, \mathfrak{D})}$. Write $\pi^{(\mathbf{w}, \mathfrak{D})} = (\pi_1^{(\mathbf{w}, \mathfrak{D})}, \dots, \pi_{\mathfrak{D}}^{(\mathbf{w}, \mathfrak{D})})$. Then for $i=1, 2, \dots, \mathfrak{D}$, it holds that $\pi_i^{(\mathbf{w}, \mathfrak{D})} = u_i$ if $w_i=0$, and $\pi_i^{(\mathbf{w}, \mathfrak{D})} = u_i \pm 1$ if $w_i=1$. Therefore, for two distinct words $\mathbf{w}, \mathbf{w}' \in \{0, 1\}^{\mathfrak{D}}$, the vertices $\pi^{(\mathbf{w}, \mathfrak{D})}$ and $\pi^{(\mathbf{w}', \mathfrak{D})}$ are distinct as well, and consequently there are $2^{\mathfrak{D}}$ such vertices. Thus, the connected component of $G_{\mathbf{r}}$ containing \mathbf{u} has at least $2^{\mathfrak{D}}$ vertices.

It follows that there are at most $n^{\mathfrak{D}}/2^{\mathfrak{D}}$ connected components of $G_{\mathbf{r}}$ containing a vertex in $\{1, 2, \dots, n-2\}^{\mathfrak{D}}$. There are at most $n^{\mathfrak{D}} - (n-2)^{\mathfrak{D}}$ connected components not containing a vertex in $\{1, 2, \dots, n-2\}^{\mathfrak{D}}$ and hence the total number of

4.2. The capacity of $\text{CHG}(2)^{\otimes \mathfrak{D}}$

connected components, ℓ , in $G_{\mathbf{r}}$ satisfies $\ell \leq n^{\mathfrak{D}}/2^{\mathfrak{D}} + n^{\mathfrak{D}} - (n-2)^{\mathfrak{D}}$. Hence,

$$|X(\mathbf{r})| \leq 2^{n^{\mathfrak{D}}/2^{\mathfrak{D}} + n^{\mathfrak{D}} - (n-2)^{\mathfrak{D}}}.$$

Since there are $2^{\mathfrak{D}n^{\mathfrak{D}-1}}$ binary vectors $\mathbf{r} \in \{0, 1\}^{|\mathbf{R}(n, \mathfrak{D})|}$, we obtain from (4.1)

$$\begin{aligned} |S_{n \times n \times \dots \times n}| &\leq \sum_{\mathbf{r}} |X(\mathbf{r})| \\ &\leq 2^{n^{\mathfrak{D}}/2^{\mathfrak{D}} + n^{\mathfrak{D}} - (n-2)^{\mathfrak{D}} + \mathfrak{D}n^{\mathfrak{D}-1}} \\ &= 2^{n^{\mathfrak{D}}(1/2^{\mathfrak{D}} + 1 - (1-2/n)^{\mathfrak{D}}) + \mathfrak{D}n^{\mathfrak{D}-1}}, \end{aligned}$$

and the result follows from (1.1). ■

Chapter 5

Multi-choice constraints and independence capacity*

In this chapter we generalize some of the concepts appearing in [37, 38] to more than 1 dimension and to non-binary alphabets. In particular we define the notion of *independence capacity* of a multidimensional constraint which, roughly, is the contribution to the capacity resulting from independence between symbols in arrays of the constraint. For the binary alphabet $\{0, 1\}$ this concept coincides with the *maximum insertion rate* defined in [38]. We also show that for a 1-dimensional constraint S of finite-type with 0 independence entropy, $\text{cap}(S^{\otimes \mathcal{D}})$ converges to 0 exponentially fast as \mathcal{D} approaches infinity.

5.1 Multi-choice constraints

Let $\widehat{\Sigma}$ denote the set of all nonempty subsets of Σ . Let S be a \mathcal{D} -dimensional constraint over Σ and for a \mathcal{D} -dimensional array $\widehat{\Gamma} \in \widehat{\Sigma}^{\mathbf{m}}$ for some $\mathbf{m} \in \mathbb{N}^{\mathcal{D}}$, define the set of possible choices of entries in $\widehat{\Gamma}$, denoted $\Phi(\widehat{\Gamma})$, by

$$\Phi(\widehat{\Gamma}) = \left\{ \Gamma \in \Sigma^{\mathbf{m}} : \text{For all } \mathbf{i} \in [\mathbf{m}], \Gamma_{\mathbf{i}} \in \widehat{\Gamma}_{\mathbf{i}} \right\}.$$

We define the *multi-choice set corresponding to S* , denoted \widehat{S} , by

$$\widehat{S} = \left\{ \widehat{\Gamma} \in \widehat{\Sigma}^{*\mathcal{D}} : \Phi(\widehat{\Gamma}) \subseteq S \right\},$$

and, as for constraints, we use the notation $\widehat{S}_{\mathbf{m}}$, for $\mathbf{m} \in \mathbb{N}^{\mathcal{D}}$, to denote the set of \mathcal{D} -dimensional arrays of size \mathbf{m} in \widehat{S} . Note that \widehat{S} is closed under taking of contiguous sub-arrays, and so it's plausible that \widehat{S} is a \mathcal{D} -dimensional constraint over $\widehat{\Sigma}$. Unfortunately, we do not know if this is true for general \mathcal{D} -dimensional constrained systems S . In this work, we are mostly interested in \widehat{S} for a constraint S which is an axial product of some \mathcal{D} 1-dimensional constraints. In this case, it

*A version of this chapter has been submitted for publication. Loudior, E., Marcus B.H. and Pavlov R. (2010) Independence Entropy of \mathbb{Z}^d Shift Spaces.

5.1. Multi-choice constraints

turns out that \widehat{S} is indeed a constrained system. This is an easy corollary of the next theorem which shows that for a 1-dimensional constraint S , \widehat{S} is indeed a 1-dimensional constraint. This is shown in [38] for $\Sigma = \{0, 1\}$. For completeness, we state and prove the theorem for general alphabets. We require the following definitions. Let Σ be a finite alphabet. For words $x, y \in \Sigma^*$ and integer n we use the conventional notation of $|x|$, xy , x^n to denote, respectively, the length of x , the word formed by concatenating y to the right of x , and the word formed by concatenating x to itself n times. We use ε to denote the empty word. Let S be a 1-dimensional constraint over Σ . For a word $x \in S$ the follower set of x , denoted $\mathcal{F}(x) = \mathcal{F}_S(x)$ is given by

$$\mathcal{F}_S(x) = \{w \in \Sigma^* : xw \in S\}.$$

If $\mathcal{G} = ((V, E), \mathcal{L})$ is a presentation of S , and for $v \in V$, we denote by $\mathcal{F}(v)$ the set of words generated by paths in \mathcal{G} starting at v , then $\mathcal{F}_S(x) = \cup_v \mathcal{F}(v)$, where the union is taken over all terminal vertices, v , of paths in \mathcal{G} generating x . It follows that the number of follower sets of a constraint is finite. In [38] the authors construct a presentation of \widehat{S} when $\Sigma = \{0, 1\}$. We generalize their construction here for arbitrary alphabets and denote it by $\widehat{\mathcal{G}}_{\mathcal{F}_S} = ((V_{\mathcal{F}_S}, E_{\mathcal{F}_S}), \mathcal{L}_{\mathcal{F}_S})$. The set of vertices $V_{\mathcal{F}_S}$ consists of all (finite) intersections of follower sets of words in S :

$$V_{\mathcal{F}_S} = \left\{ \bigcap_{i=1}^k \mathcal{F}_S(w_i) : w_1, \dots, w_k \in S, k=1, 2, \dots \right\}.$$

For a vertex $v \in V_{\mathcal{F}_S}$, $v = \bigcap_{i=1}^k \mathcal{F}_S(w_i)$, and symbol $\widehat{a} \in \widehat{\Sigma}$ with $\widehat{a} \subseteq v$ define

$$\delta(v, \widehat{a}) = \bigcap_{a \in \widehat{a}} \bigcap_{i=1}^k \mathcal{F}_S(w_i a)$$

(note that $w_i a \in S$ for every $a \in \widehat{a}$ and $i = 1, 2, \dots, k$). It's easy to verify that $\delta(v, \widehat{a}) = \{w \in \Sigma^* : aw \in v \text{ for all } a \in \widehat{a}\}$, and therefore $\delta(v, \widehat{a})$ does not depend on the choice of k and w_1, \dots, w_k . The set $E_{\mathcal{F}_S}$ is now defined by

$$E_{\mathcal{F}_S} = \{(v, \widehat{a}, \delta(v, \widehat{a})) : v \in V_{\mathcal{F}_S}, \widehat{a} \subseteq v\},$$

and for an edge $e = (v, \widehat{a}, \delta(v, \widehat{a})) \in E_{\mathcal{F}_S}$ we define $\sigma(e) = v$, $\tau(e) = \delta(v, \widehat{a})$ and $\mathcal{L}_{\mathcal{F}_S}(e) = \widehat{a}$. We are now ready to prove the aforementioned theorem.

Theorem 6. *If S is a 1-dimensional constraint over an alphabet Σ , then \widehat{S} is a 1-dimensional constraint over $\widehat{\Sigma}$ and $\widehat{\mathcal{G}}_{\mathcal{F}_S}$ is a presentation of \widehat{S} .*

5.2. Independence capacity

Proof. The theorem readily follows from the following fact, which is easily verified by induction on the length of w .

Fact 1. *For any vertex $v = \bigcap_{i=1}^k \mathcal{F}_S(w_i) \in V_{\mathcal{F}_S}$, and a word $w \in \widehat{\Sigma}^*$, there is a path starting at v and generating w in $\widehat{\mathcal{G}}$ if and only if $\Phi(w) \subseteq v$, in which case the path ends at the vertex $\bigcap_{i=1}^k \bigcap_{z \in \Phi(w)} \mathcal{F}_S(w_i z)$.*

Now, for a word $\widehat{x} \in \widehat{\Sigma}^*$ generated by a path of $\widehat{\mathcal{G}}$ starting at some vertex v , we have $\Phi(\widehat{x}) \subseteq v \subseteq S$, and therefore $\widehat{x} \in \widehat{S}$. Conversely, since $S = \mathcal{F}(\varepsilon)$, it follows from the above fact that there is a path generating any word $\widehat{x} \in \widehat{S}$ (starting from $\mathcal{F}(\varepsilon)$) in $\widehat{\mathcal{G}}$. \blacksquare

Let $S^{(1)}, \dots, S^{(\mathfrak{D})}$ be 1-dimensional constraints over Σ , and set $S = S^{(1)} \otimes \dots \otimes S^{(\mathfrak{D})}$. Then it's easy to verify that $\widehat{S} = \widehat{S}^{(1)} \otimes \dots \otimes \widehat{S}^{(\mathfrak{D})}$, and we have the following corollary:

Corollary 2. *If $S^{(1)}, \dots, S^{(\mathfrak{D})}$ are 1-dimensional constraints over Σ , and $S = S^{(1)} \otimes \dots \otimes S^{(\mathfrak{D})}$, then \widehat{S} is a \mathfrak{D} -dimensional constrained system.*

5.2 Independence capacity

In this section, we introduce the notion of “independence capacity” of a constraint. Roughly, this is the part of the capacity resulting from inter-symbol independence in elements of the constraint.

Let S be a \mathfrak{D} -dimensional constraint over Σ . For an array $\widehat{\Gamma} \in \widehat{S}$ of size \mathbf{m} , the real number $\log(|\Phi(\widehat{\Gamma})|/|\mathbf{m}|)$ can be thought of as the contribution to the capacity resulting from independence between entries in elements of S as “captured” by $\widehat{\Gamma}$. We define the independence capacity as the limit (as $\mathbf{m} \rightarrow \infty$) of the maximum possible such contribution. Precisely, observe that the mapping $\mathbf{m} \rightarrow \log \max \{ |\Phi(\widehat{\Gamma})| : \widehat{\Gamma} \in \widehat{S}_{\mathbf{m}} \}$ for $\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}$ is entry-wise subadditive. Using Lemma 1, we define the *independence capacity* of S , denoted $\text{cap}_{\text{ind}}(S)$, by

$$\begin{aligned} \text{cap}_{\text{ind}}(S) &= \lim_{\mathbf{m} \rightarrow \infty} \frac{\log \max \{ |\Phi(\widehat{\Gamma})| : \widehat{\Gamma} \in \widehat{S}_{\mathbf{m}} \}}{|\mathbf{m}|} \\ &= \inf_{\mathbf{m}} \frac{\log \max \{ |\Phi(\widehat{\Gamma})| : \widehat{\Gamma} \in \widehat{S}_{\mathbf{m}} \}}{|\mathbf{m}|}. \end{aligned} \quad (5.1)$$

As already mentioned [38] defines the independence capacity of a 1-dimensional constraint, when $\Sigma = \{0, 1\}$, and calls it the “maximum insertion rate”. It is

5.2. Independence capacity

shown that the maximum insertion rate of a 1-dimensional constraint S can be determined from a presentation $\widehat{\mathcal{G}}$ of \widehat{S} , when $\Sigma = \{0, 1\}$. The next theorem shows the generalization of this result to larger alphabets.

Let $\mathcal{G} = ((V, E), \mathcal{L})$ be a labeled graph. A cycle $(e_i)_{i=1}^\ell \subseteq E$ is *simple*, if the vertices $\tau(e_1), \dots, \tau(e_\ell)$ are distinct.

Theorem 7. *Let S be a 1-dimensional constraint over an alphabet Σ , and pick any presentation $\widehat{\mathcal{G}} = ((V, E), \mathcal{L})$ for \widehat{S} . Then*

$$\text{cap}_{\text{ind}}(X) = \max \left\{ \frac{\log |\Phi(\widehat{w})|}{|\widehat{w}|} : \widehat{w} \text{ is generated by a simple cycle of } \widehat{\mathcal{G}} \right\}. \quad (5.2)$$

Proof. Denote the RHS of (5.2) by ν^* . Let \widehat{w}_* be a word generated by a simple cycle of $\widehat{\mathcal{G}}$ such that $\nu^* = (\log |\Phi(\widehat{w}_*)|)/|\widehat{w}_*|$. Set $\ell = |\widehat{w}_*|$. For $n \in \mathbb{N}$, clearly, the word $\widehat{w}_*^n \in \widehat{S}_{n\ell}$, and $(\log |\Phi(\widehat{w}_*^n)|)/(n\ell) = \nu^*$. It follows that $\nu^* \leq (\log \max\{|\Phi(\widehat{w})| : \widehat{w} \in \widehat{S}_{n\ell}\})/(n\ell)$. Taking the limit as $n \rightarrow \infty$, we obtain $\nu^* \leq \text{cap}_{\text{ind}}(S)$. To complete the proof we show that $\text{cap}_{\text{ind}}(S) \leq \nu^*$. We first claim that if $\widehat{w} \in \widehat{S}^*$ is a word generated by a (possibly non-simple) cycle of $\widehat{\mathcal{G}}$ then

$$\nu^* \geq \frac{\log |\Phi(\widehat{w})|}{|\widehat{w}|}. \quad (5.3)$$

This is easily proved by induction on $|\widehat{w}|$. If $|\widehat{w}| = 1$, then the cycle generating \widehat{w} is simple and obviously (5.3) holds. For $|\widehat{w}| > 1$, let $\pi = (e_i)_{i=1}^\ell$ be a cycle of $\widehat{\mathcal{G}}$ generating \widehat{w} . Obviously (5.3) holds if π is simple. Otherwise, there exist integers $1 \leq j < k \leq \ell$ such that $\tau(e_j) = \tau(e_k)$. So both $\alpha = e_{j+1}, \dots, e_k$ and $\beta = e_1, \dots, e_j, e_{k+1}, \dots, e_\ell$ are cycles in $\widehat{\mathcal{G}}$. Let \widehat{x}, \widehat{y} denote the words generated by α and β respectively. Then using the induction hypothesis on $|\widehat{x}|$ and $|\widehat{y}|$, we get

$$\begin{aligned} \frac{\log |\Phi(\widehat{w})|}{|\widehat{w}|} &= \frac{|\widehat{x}| \log |\Phi(\widehat{x})|}{|\widehat{w}| |\widehat{x}|} + \frac{|\widehat{y}| \log |\Phi(\widehat{y})|}{|\widehat{w}| |\widehat{y}|} \\ &\leq \frac{|\widehat{x}|}{|\widehat{w}|} \nu^* + \frac{|\widehat{y}|}{|\widehat{w}|} \nu^* \\ &= \nu^*. \end{aligned}$$

Now, for $n \in \mathbb{N}$, let $\widehat{z}(n) \in \widehat{S}_n$ be a word such that $|\Phi(\widehat{z}(n))| = \max\{|\Phi(\widehat{w})| : \widehat{w} \in \widehat{S}_n\}$. Let $(e_i^{(n)})_{i=0}^{n-1}$ be a path in $\widehat{\mathcal{G}}$ generating $\widehat{z}(n)$. Then using [37, Lemma 13] (and removing “0-length” cycles), it may be decomposed as follows. There exist an integer $0 \leq m \leq |V|$ and $2m$ integers $0 \leq s_1 \leq t_1 < s_2 \leq t_2 < \dots < s_m \leq t_m < n$

5.3. Independence capacity and axial products

such that for each $k = 1, \dots, m$, $(e_i^{(n)})_{i=s_k}^{t_k}$ is a cycle, and $n - \sum_k (t_k - s_k + 1) \leq |V|$. Set $X = \bigcup_{k=1}^m \{s_k, \dots, t_k\}$. Using (5.3), we have

$$\begin{aligned} \frac{\log |\Phi(\widehat{z}(n))|}{n} &= \sum_{k=1}^m \frac{\log |\Phi(\mathcal{L}(e_{s_k}^{(n)}) \dots \mathcal{L}(e_{t_k}^{(n)}))|}{n} + \sum_{i \in [n] \setminus X} \frac{\log |\Phi(\mathcal{L}(e_i^{(n)}))|}{n} \\ &\leq \sum_{k=1}^m \left(\frac{\log |\Phi(\mathcal{L}(e_{s_k}^{(n)}) \dots \mathcal{L}(e_{t_k}^{(n)}))|}{t_k - s_k + 1} \cdot \frac{t_k - s_k + 1}{n} \right) + \frac{|V|}{n} \log |\Sigma| \\ &\leq \nu^* + \frac{|V|}{n} \log |\Sigma|. \end{aligned}$$

The result follows by taking the limit as $n \rightarrow \infty$ of both sides of the last inequality. \blacksquare

We next show that the independence capacity of a constraint cannot exceed its capacity.

Theorem 8. *Let S be a \mathfrak{D} -dimensional constraint over Σ then $\text{cap}_{\text{ind}}(S) \leq \text{cap}(S)$.*

Proof. For $\mathbf{m} \in \mathbb{N}^{\mathfrak{D}}$, let $\widehat{z}(\mathbf{m}) \in \widehat{S}_{\mathbf{m}}$ be a configuration such that $|\Phi(\widehat{z}(\mathbf{m}))| = \max\{|\Phi(\widehat{w})| : \widehat{w} \in \widehat{S}_{\mathbf{m}}\}$. Since, $\Phi(\widehat{z}(\mathbf{m})) \subseteq S_{\mathbf{m}}$, we get $(1/|\mathbf{m}|) \log |\Phi(\widehat{z}(\mathbf{m}))| \leq (1/|\mathbf{m}|) \log |S_{\mathbf{m}}|$. Taking the limit of both sides as $\mathbf{m} \rightarrow \infty$, we obtain the result. \blacksquare

5.3 Independence capacity and axial products

In this section we show a relation between the independence capacity of an axial product of 1-dimensional constraints to the independence capacities of the 1-dimensional constraints. A similar relation holds for the (conventional) capacities. The relation is stated in the next theorem.

Theorem 9. *Let $S^{(1)}, \dots, S^{(\mathfrak{D})}$ be 1-dimensional constraints over Σ . Then the following statements hold:*

1. $\text{cap}_{\text{ind}}(S^{(1)} \otimes \dots \otimes S^{(\mathfrak{D})}) \leq \min_i \text{cap}_{\text{ind}}(S^{(i)})$.
2. $\text{cap}(S^{(1)} \otimes \dots \otimes S^{(\mathfrak{D})}) \leq \min_i \text{cap}(S^{(i)})$.
3. *If $S^{(1)} = \dots = S^{(\mathfrak{D})} = S$ then $\text{cap}_{\text{ind}}(S^{\otimes \mathfrak{D}}) = \text{cap}_{\text{ind}}(S)$.*

5.3. Independence capacity and axial products

Remark 1. [37] proves Part 3 for binary 1-dimensional constraints. The proof uses the existence of a word $\widehat{w} \in \widehat{S}$ with $\text{cap}_{\text{ind}}(S) = \log(|\Phi(\widehat{w})|)/|\widehat{w}|$ and $\widehat{w}^n \in \widehat{S}$ for all $n \in \mathbb{N}$. The proof given here does not rely on the existence of such a word.

Proof. Part 1. Denote by \widehat{S} the multi-choice constraint corresponding to $S^{(1)} \otimes \dots \otimes S^{(\mathfrak{D})}$. Fix $i \in \{1, \dots, \mathfrak{D}\}$. For $k \in \mathbb{N}$, let $\mathbf{m}_k = (m_k^{(1)}, \dots, m_k^{(\mathfrak{D})}) \subseteq \mathbb{N}^{\mathfrak{D}}$, be the \mathfrak{D} -tuple with $m_k^{(j)} = 1$, for $j \in \{1, \dots, \mathfrak{D}\} \setminus \{i\}$, and $m_k^{(i)} = k$. Every array in $\widehat{S}_{\mathbf{m}_k}$ is essentially a word in $\widehat{S}^{(i)}_k$ and vice versa. Hence, $(\log \max\{|\Phi(\widehat{w})| : \widehat{w} \in \widehat{S}_{\mathbf{m}_k}\})/|\mathbf{m}_k| = (\log \max\{|\Phi(\widehat{w})| : \widehat{w} \in \widehat{S}^{(i)}_k\})/k$. By (5.1) it follows that $\text{cap}_{\text{ind}}(S^{(1)} \otimes \dots \otimes S^{(\mathfrak{D})}) \leq \text{cap}_{\text{ind}}(S^{(i)})$, and the theorem follows.

Part 2. The proof is similar to the proof of Part 1, so we omit it here.

Part 3. Let $T = S^{\otimes \mathfrak{D}}$. By Part 1, it's enough to show $\text{cap}_{\text{ind}}(T) \geq \text{cap}_{\text{ind}}(S)$. Let $f: \mathbb{N} \rightarrow \mathbb{N}$ be any function satisfying $\lim_{i \rightarrow \infty} (f(i)/i) = \infty$. For $i \in \mathbb{N}$, let $\mathbf{m}_i = (i, i, \dots, i, f(i)) \in \mathbb{N}^{\mathfrak{D}}$ be the \mathfrak{D} -tuple with every entry but the last equal to i , and the last entry equal to $f(i)$. Set $\ell(i) = (\mathfrak{D}-1)(i-1) + f(i)$, and let $\widehat{z}^{(i)} = \widehat{z}_0^{(i)} \dots \widehat{z}_{\ell(i)-1}^{(i)} \in \widehat{S}_{\ell(i)}$ be a word such that $|\Phi(\widehat{z}^{(i)})| = \max\{|\Phi(\widehat{w})| : \widehat{w} \in \widehat{S}_{\ell(i)}\}$. Define the \mathfrak{D} -dimensional array $\widehat{\Gamma}^{(i)} \in \widehat{\Sigma}^{\mathbf{m}_i}$ by

$$\widehat{\Gamma}_{\mathbf{j}}^{(i)} = \widehat{z}_{\psi(\mathbf{j})}^{(i)}, \quad \mathbf{j} \in [\mathbf{m}_i],$$

where $\psi: \mathbb{Z}^{\mathfrak{D}} \rightarrow \mathbb{Z}$ is given by $\psi(j_1, \dots, j_{\mathfrak{D}}) = \sum_k j_k$. Observe that every row of $\widehat{\Gamma}^{(i)}$ is a (contiguous) sub-word of $\widehat{z}^{(i)}$; it follows that $\widehat{\Gamma}^{(i)} \in \widehat{T}_{\mathbf{m}_i}$. Consequently,

$$|\Phi(\widehat{\Gamma}^{(i)})| \leq \max\{|\Phi(\widehat{\Gamma})| : \widehat{\Gamma} \in \widehat{T}_{\mathbf{m}_i}\}. \quad (5.4)$$

We next lower bound $\log |\Phi(\widehat{\Gamma}^{(i)})|$. Set $X = \{j \in \mathbb{Z} : (i-1)(\mathfrak{D}-1) \leq j < f(i)\}$, and for $Y \subseteq \mathbb{Z}$ denote by $\psi^{-1}(Y) = \{\mathbf{j} \in \mathbb{Z}^{\mathfrak{D}} : \psi(\mathbf{j}) \in Y\}$. Then

$$\begin{aligned} \log |\Phi(\widehat{\Gamma}^{(i)})| &= \sum_{\mathbf{j} \in [\mathbf{m}_i]} \log |\Phi(\widehat{\Gamma}_{\mathbf{j}}^{(i)})| \\ &= \sum_{k \in [\ell(i)]} \left(|\psi^{-1}(\{k\}) \cap [\mathbf{m}_i]| \cdot \log |\Phi(\widehat{z}_k^{(i)})| \right) \\ &\geq i^{\mathfrak{D}-1} \sum_{k \in X} \log |\Phi(\widehat{z}_k^{(i)})| \\ &= i^{\mathfrak{D}-1} \left(\sum_{k \in [\ell(i)]} \log |\Phi(\widehat{z}_k^{(i)})| - \sum_{k \in [\ell(i)] \setminus X} \log |\Phi(\widehat{z}_k^{(i)})| \right) \\ &\geq i^{\mathfrak{D}-1} \log |\Phi(\widehat{z}^{(i)})| - 2(\mathfrak{D}-1)(i-1)i^{\mathfrak{D}-1} \log |\Sigma|, \end{aligned}$$

5.4. Independence capacity and $\lim_{\mathfrak{D} \rightarrow \infty} \text{cap}(S^{\otimes \mathfrak{D}})$

where we used the fact that for $k \in X$, it holds that $\psi^{-1}(\{k\}) \cap [\mathbf{m}_i] = \{(j_1, \dots, j_{\mathfrak{D}-1}, k - \sum_t j_t) : (j_1, \dots, j_{\mathfrak{D}-1}) \in [i]^{\mathfrak{D}-1}\}$. Combining the last inequality with (5.4), we have

$$\begin{aligned} \frac{\log \max\{|\Phi(\widehat{\Gamma})| : \widehat{\Gamma} \in \widehat{T}_{\mathbf{m}_i}\}}{|\mathbf{m}_i|} &\geq \frac{i^{\mathfrak{D}-1} \log |\Phi(\widehat{z}^{(i)})| - 2(\mathfrak{D}-1)(i-1)i^{\mathfrak{D}-1} \log |\Sigma|}{|\mathbf{m}_i|} \\ &\geq \frac{\log |\Phi(\widehat{z}^{(i)})|}{\ell(i)} - \frac{2(\mathfrak{D}-1)(i-1) \log |\Sigma|}{f(i)}. \end{aligned}$$

Taking the limit of both sides as $i \rightarrow \infty$, we obtain $\text{cap}_{\text{ind}}(T) \geq \text{cap}_{\text{ind}}(S)$. ■

5.4 Independence capacity and $\lim_{\mathfrak{D} \rightarrow \infty} \text{cap}(S^{\otimes \mathfrak{D}})$

Combining the fact that $\text{cap}(S^{\otimes \mathfrak{D}})$ is non-increasing in \mathfrak{D} with Theorem 8, we have $\text{cap}(S^{\otimes 1}) \geq \text{cap}(S^{\otimes 2}) \geq \dots \geq \text{cap}_{\text{ind}}(S)$. We denote the limit $\lim_{\mathfrak{D} \rightarrow \infty} \text{cap}(S^{\otimes \mathfrak{D}})$ by $\text{cap}_{\infty}(S)$. Obviously, $\text{cap}_{\infty}(S) \geq \text{cap}_{\text{ind}}(S)$ and for all the constraints for which we know the value of cap_{∞} , it turns out to be equal to cap_{ind} . We list these constraints.

1. RLL(d, k) with $k \leq 2d$. For this family of constraints, cap_{ind} turns out to be 0 [38]. In [20], it is shown that $\text{cap}(\infty) = 0$ as well.
2. RLL(d, ∞). For this family of constraints, cap_{ind} turns out to be $1/(d+1)$ [38]. [36] shows that

$$\text{cap}(\text{RLL}(d, \infty)^{\otimes \mathfrak{D}}) = \frac{1}{d+1} + O\left(\frac{\log^2(\mathfrak{D}(d+1))}{\mathfrak{D}(d+1)}\right).$$

Additionally, for $d = 1$, [29] shows that, for sufficiently large \mathfrak{D} ,

$$\frac{1}{2} + \frac{1}{2} \cdot 2^{-2\mathfrak{D}} \leq \text{cap}(\text{RLL}(1, \infty)^{\otimes \mathfrak{D}}) \leq \frac{1}{2} + 2^{o(\mathfrak{D})} 2^{-2\mathfrak{D}}.$$

3. The 3-checkerboard constraint CHK is defined over the alphabet $\{0, 1, 2\}$, and consists of all words in which every 2 adjacent symbols are distinct. The independence capacity of this constraint turns out to be $1/2$, and [29] shows that $\text{cap}_{\infty}(\text{CHK}) = 1/2$ as well.

The equality of cap_{∞} and cap_{ind} was first noticed empirically by Chaichanavong and Poo in [37], where they ask if this is true for all 1-dimensional constraints. The following theorem gives a partial answer.

5.4. Independence capacity and $\lim_{\mathfrak{D} \rightarrow \infty} \text{cap}(S^{\otimes \mathfrak{D}})$

Theorem 10. *Let S be a 1-dimensional constraint having finite memory m over Σ with $\text{cap}_{\text{ind}}(S) = 0$. Then*

$$\text{cap}(S^{\otimes (\mathfrak{D}+1)}) \leq \frac{m}{m+1} \text{cap}(S^{\otimes \mathfrak{D}}).$$

In particular, $\text{cap}_{\infty}(S) = 0$.

To prove the theorem we need the following definitions on labeled graphs. For a labeled graph $\mathcal{G} = (V, E, \mathcal{L})$ and a subset $U \subseteq V$, we call the graph $(U, E_U, \mathcal{L}|_{E_U})$ where $E_U = \{e \in E : \sigma(e) \in U, \tau(e) \in U\}$ the *subgraph of \mathcal{G} induced by U* . For two vertices $u, v \in V$ we say that u is *reachable from v* if there is a path in \mathcal{G} that starts in u and ends in v . We write $u \xrightarrow{\mathcal{G}} v$ if u is reachable from v and v is reachable from u . If $u \xrightarrow{\mathcal{G}} v$ does not hold we write $u \not\xrightarrow{\mathcal{G}} v$. The relation $\xrightarrow{\mathcal{G}}$ is an equivalence relation on the vertices of \mathcal{G} and the equivalence classes are called the *irreducible components* of \mathcal{G} . For an irreducible component of \mathcal{G} we shall sometime also call the subgraph of \mathcal{G} that it induces an irreducible component. Obviously, a graph \mathcal{G} is irreducible iff it has only one irreducible component. A labeled graph \mathcal{G} has *memory m* , for some nonnegative integer m , if all paths of length m generating the same word terminate at the same vertex. Clearly, if \mathcal{G} has memory m , then it also has memory n for any $n > m$. The following proposition relates memory of a constraint to memory of a graph.

Proposition 7. *If S is a 1-dimensional constraint with finite memory m , then there exists a presentation \mathcal{G} of S with memory m .*

Proof. We construct the “follower-set graph of S ”, $\mathcal{G} = ((V, E), \mathcal{L})$, as follows. $V = \{\mathcal{F}_S(x) : x \in S\}$, and for a vertex $\mathcal{F}_S(x) \in V$ and symbol $a \in \mathcal{F}_S(x) \cap \Sigma$, define $\delta(\mathcal{F}_S(x), a) = \mathcal{F}_S(xa)$. It’s easy to verify that $\delta(\cdot, \cdot)$ is well-defined. The set of edges E is now given by $E = \{(u, a, \delta(u, a)) : u \in V, a \in u \cap \Sigma\}$, and for an edge $e = (u, a, v) \in E$, $\sigma(e) = u$, $\tau(e) = v$ and $\mathcal{L}(e) = a$. Clearly, \mathcal{G} is deterministic. The following fact is easily verified

Fact 2. *For every vertex $\mathcal{F}_S(x) \in V$, a word w is generated from $\mathcal{F}_S(x)$ in \mathcal{G} if and only if $w \in \mathcal{F}_S(x)$ in which case the path generating it terminates at $\mathcal{F}_S(xw)$.*

As $\mathcal{F}_S(\varepsilon) = S$ it follows from this fact that \mathcal{G} is a presentation of S . We show that \mathcal{G} has memory m . Let $w \in S$ with $|w| = m$ be generated by some path in \mathcal{G} starting from a vertex $\mathcal{F}_S(x)$, then by the above fact, the path terminates at $\mathcal{F}_S(xw)$. We claim that $\mathcal{F}_S(xw) = \mathcal{F}_S(w)$. If $x = \varepsilon$ this is obviously true, so assume $x \neq \varepsilon$. Clearly, $\mathcal{F}_S(xw) \subseteq \mathcal{F}_S(w)$. On the other hand, let $y \in \mathcal{F}_S(w)$. Note that by our assumption, $|xwy| > m$ and every contiguous sub-word, with length $m+1$, of xwy is a contiguous sub-word of xw or a contiguous sub-word of wy and

therefore satisfies S . Since S has memory m , this implies that xwy satisfies S as well, and hence $y \in \mathcal{F}_S(xw)$. Thus $\mathcal{F}_S(w) = \mathcal{F}_S(xw)$ as claimed. Therefore all paths generating w terminate at $\mathcal{F}_S(w)$ and it follows that \mathcal{G} has memory m . ■

Proof of Theorem 10. Let $\mathcal{G} = ((V, E), \mathcal{L})$ be a presentation of S with memory m , and for a word $x \in S$ with $|x| \geq m$, denote by $v(x)$ the terminal state of any path generating x in G . We will need the next two lemmas. The first shows that if $\text{cap}_{\text{ind}}(S) = 0$, knowledge of long enough prefixes and suffixes of a word in S is often sufficient to determine the middle of the word.

Lemma 4. *Let $x, y \in S$ be words of length m such that $v(x) \stackrel{\mathcal{G}}{\leftrightarrow} v(y)$. Then there is at most one word of the form xay , where $a \in \Sigma$, in S .*

Proof. Assume to the contrary that there are two such words $xay, xby \in S$ where $a, b \in \Sigma$ and $a \neq b$. Therefore there are two paths in \mathcal{G} , $(e_i)_{i=1}^{2m+1}, (f_i)_{i=1}^{2m+1} \subseteq E$ generating xay and xby respectively. Since the paths $(e_i)_{i=1}^m$ and $(f_i)_{i=1}^m$ both generate x , it follows that $\sigma(e_{m+1}) = \sigma(f_{m+1}) = v(x)$. Similarly, since both paths $(e_i)_{i=m+2}^{2m+1}$ and $(f_i)_{i=m+2}^{2m+1}$ generate y , it follows that $\tau(e_{2m+1}) = \tau(f_{2m+1}) = v(y)$. Therefore the paths $(e_i)_{i=m+1}^{2m+1}, (f_i)_{i=m+1}^{2m+1}$ both start at $v(x)$, both end at $v(y)$, and generate ay and by respectively. Since, by our assumption, $v(x)$ and $v(y)$ are in the same irreducible component, there is a path generating some word $z \in \Sigma^*$ starting at $v(y)$ and ending at $v(x)$. Concatenating this path to the end of $(e_i)_{i=m+1}^{2m+1}$ and $(f_i)_{i=m+1}^{2m+1}$ we obtain two cycles—both starting and ending at $v(x)$ —one generating ayz and the other generating byz . Consequently, for all $k \in \mathbb{N}$ and sequence $(c_i)_{i=1}^k \subseteq \{a, b\}$, there is a cycle in \mathcal{G} generating the word $c_1 y z c_2 y z \dots c_k y z$. Let $|z| = \ell$, $y = y_0 \dots y_{m-1}$ and $z = z_0 \dots z_{\ell-1}$, where $y_i, z_j \in \Sigma$ for $i \in [m], j \in [\ell]$, and denote by \hat{y}, \hat{z} the words over $\hat{\Sigma}$ given by $\hat{y} = \{y_1\} \dots \{y_{m-1}\}$ and $\hat{z} = \{z_1\} \dots \{z_{\ell-1}\}$. It follows that for all $k \in \mathbb{N}$, $(\{a, b\} \hat{y} \hat{z})^k \in \hat{S}$. But this easily implies that $\text{cap}_{\text{ind}}(S) \geq 1/(m+\ell+1)$, contradicting our assumption. ■

Let $C_{\mathcal{G}}$ denote the number of irreducible components of \mathcal{G} . The next lemma bounds the number of appearances of words of the form xay in a certain word of S , where $a \in \Sigma$, $|x| = |y| = m$ and $v(x) \stackrel{\mathcal{G}}{\not\leftrightarrow} v(y)$.

Lemma 5. *For $\ell \in \mathbb{N}$ and word $z \in S_{\ell(m+1)+m}$, let $y^{(0)}, \dots, y^{(\ell)} \in \Sigma^m$ and $a^{(0)}, \dots, a^{(\ell-1)} \in \Sigma$ be defined by*

$$z = y^{(0)} a^{(0)} y^{(1)} a^{(1)} \dots y^{(\ell-1)} a^{(\ell-1)} y^{(\ell)}.$$

Then

$$\left| \left\{ i \in [\ell] : v(y^{(i)}) \stackrel{\mathcal{G}}{\not\leftrightarrow} v(y^{(i+1)}) \right\} \right| < C_{\mathcal{G}}.$$

5.4. Independence capacity and $\lim_{\mathfrak{D} \rightarrow \infty} \text{cap}(S^{\otimes \mathfrak{D}})$

Proof. Assume to the contrary that there are $0 \leq i_1 < i_2 < \dots < i_{C_G} < \ell$ such that $v(y^{(i_j)}) \not\stackrel{\mathcal{G}}{\rightarrow} v(y^{(i_{j+1})})$, for every $j \in \{1, \dots, C_G\}$. Then the sequence $v(y^{(i_1)}), v(y^{(i_2)}), \dots, v(y^{(i_{C_G})}), v(y^{(i_{C_G}+1)})$ has C_G+1 vertices and therefore contains two belonging to the same irreducible component, say $v(y^{(s)})$ and $v(y^{(t)})$, for some integers $0 \leq s < t \leq \ell$, with $s \in \{i_1, \dots, i_{C_G}\}$. Since the word $y^{(s)} a^{(s)} y^{(s+1)} \in S$, it follows that $v(y^{(s+1)})$ is reachable from $v(y^{(s)})$. Similarly, since $y^{(s+1)} a^{(s+1)} y^{(s+2)} a^{(s+2)} \dots a^{(t-1)} y^{(t)} \in S$, it holds that $v(y^{(t)})$ is reachable from $v(y^{(s+1)})$. But since $v(y^{(t)})$ is in the same irreducible component as $v(y^{(s)})$, it follows that $v(y^{(s)})$ is reachable from $v(y^{(s+1)})$ as well. Thus $v(y^{(s)}) \stackrel{\mathcal{G}}{\leftrightarrow} v(y^{(s+1)})$ which contradicts $s \in \{i_1, \dots, i_{C_G}\}$. \blacksquare

We can now prove the theorem. This part of the proof is a generalization of [20, Lemma 3]. Figure 5.1 illustrates the proof. Let ℓ be a positive integer. Denote by $\boldsymbol{\ell} \in \mathbb{N}^{\mathfrak{D}}$ the \mathfrak{D} -tuple with every entry equal to ℓ , and let $\mathbf{m}_\ell \in \mathbb{N}^{\mathfrak{D}+1}$ be given by $\mathbf{m}_\ell = (\ell(m+1) + m, \ell, \ell, \dots, \ell)$. Set $T = S^{\otimes (\mathfrak{D}+1)}$. We will give an upper bound on $|T_{\mathbf{m}_\ell}|$. Let X be the set $[\ell(m+1) + m] \setminus \{i(m+1) + m : i \in [\ell]\}$ and \mathcal{Y} the Cartesian product $X \times [\ell]$. For an array $\Gamma \in T_{\mathbf{m}_\ell}$ let $\Gamma|_{\mathcal{Y}} : \mathcal{Y} \rightarrow \Sigma$ denote the mapping given by $\Gamma|_{\mathcal{Y}}(\mathbf{x}) = \Gamma_{\mathbf{x}}$. Let $B_{\mathcal{Y}} = \{\Gamma|_{\mathcal{Y}} : \Gamma \in T_{\mathbf{m}_\ell}\}$ denote the set of all such mappings. For a mapping $\Delta \in B_{\mathcal{Y}}$ we define the set $Z(\Delta) \subseteq T_{\mathbf{m}_\ell}$ by $Z(\Delta) = \{\Gamma \in T_{\mathbf{m}_\ell} : \Gamma|_{\mathcal{Y}} = \Delta\}$. Clearly,

$$\bigcup_{\Delta \in B_{\mathcal{Y}}} Z(\Delta) = T_{\mathbf{m}_\ell}. \quad (5.5)$$

Let $\Delta \in B_{\mathcal{Y}}$, and fix $\mathbf{j} \in [\ell]$. For $i \in [\ell+1]$, let $y^{(i, \mathbf{j}, \Delta)} \in \Sigma^m$ be the word given by

$$y^{(i, \mathbf{j}, \Delta)} = \Delta(i(m+1), \mathbf{j}) \Delta(i(m+1) + 1, \mathbf{j}) \dots \Delta(i(m+1) + m - 1, \mathbf{j})$$

and for $\Gamma \in Z(\Delta)$ let $w^{(\Gamma, \mathbf{j})} \in \Sigma^{\ell(m+1) + m}$ be the word given by

$$w^{(\Gamma, \mathbf{j})} = \Gamma_{(0, \mathbf{j})} \Gamma_{(1, \mathbf{j})} \dots \Gamma_{(\ell(m+1) + m - 1, \mathbf{j})}.$$

Note that for such Γ , since $\Gamma \in T_{\mathbf{m}_\ell}$, $w^{(\Gamma, \mathbf{j})} \in S$, and, since $\Gamma|_{\mathcal{Y}} = \Delta$, we may write

$$w^{(\Gamma, \mathbf{j})} = y^{(0, \mathbf{j}, \Delta)} a^{(0, \mathbf{j}, \Gamma)} y^{(1, \mathbf{j}, \Delta)} a^{(1, \mathbf{j}, \Gamma)} \dots y^{(\ell-1, \mathbf{j}, \Delta)} a^{(\ell-1, \mathbf{j}, \Gamma)} y^{(\ell, \mathbf{j}, \Delta)},$$

where $a^{(i, \mathbf{j}, \Gamma)} = \Gamma_{(i(m+1) + m, \mathbf{j})}$ for $i \in [\ell]$. Now, if $i \in [\ell]$ such that $v(y^{(i, \mathbf{j}, \Delta)}) \stackrel{\mathcal{G}}{\leftrightarrow} v(y^{(i+1, \mathbf{j}, \Delta)})$ then, by Lemma 4, all $\Gamma \in Z(\Delta)$ have the same $a^{(i, \mathbf{j}, \Gamma)}$. On the other hand, by Lemma 5, since $Z(\Delta) \neq \emptyset$, it holds that $|\{i \in [\ell] : v(y^{(i, \mathbf{j}, \Delta)}) \stackrel{\mathcal{G}}{\leftrightarrow} v(y^{(i+1, \mathbf{j}, \Delta)})\}| < C_G$. It follows that $|\{w^{(\Gamma, \mathbf{j})} : \Gamma \in Z(\Delta)\}| \leq |\Sigma|^{C_G}$, and consequently

$$|Z(\Delta)| \leq \prod_{\mathbf{j} \in [\ell]} |\{w^{(\Gamma, \mathbf{j})} : \Gamma \in Z(\Delta)\}| \leq |\Sigma|^{C_G \ell^{\mathfrak{D}}}. \quad (5.6)$$

5.4. Independence capacity and $\lim_{\mathfrak{D} \rightarrow \infty} \text{cap}(S^{\otimes \mathfrak{D}})$

Since for any $\Delta \in B_{\mathfrak{Y}}$, and $i \in X$, the \mathfrak{D} -dimensional array $\Lambda(\Delta, i) \in \Sigma^{\ell}$ with entries given by $(\Lambda(\Delta, i))_{\mathbf{j}} = \Delta(i, \mathbf{j})$ is clearly in $S^{\otimes \mathfrak{D}}$, it follows that $|B_{\mathfrak{Y}}| \leq |(S^{\otimes \mathfrak{D}})_{\ell}|^{|X|}$. Combining this with (5.6) and (5.5), we have

$$|T_{\mathbf{m}_{\ell}}| = \sum_{\Delta \in B_{\mathfrak{Y}}} |Z(\Delta)| \leq |\Sigma|^{C_{\mathcal{G}} \ell^{\mathfrak{D}}} |(S^{\otimes \mathfrak{D}})_{\ell}|^{(\ell+1)m}.$$

Taking the logarithm of both sides and dividing by $|\mathbf{m}_{\ell}|$, we obtain

$$\frac{\log |T_{\mathbf{m}_{\ell}}|}{|\mathbf{m}_{\ell}|} \leq \frac{(\ell+1)m}{\ell(m+1)+m} \frac{\log |(S^{\otimes \mathfrak{D}})_{\ell}|}{\ell^{\mathfrak{D}}} + \frac{C_{\mathcal{G}} \log |\Sigma|}{\ell(m+1)+m}$$

The theorem follows by taking the limit as $\ell \rightarrow \infty$. ■

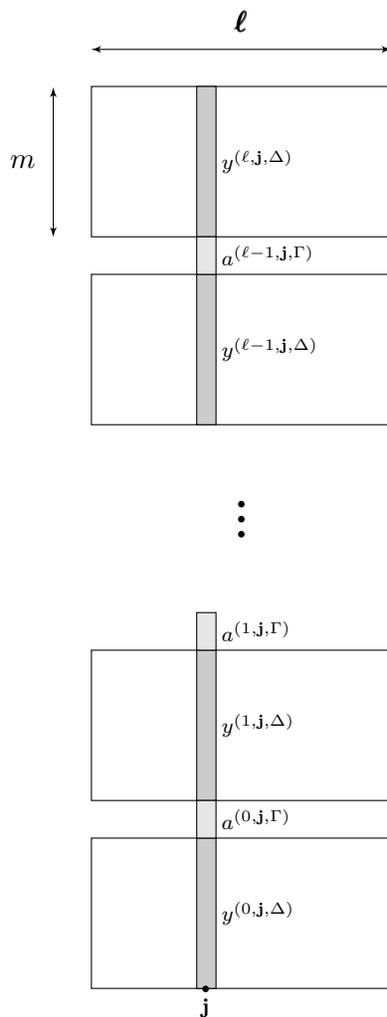


Figure 5.1: Proof of Theorem 10.

As an application of Theorem 10, consider the the family of *multiple-spaced runlength constraints*. Each of these constraints is denoted $\text{RLL}(d, k, s)$ with d, k and s nonnegative integers and $d \leq k$. It is defined over the alphabet $\{0, 1\}$ and consists of all words in $\text{RLL}(d, k)$ for which the length of each run of ‘0’s delimited by ‘1’s on both ends is a multiple of s . Fix such integers d, k , and s with $s \geq 2$, and let $S = \text{RLL}(d, k, s)$. It can be verified that each word of \widehat{S} has at most two

5.5. Open questions

letters equal to $\{0, 1\}$, and it follows that $\text{cap}_{\text{ind}}(S) = 0$. As the memory of S is (at most) k , by Theorem 10, we have the following corollary.

Corollary 3. *Let d, k, s be nonnegative integers such that $d \leq k$ and $s \geq 2$. Then*

$$\text{cap}(\text{RLL}(d, k, s)^{\otimes \mathfrak{D}}) = \left(\frac{k}{k+1} \right)^{\mathfrak{D}-1} \cdot \text{cap}(\text{RLL}(d, k, s)).$$

5.5 Open questions

Is it true that $\text{cap}_{\infty}(S) = \text{cap}_{\text{ind}}(S)$ for every 1-dimensional constraint S ? For a 1-dimensional constraint S , what can be said about the rate of convergence of $\text{cap}(S^{\otimes \mathfrak{D}})$ to $\text{cap}_{\infty}(S)$? Finally, is \widehat{S} a \mathfrak{D} -dimensional constraint, for every \mathfrak{D} -dimensional constraint S , when $\mathfrak{D} \geq 2$?

Chapter 6

The tradeoff function for binary 1-dimensional constraints*

This chapter deals with the tradeoff function for 1-dimensional binary constraints. As mentioned in Chapter 1, the tradeoff function for a 1-dimensional constraint evaluated at 0 equals the constraint's capacity and thus is a more general notion than capacity. The motivation for defining this function comes from the application of 1-dimensional constraints in digital storage systems. We describe the motivation in more detail in the next section. Later on, we give the precise definition of the tradeoff function for a general 1-dimensional binary constraint. The rest of the chapter shows our computation of this function for two families of RLL(d, k) constraints.

6.1 A brief overview of digital recording

In digital storage systems, user data is written to the device in the form of a binary sequence. Typically, not every binary sequence may be written reliably to the device and therefore, only a subset of all possible sequences is “allowed” to be written. The set of all the “allowed” sequences is usually modeled as a 1-dimensional binary constraint. In practice, limiting the written sequence to this “allowed” set is often not sufficient to guarantee the required reliability, and an error-correcting-code or ECC is used as well. Consequently, user data (represented as an arbitrary stream of ‘0’s and ‘1’s) is encoded twice before written to media. First the data is encoded to a codeword of an error-correcting-code or ECC and then the resultant codeword is encoded to an “allowed sequence” of some 1-dimensional binary constraint. In this context, the constraint is sometimes called a “modulation code” and the encoding of the ECC codeword to a constrained sequence is known as “modulation encoding”. When reading back the data the process is reversed: the data read from the device is first decoded by the modulation code decoder and then the ECC decoder is used to recover the source data, attempting to correct any errors

*A version of this chapter has been submitted for publication. Loudior, E. (2010) The Tradeoff Function for a Class of RLL(d, k) Constraints.

that may have occurred when the data was read. Roughly speaking, the rate of a modulation encoder is the ratio between the length of its input to the length of its output; it is typically strictly smaller than 1. In storage systems it is desirable that the rate be as high as possible to maximize storage.

This scheme suffers from a couple of disadvantages. First, a small number of errors that are present after reading the data from the device may turn into a burst of errors at the output of the modulation decoder, which may overwhelm the ECC decoder. Second, since the modulation code decoder is typically a “hard” decoder—meaning that it outputs “hard” ‘0’s or ‘1’s rather than probabilities or likelihoods—any soft or probabilistic information that might have been available after reading the data from the device is not readily available to the ECC decoder, thereby limiting its error correction capability.

In [40], [5], [38] and the references therein, several encoding schemes are proposed to overcome these disadvantages. Here, we focus on one of these schemes, in which the order of the two encodings mentioned above is reversed. The source data is first encoded with a modulation code into a constrained sequence, but instead of using the original constraint, we encode it to a sequence of the multi-choice constraint corresponding to the original constraint, where we use 0, 1, and \square in place of the symbols $\{0\}$, $\{1\}$, and $\{0, 1\}$ of $\widehat{\Sigma}$, respectively. The entries containing the ‘ \square ’s are “unconstrained” in the sense that replacing (or “filling”) them with any values in $\{0, 1\}$ independently would result in a sequence satisfying the constraint. Next, a systematic ECC with a suitable redundancy is applied, placing the redundancy (parity-check) bits in these unconstrained positions. Clearly, this addresses both of the disadvantages listed above.

In this scheme, since the error correction capability of the ECC depends on the number of redundancy bits, it is desirable that the number of unconstrained positions be as large as possible. On the other hand, increasing the number of unconstrained positions at the output of the modulation encoder naturally reduces its rate, as no user information is encoded in the unconstrained positions. In [38] the authors study the *tradeoff function* that defines for a given “density” of unconstrained positions, called the *insertion-rate*, the maximum overall rate of the encoding; knowing this function is obviously important to the design of efficient digital storage systems employing this scheme. Currently, there are only very few constraints for which the tradeoff function has been computed explicitly.

As mentioned in Chapter 1, the $\text{RLL}(d, k)$ constraint is widely used in digital storage systems employing optical or magnetic recording. Another constraint used in practice is the *maximum transition run* or MTR constraint, denoted $\text{MTR}(j, k)$ for some nonnegative integers j, k . This constraint consists of all binary sequences in which the length of each run of ‘1’s is at most j and the length of each run of ‘0’s is at most k . More details on these constraints as well as other constraints used in

practice may be found in [18] and [32]. We give a precise definition of the tradeoff function in Section 6.3.

6.2 Previous work

In [5], the tradeoff functions for $\text{RLL}(0, 1)$ and $\text{RLL}(0, 2)$ are determined. In [37], Poo computed the tradeoff function for $\text{RLL}(0, 3)$ for insertion rates between 0 and $1/4$, and the tradeoff function for $\text{RLL}(d, 2d + 1)$ for any d . In [38] the authors compute the tradeoff function for $\text{MTR}(2, 2)$. For completeness, we present these functions in Theorem 11. Lower bounds on the tradeoff function for $\text{RLL}(0, k)$ are given in [19] and [21]. In this chapter, we determine the tradeoff functions for two other families of constraints: $\text{RLL}(d, 2d+2)$, and $\text{RLL}(d, \infty)$. Our results are stated precisely in Theorems 12 and 13. For $\text{RLL}(d, 2d+2)$, we find a curious dichotomy in the shape of the tradeoff function between different ranges of values of d . The function is always piecewise linear; yet it consists of 2 linear “segments” for $1 \leq d \leq 16$ and 3 segments for $d \geq 17$.

This chapter is organized as follows. In Section 6.3 we define the tradeoff function and related concepts as well as summarize some of its known properties. We also state the previously known tradeoff functions and our new results. In Section 6.4 we show the derivation of the tradeoff function for $\text{RLL}(d, \infty)$ and in Section 6.5 we show the derivation of the tradeoff function for $\text{RLL}(d, 2d+2)$.

6.3 Background and definitions

For the rest of this chapter, fix $\Sigma = \{0, 1\}$. Let S be a 1-dimensional constraint over Σ and $\widehat{\Sigma}$, Φ , \widehat{S} and $\widehat{\mathcal{G}}_{\mathcal{F}_S}$ be as defined in Chapter 5. As already stated, in this chapter we use ‘0’, ‘1’ and ‘ \square ’ in place of $\{0\}$, $\{1\}$ and $\{0, 1\}$, respectively. So $\widehat{\Sigma} = \{0, 1, \square\}$, and $\Phi(x)$, for $x \in \widehat{\Sigma}^*$, can be thought of as the set of all possible “fillings” of the ‘ \square ’s of x with bits. We also sometimes omit the subscript \mathcal{F}_S from $\widehat{\mathcal{G}}_{\mathcal{F}_S}$ to simplify notation. For a word $w \in \widehat{\Sigma}^*$, let $\#\square(w)$ denote the number of ‘ \square ’s in w . Observe that $\log |\Phi(w)| = \#\square(w)$ and hence

$$\text{cap}_{\text{ind}}(S) = \lim_{m \rightarrow \infty} \frac{\max\{\#\square(w) : w \in \widehat{S}_m\}}{m},$$

so $\text{cap}_{\text{ind}}(S)$ is the asymptotic maximum density of ‘ \square ’s in words of \widehat{S} . In this chapter we use the following notation for sequences. For a set T and nonnegative integer n we denote a sequence b_1, \dots, b_n of n elements of T by (b_i) , and index its elements by $\{1, 2, \dots, n\}$. We refer to n as the *length* of the sequence

6.3. Background and definitions

and denote it by $|(b_i)|$. We abuse notation and write $(b_i) \subseteq T$ to mean that $b_i \in T$ for all $i \in \{1, 2, \dots, |(b_i)|\}$. For two sequences $(b_i), (d_i) \subseteq T$ we denote by $(b_i)(d_i)$ the sequence formed by concatenating the sequences (b_i) and (d_i) , that is the sequence $b_1, b_2, \dots, b_{|(b_i)|}, d_1, d_2, \dots, d_{|(d_i)|}$. For a nonnegative integer m , the notation $(b_i)^m$ is used to denote the sequence formed by concatenating (b_i) to itself m times (as usual $(b_i)^0$ is the empty sequence). We shall also consider infinite sequences b_1, b_2, \dots with elements in T and denote such sequences by $(b_i)_{i=1}^\infty$. For a sequence (M_i) of m real nonnegative square matrices all having the same size, we write $\lambda((M_i))$ to mean $\lambda(\prod_i M_i)$.

Let S be a constraint over Σ . For a positive integer n and a subset $I \subseteq [n]$ we define $\mathcal{M}(I, n) = \mathcal{M}_S(I, n) = |\{w_0 w_1 \dots w_{n-1} \in \widehat{S}_n : \forall i \in [n], w_i = \square \Leftrightarrow i \in I\}|$. Also, for any real number $\rho \in [0, 1]$, define the set \mathcal{I}_ρ by

$$\mathcal{I}_\rho = \left\{ (I_j)_{j=1}^\infty : I_j \subseteq [j] \text{ for all } j, \text{ and } \lim_{j \rightarrow \infty} \frac{|I_j|}{j} = \rho \right\}.$$

Then the *tradeoff function* of S , $f_S : [0, 1] \rightarrow [0, 1] \cup \{-\infty\}$ is given by

$$f_S(\rho) = \sup_{(I_j) \in \mathcal{I}_\rho} \limsup_{j \rightarrow \infty} \frac{\log \mathcal{M}_S(I_j, j)}{j}.$$

A 1-dimensional constraint is *irreducible* if it has an irreducible presentation. A graph (labeled graph) is called *trivial* if it has exactly one vertex and no edges. Every RLL(d, k) constraint is irreducible. In [38] it is shown that if S is an irreducible finite-type constraint (that has infinitely many words), then $\widehat{\mathcal{G}}_{\mathcal{F}_S}$ has exactly one non-trivial irreducible component. Here, we denote this component by $\widehat{\mathcal{G}}_\star = (\widehat{V}_\star, \widehat{E}_\star, \widehat{\mathcal{L}}_\star)$. For a subset $Q \subseteq \widehat{\Sigma}$, let \widehat{E}_\star^Q denote the subset of \widehat{E}_\star consisting of the edges whose label is in Q . We denote by $\widehat{\mathcal{G}}_\star^{\{0,1\}}$ the subgraph of $\widehat{\mathcal{G}}_\star$ given by $\widehat{\mathcal{G}}_\star^{\{0,1\}} = (\widehat{V}_\star, \widehat{E}_\star^{\{0,1\}}, \widehat{\mathcal{L}}_\star|_{\widehat{E}_\star^{\{0,1\}}})$ and by $\widehat{\mathcal{G}}_\star^{\{\square\}}$ the subgraph of $\widehat{\mathcal{G}}_\star$ given by $\widehat{\mathcal{G}}_\star^{\{\square\}} = (\widehat{V}_\star, \widehat{E}_\star^{\{\square\}}, \widehat{\mathcal{L}}_\star|_{\widehat{E}_\star^{\{\square\}}})$. We define $A^{\{0,1\}}(S) = A(\widehat{\mathcal{G}}_\star^{\{0,1\}})$ and $A^{\{\square\}}(S) = A(\widehat{\mathcal{G}}_\star^{\{\square\}})$. Let $(M_i) \subseteq \{A^{\{0,1\}}(S), A^{\{\square\}}(S)\}$ be a sequence of length n . We say that a path $(e_i) \subseteq \widehat{E}_\star$ of $\widehat{\mathcal{G}}_\star$ *matches* (M_i) if it has length n , and for every $1 \leq i \leq n$, $\widehat{\mathcal{L}}_\star(e_i) = \square$ iff $M_i = A^{\{\square\}}(S)$. Note that for any $s, t \in \widehat{V}_\star$, the entry $(\prod_i M_i)_{(s,t)}$ is the number of paths in $\widehat{\mathcal{G}}_\star$ starting at s , ending at t and matching (M_i) . For a finite sequence $(M_i) \subseteq \{A^{\{0,1\}}(S), A^{\{\square\}}(S)\}$, we denote by $\varrho((M_i))$ the density of $A^{\{\square\}}(S)$ in (M_i) , namely

$$\varrho((M_i)) = \frac{|\{1 \leq i \leq |(M_i)| : M_i = A^{\{\square\}}(S)\}|}{|(M_i)|}.$$

Let S be a 1-dimensional constraint over Σ . We list the following known facts about the tradeoff function f_S .

6.3. Background and definitions

- For $0 \leq \rho \leq \text{cap}_{\text{ind}}(S)$, $f_S(\rho) \geq 0$, and for $\text{cap}_{\text{ind}}(S) < \rho \leq 1$, $f_S(\rho) = -\infty$.
- $f_S(0) = \text{cap}(S)$.
- f_S is decreasing in $[0, \text{cap}_{\text{ind}}(S)]$. Moreover, for any $0 \leq \rho_1 < \rho_2 \leq \text{cap}_{\text{ind}}(S)$,

$$f_S(\rho_1) - f_S(\rho_2) \geq \rho_2 - \rho_1.$$

- f_S is left-continuous in $[0, \text{cap}_{\text{ind}}(S)]$.
- If S is irreducible and finite-type then f_S is concave and continuous in $[0, \text{cap}_{\text{ind}}(S)]$. Furthermore, for all rational $\rho \in [0, 1]$

$$f_S(\rho) = \sup_{(M_i)} \frac{\log \lambda((M_i))}{|(M_i)|}, \quad (6.1)$$

where the sup is taken over all sequences $(M_i) \subseteq \{A^{\{0,1\}}(S), A^{\{\square\}}(S)\}$ with $\varrho((M_i)) = \rho$.

- For integers $0 \leq d \leq k$

$$\text{cap}_{\text{ind}}(\text{RLL}(d, k)) = \frac{\lfloor (k-d)/(d+1) \rfloor}{\lfloor (k+1)/(d+1) \rfloor (d+1)}, \quad (6.2)$$

$$\text{and} \quad \text{cap}_{\text{ind}}(\text{RLL}(d, \infty)) = \frac{1}{d+1}. \quad (6.3)$$

See [38] for proofs.

As mentioned in the introduction, there are a few constraints for which the tradeoff function has been computed explicitly. These are summarized in the next theorem, along with the references to the respective papers. For a finite sequence $(\mathbf{x}_i) \subseteq \mathbb{R}^2$ of points such that $\mathbf{x}_i = (x_i, y_i)$ and $x_1 < x_2 < \dots < x_n$, we define $\mathbb{L}_{(\mathbf{x}_i)} : [x_1, x_n] \rightarrow \mathbb{R}$ to be the function whose graph is the piecewise linear curve connecting these points in sequence; namely, the function that satisfies

$$\mathbb{L}_{(\mathbf{x}_i)}(x) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} (x - x_i) + y_i, \quad x_i \leq x \leq x_{i+1},$$

for all $1 \leq i < n$.

Theorem 11. *Let $S^{(1)} = \text{RLL}(0, 1)$, $S^{(2)} = \text{RLL}(0, 2)$, $S^{(3)} = \text{RLL}(0, 3)$, $S^{(4)} = \text{RLL}(d, 2d+1)$, and $S^{(5)} = \text{MTR}(2, 2)$. Let f_i be the function $f_{S^{(i)}}$ restricted to $[0, \text{cap}_{\text{ind}}(S^{(i)})]$. Then the following statements hold:*

1. $f_1 = \mathbb{L}_{(0, \text{cap}(S^{(1)}))}(\frac{1}{2}, 0)$ (shown in [5]).

6.3. Background and definitions

2. $f_2 = \mathbb{L}_{(0, \text{cap}(S^{(2)})), (\frac{1}{3}, \frac{2}{3} \text{cap}(S^{(1)})), (\frac{2}{3}, 0)}$ (shown in [5]).
3. $f_3(\rho) = \mathbb{L}_{(0, \text{cap}(S^{(3)})), (\frac{1}{4}, \frac{3}{4} \text{cap}(S^{(2)}))}(\rho)$, for $0 \leq \rho \leq \frac{1}{4}$ (shown in [37]).
4. $f_4 = \mathbb{L}_{(0, \text{cap}(S^{(4)})), (\frac{1}{2(d+1)}, 0)}$ (shown in [37]).
5. $f_5 = \mathbb{L}_{(0, \text{cap}(S^{(5)})), (\frac{1}{3}, 0)}$ (shown in [38]).

We now state our new results.

Theorem 12. *Let d be a nonnegative integer and $S = \text{RLL}(d, \infty)$. Set*

$$\mathbf{p}_1 = (0, \text{cap}(S)), \quad \mathbf{p}_2 = \left(\frac{1}{d+1}, 0 \right) = (\text{cap}_{\text{ind}}(S), 0).$$

Then

$$f_S(\rho) = \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}(\rho), \text{ for } \rho \in [0, \text{cap}_{\text{ind}}(S)].$$

Theorem 13. *Let d be a positive integer and $S = \text{RLL}(d, 2d+2)$. Set*

$$\begin{aligned} \mathbf{p}_1 &= (0, \text{cap}(S)), & \mathbf{p}_2 &= \left(\frac{3}{6d+8}, \frac{\log 3}{6d+8} \right), \\ \mathbf{p}_3 &= \left(\frac{2}{4d+5}, \frac{1}{4d+5} \right), & \mathbf{p}_4 &= \left(\frac{2}{4d+4}, 0 \right) = (\text{cap}_{\text{ind}}(S), 0). \end{aligned}$$

Then the following statements hold:

1. If $1 \leq d \leq 16$ then

$$f_S(\rho) = \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_4}(\rho), \text{ for } \rho \in [0, \text{cap}_{\text{ind}}(S)].$$

2. If $17 \leq d$ then

$$f_S(\rho) = \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4}(\rho), \text{ for } \rho \in [0, \text{cap}_{\text{ind}}(S)].$$

The proofs are given in the next sections. The graphs of the tradeoff functions for $\text{RLL}(d, \infty)$ and $\text{RLL}(d, 2d+2)$ for $\rho \in [0, \text{cap}_{\text{ind}}(S)]$ are sketched in Figure 6.1. We need some properties of nonnegative matrices which we summarize here. If (M'_i) is formed by cyclically shifting (M_i) (that is, there exists an integer o such that for all i , $M'_i = M_{((i-1+o) \bmod m)+1}$, where for an integer j , $j \bmod m$ denotes the unique integer $k \in [m]$ such that $k \equiv j \pmod{m}$) then $\prod_i M_i$ and $\prod_i M'_i$ have the same characteristic polynomial ([33, 2.15.15]); in particular, $\lambda((M'_i)) = \lambda((M_i))$. If M and N are two nonnegative square matrices with $M \leq N$ then $\lambda(M) \leq \lambda(N)$

6.3. Background and definitions

([33, 5.7.5]). The *support graph* of an $m \times m$ nonnegative matrix M denoted $G_M = (V_M, E_M)$ is the directed graph with vertices $V_M = \{1, 2, \dots, m\}$ and edges $E_M = \{(i, j) \in V_M \times V_M : M_{i,j} > 0\}$, where for an edge $e = (i, j) \in E_M$, $\sigma(e) = i$ and $\tau(e) = j$. Such a matrix is called primitive if its support graph is primitive. For a primitive matrix M , the limit $\lim_{g \rightarrow \infty} (M^g) / (\lambda(M)^g)$ (where the limit is taken entry-wise), exists and is strictly positive in each entry ([33, 5.9.7]).

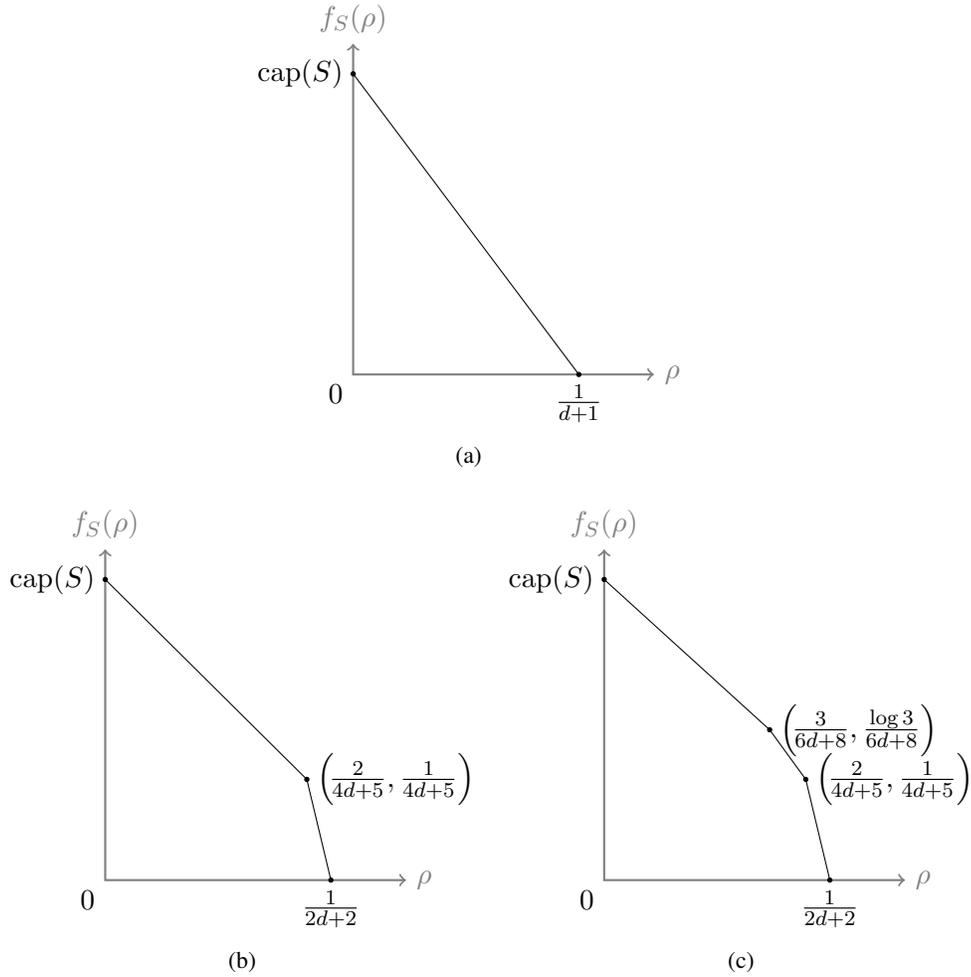


Figure 6.1: The graphs of $f_S(\rho)$ for $\rho \in [0, \text{cap}_{\text{ind}}(S)]$ (not to scale): (a) $S = \text{RLL}(d, \infty)$; (b) $S = \text{RLL}(d, 2d+2)$ and $1 \leq d \leq 16$; (c) $S = \text{RLL}(d, 2d+2)$ and $17 \leq d$.

6.4 Proof of Theorem 12

The proof is similar to the proof of [37, Proposition 45]. Set $\mathcal{C}_d = \text{cap}(S)$, and let $\widehat{\mathcal{G}} = \widehat{\mathcal{G}}_{\mathcal{F}_S} = (\widehat{V}, \widehat{E}, \widehat{\mathcal{L}})$ be the presentation of \widehat{S} defined in Section 6.3. Note that S is irreducible and of finite-type. We denote by a_i the follower set $\mathcal{F}_S(10^i)$ for $0 \leq i \leq d$. It can be verified that $\widehat{V} = \{a_0, a_1, \dots, a_d\}$. The graph $\widehat{\mathcal{G}}$ is given in Figure 6.2.

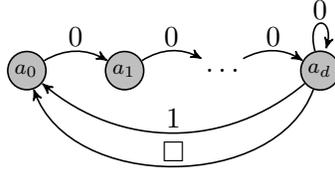


Figure 6.2: The graph $\widehat{\mathcal{G}}_{\mathcal{F}_{\text{RLL}}(d, \infty)}$.

Clearly, $\widehat{\mathcal{G}}_\star = \widehat{\mathcal{G}}_{\mathcal{F}_S}$ in this case. Let $A = A^{\{0,1\}}(S)$, $B = A^{\{\square\}}(S)$, and set $C = BA^d$ and $h_d = \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}$. For a sequence $(N_i) \subseteq \{A, C\}$ we denote by $\varepsilon((N_i))$ the “expanded” sequence, with elements in $\{A, B\}$, formed by substituting the sequence $(B)(A)^d$ for every element C in (N_i) .

It follows from (6.3) and the rest of the discussion in Section 6.3 that $f_S(0) = h_d(0)$ and $f_S(\rho) \geq h_d(\rho)$ for $\rho \in (0, \text{cap}_{\text{ind}}(S)]$. Hence it’s enough to show $f_S(\rho) \leq h_d(\rho)$ for $\rho \in (0, \text{cap}_{\text{ind}}(S)]$. Since both f_S and h_d are continuous in that interval, it’s enough to show the latter inequality for all rational $\rho \in (0, \text{cap}_{\text{ind}}(S)]$. Let ρ be such a rational. By (6.1) it suffices to show that for any sequence $(M_i) \subseteq \{A, B\}$ with $\varrho((M_i)) = \rho$, we have

$$\frac{\log \lambda((M_i))}{|(M_i)|} \leq h_d(\varrho((M_i))). \quad (6.4)$$

Let (M_i) be such a sequence. Set $m = |(M_i)|$ and consider the sequence $(X_i) = (M_i)^2$. Note that in any path of $\widehat{\mathcal{G}}$ the number of edges between a pair of edges labelled with a ‘ \square ’ must be at least d . It follows that if there exist integers $1 \leq i < j \leq 2m$, with $j - i - 1 < d$, such that $X_i = X_j = B$, then no paths of $\widehat{\mathcal{G}}$ match (X_i) or equivalently $(\prod_{i=1}^m M_i)^2 = 0$. The latter equality implies $\lambda((M_i)) = 0$, and therefore (6.4) holds. So assume no such integers exist. It can be verified that in this case we may cyclically shift (M_i) such that $(M_i) = \varepsilon((N_i))$ for some $(N_i) \subseteq \{A, C\}$ with $N_1 = C$; such a cyclic shift does not change either side of (6.4). We denote $|(N_i)|$

6.4. Proof of Theorem 12

by n . Now, it's easy to verify that there is exactly one path matching $(B)(A)^d$ and it starts and ends at a_d ; hence the only nonzero entry of C is $(C)_{(a_d, a_d)}$ and it is equal to 1. It follows that

$$C^k = C, \quad \text{for } k = 1, 2, \dots \quad (6.5)$$

$$CQC \leq C^2Q, \quad \text{for any nonnegative } |\widehat{V}| \times |\widehat{V}| \text{ matrix } Q. \quad (6.6)$$

Let s be the number of elements in (N_i) equal to C ; clearly, $s = m\rho > 0$. Let $(N'_i) = (C)^s(A)^{n-s}$; then by (6.5) and (6.6), we have $\lambda((M_i)) = \lambda((N_i)) \leq \lambda((N'_i)) = \lambda(CA^{n-s})$. We order the entries of (CA^{n-s}) as follows. For every $i, j \in [d]$ we place the element $(CA^{n-s})_{(a_i, a_j)}$ in the i th row and j th column. Observe, that using this ordering, the matrix CA^{n-s} is lower triangular with $(CA^{n-s})_{(i, i)} = 0$ for $0 \leq i < d$. It follows that $\lambda(CA^{n-s}) = (CA^{n-s})_{(a_d, a_d)} = (A^{n-s})_{(a_d, a_d)}$. Let Γ be the set of all paths in $\widehat{\mathcal{G}}$ starting and terminating in a_d and matching $(A)^{n-s}$, and for a nonnegative integer g define the set $\Delta(g) \subseteq S$ by

$$\Delta(g) = \{w \in S_g : w \text{ does not end with } 10^i, 0 \leq i \leq d-1\}.$$

Then it's not hard to check that $\Delta(n-s)$ is precisely the set of words generated by paths in Γ . Since $\widehat{\mathcal{G}}$ is deterministic, we have that $|\Gamma| = |\Delta(n-s)|$, so $\lambda((N'_i)) = |\Delta(n-s)|$. Now, observe that for $g \geq d$, any word in S_{g-d} can be extended to a word in $\Delta(g)$ by adding '0's; thus for all $g \geq d$, $|S_{g-d}| \leq |\Delta(g)| \leq |S_g|$. It follows that

$$\lim_{g \rightarrow \infty} \frac{\log |\Delta(g)|}{g} = C_d.$$

On the other hand, note that for any nonnegative integers g_1, g_2 and words $w \in \Delta(g_1), x \in \Delta(g_2)$, the word $wx \in \Delta(g_1 + g_2)$; it follows that $\log |\Delta(\cdot)|$ is superadditive, and therefore by Lemma 1 we have

$$\lim_{g \rightarrow \infty} \frac{\log |\Delta(g)|}{g} = \sup_{g \geq 1} \frac{\log |\Delta(g)|}{g}.$$

In particular,

$$\frac{\log |\Delta(n-s)|}{n-s} \leq C_d,$$

and therefore,

$$\begin{aligned}
 \frac{\log \lambda((M_i))}{|(M_i)|} &\leq \frac{\log \lambda((N'_i))}{m} \\
 &= \frac{\log |\Delta(n-s)|}{n-s} \frac{n-s}{m} \\
 &\leq C_d \frac{n-s}{m} = C_d \frac{m-s(d+1)}{m} \\
 &= C_d (1 - \rho(d+1)) = h_d(\rho).
 \end{aligned}$$

This completes the proof. \blacksquare

6.5 Proof of Theorem 13

In this section we prove Theorem 13. As the proof is rather involved, we show an outline of the proof in Section 6.5.1, relying on several propositions whose proofs we defer to Section 6.5.2. Throughout this section, S , $\mathbf{p}_1, \dots, \mathbf{p}_4$ are as defined in the statement of the theorem, and we set $C_d = \text{cap}(\text{RLL}(d, 2d+2))$. Note that S is irreducible and of finite-type; let $\widehat{\mathcal{G}}_\star = (\widehat{V}_\star, \widehat{E}_\star, \widehat{\mathcal{L}}_\star)$ denote the unique non-trivial component of $\widehat{\mathcal{G}}_{\mathcal{F}_S}$. Then it can be verified that $\widehat{\mathcal{G}}_\star$ is the subgraph of $\widehat{\mathcal{G}}$ induced by \widehat{V}_\star , where

$$\begin{aligned}
 \widehat{V}_\star &= \{ \mathcal{F}_S(10^i) : 0 \leq i \leq 2d+2 \} \cup \\
 &\quad \{ \mathcal{F}_S(10^i) \cap \mathcal{F}_S(10^{i+d+1}) : 0 \leq i \leq d-1 \} \cup \\
 &\quad \{ \mathcal{F}_S(10^i) \cap \mathcal{F}_S(10^{i+d+2}) : 0 \leq i \leq d-1 \}.
 \end{aligned}$$

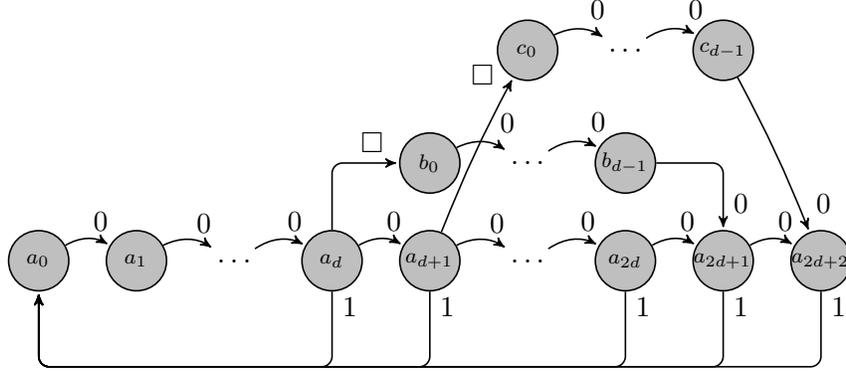
For the purpose of this proof we use the abbreviations

$$\begin{aligned}
 a_i &= \mathcal{F}_S(10^i), & 0 \leq i \leq 2d+2, \\
 b_i &= \mathcal{F}_S(10^i) \cap \mathcal{F}_S(10^{i+d+1}), & 0 \leq i \leq d-1, \text{ and} \\
 c_i &= \mathcal{F}_S(10^i) \cap \mathcal{F}_S(10^{i+d+2}), & 0 \leq i \leq d-1.
 \end{aligned}$$

The graph $\widehat{\mathcal{G}}_\star$ is shown in Figure 6.3.

Let $A = A^{\{0,1\}}(S)$, $B = A^{\{\square\}}(S)$ and set $C = A^d B A^{d+1}$; we index the entries of C by \widehat{V}_\star^2 . For a sequence of matrices $(M_i) \subseteq \{A, C\}$ we denote by $\varepsilon((M_i))$ the sequence with elements in $\{A, B\}$ formed by substituting the sequence $(A)^d (B) (A)^{d+1}$ for every element C in (M_i) . Let $h_d : [0, \text{cap}_{\text{ind}}(S)] \rightarrow [0, 1]$ be given by

$$h_d = \begin{cases} \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_4} & \text{if } 1 \leq d \leq 16 \\ \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4} & \text{if } 17 \leq d \end{cases}.$$


 Figure 6.3: The non-trivial component of $\widehat{\mathcal{G}}_{\mathcal{F}_{\text{RLL}}(d, 2d+2)}$.

6.5.1 Outline of proof

Let $i, j \in \widehat{V}_*$. Since there are only 3 paths in $\widehat{\mathcal{G}}_*$ matching $(A)^d(B)(A)^{d+1}$: a path starting at vertex a_0 and ending at vertex a_{2d+2} , a path starting at vertex a_0 and ending at vertex a_0 and a path starting at vertex a_1 and ending at vertex a_0 , it follows that the entries of C and C^2 are given by

$$C_{(i,j)} = \begin{cases} 1 & \text{if } i = a_0 \text{ and } j \in \{a_0, a_{2d+2}\} \\ 1 & \text{if } i = a_1 \text{ and } j = a_0 \\ 0 & \text{otherwise} \end{cases}, \quad i, j \in V, \quad (6.7)$$

$$(C^2)_{(i,j)} = \begin{cases} 1 & \text{if } i \in \{a_0, a_1\} \text{ and } j \in \{a_0, a_{2d+2}\} \\ 0 & \text{otherwise} \end{cases}, \quad i, j \in V. \quad (6.8)$$

The following facts easily follow:

Fact 3. $C^2 \geq C$.

Fact 4. For all integers $k \geq 2$, $C^k = C^2$.

Fact 5. $C^2 = \mathbf{c}\mathbf{r}$, where \mathbf{c} and \mathbf{r} are the column and row vectors, respectively, of size $|\widehat{V}_*|$ with entries indexed by \widehat{V}_* and given by

$$\begin{aligned} (\mathbf{c})_i &= \begin{cases} 1 & \text{if } i \in \{a_0, a_1\} \\ 0 & \text{otherwise} \end{cases} \\ (\mathbf{r})_i &= \begin{cases} 1 & \text{if } i \in \{a_0, a_{2d+2}\} \\ 0 & \text{otherwise} \end{cases} \end{aligned}, \quad i \in \widehat{V}_*.$$

Hence C^2 is a (nonnegative) rank-1 matrix. The following proposition shows how to compute the Perron eigenvalue of such a matrix.

Proposition 8. *Let M be an $m \times m$ real nonnegative matrix of the form $M = \mathbf{a}\mathbf{b}$ with \mathbf{a} and \mathbf{b} column and row vectors of size m , respectively. Then the following statements hold:*

1. *For any real nonnegative $m \times m$ matrix N , $\lambda(MN) = \mathbf{b}N\mathbf{a}$.*
2. *For any real nonnegative $m \times m$ matrices N_1, N_2 ,*

$$\lambda(MN_1MN_2) = \lambda(MN_1)\lambda(MN_2).$$

Now, observe, that by (6.2), $\text{cap}_{\text{ind}}(S) = 1/(2d+2)$; hence it follows from the discussion in Section 6.3 that

$$f_S(0) = h_d(0), \tag{6.9}$$

$$\text{and } f_S(1/(2d+2)) \geq h_d(1/(2d+2)). \tag{6.10}$$

So it's enough to show $f_S(\rho) = h_d(\rho)$ for all $\rho \in (0, \text{cap}_{\text{ind}}(S)]$. The following proposition characterizes h_d .

Proposition 9. $h_d : [0, 1/(2d+2)] \rightarrow \mathbb{R}$ is the smallest function satisfying:

1. h_d is concave.
2. $h_d(0) \geq \mathcal{C}_d$.
3. $h_d(3/(6d+8)) \geq \log 3/(6d+8)$
4. $h_d(2/(4d+5)) \geq 1/(4d+5)$
5. $h_d(1/(2d+2)) \geq 0$

We first show that for every $\rho \in [0, \text{cap}_{\text{ind}}(S)]$, $f_S(\rho) \geq h_d(\rho)$. Consider the two sequences of matrices $(Y_i) = \varepsilon((C, C, A))$ and $(Z_i) = \varepsilon((C, C, A, C, A))$. Clearly,

$$\varrho((Y_i)) = \frac{2}{4d+5}, \tag{6.11}$$

$$\varrho((Z_i)) = \frac{3}{6d+8}. \tag{6.12}$$

Now, let \mathbf{c}, \mathbf{r} be the vectors defined in Fact 5; by Proposition 8 and Fact 5,

$$\begin{aligned}\lambda((Y_i)) &= \lambda(C^2 A) \\ &= \mathbf{r} \mathbf{A} \mathbf{c} \\ &= \sum_{\substack{i \in \{a_0, a_{2d+2}\}, \\ j \in \{a_0, a_1\}}} A_{(i,j)} = 2,\end{aligned}\tag{6.13}$$

and

$$\begin{aligned}\lambda((Z_i)) &= \lambda(C^2 A C A) \\ &= \mathbf{r} A C A \mathbf{c}.\end{aligned}$$

Observe that for $i \in \widehat{V}_*$, the entry $(\mathbf{r}A)_i$ is the number of paths of length 1 in $\widehat{\mathcal{G}}_*^{\{0,1\}}$ that begin in either a_0 or a_{2d+2} and end at i . It follows that $\mathbf{r}A = \mathbf{c}^\dagger$ (where \mathbf{c}^\dagger is the transpose of \mathbf{c}). Using (6.7), we have

$$\begin{aligned}\lambda((Z_i)) &= \mathbf{c}^\dagger C A \mathbf{c} \\ &= 2A_{(a_0, a_0)} + 2A_{(a_0, a_1)} + A_{(a_{2d+2}, a_0)} + A_{(a_{2d+2}, a_1)} = 3.\end{aligned}\tag{6.14}$$

By (6.1), for any sequence $(X_i) \subseteq \{A, B\}$, $f_S(\varrho((X_i))) \geq \lambda((X_i))/|(X_i)|$; hence by (6.11), (6.12), (6.13), and (6.14) above, we get

$$f_S(2/(4d+5)) \geq 1/(4d+5),\tag{6.15}$$

$$f_S(3/(6d+8)) \geq (\log 3)/(6d+8).\tag{6.16}$$

Since f_S is concave in $[0, \text{cap}_{\text{ind}}(S)]$, equality (6.9), and inequalities (6.10), (6.15), and (6.16) imply that $f_S(\rho) \geq h_d(\rho)$ for all $\rho \in [0, \text{cap}_{\text{ind}}(S)]$.

In the remainder of this section we show $f_S(\rho) \leq h_d(\rho)$ for $\rho \in (0, \text{cap}_{\text{ind}}(S)]$. Since both f_S and h_d are continuous in this interval, it's enough to show $f_S(\rho) \leq h_d(\rho)$ for all rational $\rho \in [0, \text{cap}_{\text{ind}}(S)]$. Let ρ be such a rational. By (6.1), it's enough to show that for all finite sequences $(M_i) \subseteq \{A, B\}$ with $\varrho((M_i)) = \rho$, we have

$$\frac{\log \lambda((M_i))}{|(M_i)|} \leq h_d(\varrho((M_i))).\tag{6.17}$$

Let (M_i) be such a sequence. Set $n = |(M_i)|$ and consider the sequence $(X_i) = (M_i)^2$. Note that in any path of $\widehat{\mathcal{G}}_*$ the number of edges between a pair of consecutive edges labelled with a ' \square ' must be at least $2d+1$. It follows that if there exist nonnegative integers $1 \leq i < j \leq 2n$, with $j-i-1 < 2d+1$, such that $X_i = X_j = B$, then no paths of $\widehat{\mathcal{G}}_*$ match (X_i) or equivalently $(\prod_{i=1}^n M_i)^2 = 0$. This implies

$\lambda((M_i)) = 0$, and therefore (6.17) holds. So assume no such integers exist. It can be verified that in this case we may cyclically shift (M_i) such that $(M_i) = \varepsilon((N_i))$ for some $(N_i) \subseteq \{A, C\}$; such a cyclic shift does not change either side of (6.17). Since we assumed $\rho > 0$, the sequence (N_i) must have at least one element equal to C . If every element of (N_i) is equal to C , then $\rho = \text{cap}_{\text{ind}}(S)$ and

$$\begin{aligned} \lambda((N_i)) &= \lambda(C^{|(N_i)|}) \\ &= (\lambda(C^2))^{|(N_i)|/2} = (\mathbf{rc})^{|(N_i)|/2} = 1, \end{aligned}$$

where \mathbf{c}, \mathbf{r} are the vectors defined in Fact 5 and we used Proposition 8. Therefore (6.17) holds with equality in this case. So we assume (N_i) has an element equal to C and an element equal to A . By cyclically shifting (M_i) and (N_i) , if necessary, we may assume (N_i) is either of the form

$$\begin{aligned} (C)^{s_1}(A)^{g_1}(C)^{s_2}(A)^{g_2} \dots (C)^{s_k}(A)^{g_k}, \\ k \geq 1, g_1, \dots, g_k \geq 1, s_1 \geq 2, \text{ and } s_2, \dots, s_k \geq 1, \end{aligned} \quad (6.18)$$

or the form

$$\begin{aligned} (C)(A)^{g_1}(C)(A)^{g_2} \dots (C)(A)^{g_k}, \\ k \geq 1 \text{ and } g_1, \dots, g_k \geq 1. \end{aligned} \quad (6.19)$$

We claim it's enough to show that (6.17) holds for $(M_i) = \varepsilon((N_i))$, where (N_i) is of the form (6.18). Indeed, assume that (6.17) holds for all sequences $(M_i) = \varepsilon((N_i))$ such that $(N_i) \subseteq \{A, C\}$ is of the form (6.18), and let $(M_i) = \varepsilon((N_i))$ with (N_i) a sequence of the form (6.19). Pick a positive integer m , and set $(X_i) = \varepsilon((C)(N_i)^m)$. Then

$$\begin{aligned} \frac{\log \lambda((M_i))}{|(M_i)|} &= \frac{\log \lambda((M_i)^m)}{m|(M_i)|} = \frac{\log \lambda((CA^{g_1}CA^{g_2} \dots CA^{g_k})^m)}{m|(M_i)|} \\ &\leq \frac{\log \lambda(C(CA^{g_1}CA^{g_2} \dots CA^{g_k})^m)}{m|(M_i)|} \end{aligned} \quad (6.20)$$

$$\begin{aligned} &= \frac{\log \lambda((X_i))}{|(X_i)|} \frac{|(X_i)|}{m|(M_i)|} \\ &\leq h_d(\varrho((X_i))) \frac{|(X_i)|}{m|(M_i)|} \\ &= h_d \left(\frac{1+m|(M_i)|\varrho((M_i))}{2d+2+m|(M_i)|} \right) \frac{2d+2+m|(M_i)|}{m|(M_i)|} \end{aligned} \quad (6.21)$$

where (6.20) follows from Fact 3 and (6.21) follows from our assumption, as $(C)(N_i)^m$ is of the form (6.18). Since h_d is continuous, taking the limit of the RHS as m approaches infinity, we get that (6.17) holds for (M_i) .

6.5. Proof of Theorem 13

So henceforth, we assume (N_i) is of the form (6.18). We now further transform (N_i) by reducing runs of C elements with lengths greater than 2; that is, we (possibly) change (N_i) to be the sequence

$$(C)^u (C)^2 (A)^{g_1} (C)^{s'_2} (A)^{g_2} (C)^{s'_3} (A)^{g_3} \dots (C)^{s'_k} (A)^{g_k},$$

where each $s'_j = \min\{2, s_j\}$ and $u = \sum_j (s_j - s'_j)$. We also update (M_i) so that it still satisfies $(M_i) = \varepsilon((N_i))$. Clearly this does not change the RHS of (6.17) and by Fact 4 the LHS remains the same, as well. Now, the sequence (N_i) may be rewritten as $(N_i) = (C)^u (O_i^{(1)}) (O_i^{(2)}) \dots (O_i^{(m)})$, where each $(O_i^{(j)}) \subseteq \{A, C\}$ is given by

$$(O_i^{(j)}) = (C)^2 (A)^{g_{j,1}} (C) (A)^{g_{j,2}} (C) (A)^{g_{j,3}} \dots (C) (A)^{g_{j,k_j}}, \quad (6.22)$$

$$k_j \geq 1, \text{ and } g_{j,1}, \dots, g_{j,k_j} \geq 1.$$

Observe that by Proposition 8 we have: $\lambda(N_i) = \lambda(\prod_{j=1}^m \prod_i O_i^{(j)}) = \prod_{j=1}^m \lambda((O_i^{(j)}))$. We will use the following two propositions to further transform (N_i) .

Proposition 10. *For integers $k \geq 2$, $1 \leq i < k$, $g_1, g_2, \dots, g_{i-1}, g_{i+2}, \dots, g_k \geq 1$ and $g_i, g_{i+1} \geq 2$*

$$\lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_{i-1}} C A^{g_i} C A^{g_{i+1}} C A^{g_{i+2}} \dots C A^{g_k}) \leq$$

$$\lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_{i-1}} C A^{g_i + g_{i+1}} C A^{g_{i+2}} \dots C A^{g_k})$$

Proposition 11. *For integers $k \geq 2$, $1 \leq i < k$, $s, g_1, \dots, g_{i-1}, g_{i+2}, \dots, g_k \geq 1$ and $g_i, g_{i+1} \geq 2$*

$$\lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_{i-1}} C A^{g_i} (C A)^s C A^{g_{i+1}} C A^{g_{i+2}} \dots C A^{g_k}) \leq$$

$$\lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_{i-1}} C A^{g_i + g_{i+1}} C A^{g_{i+2}} \dots C A^{g_k} C^2 A (C A)^{s-1})$$

For each $j=1, 2, \dots, m$ we transform $(O_i^{(j)})$ in turn, resulting in a new sequence $(\tilde{O}_i^{(j)}) \subseteq \{A, C\}$. We first replace occurrences of (contiguous) subsequences of the form $(A)^{g_1} (C) (A)^{g_2}$, for some $g_1, g_2 \geq 2$, in our sequence with $(A)^{g_1 + g_2}$. Each such replacement decreases the number of elements equal to C by 1, does not change the number of elements equal to A , and by Proposition 10 does not decrease the λ of the sequence. We continue to do this until no more occurrences of such sequences exist. Let q_j be the number of the replacements we performed. Next, we consider every occurrence of a (contiguous) subsequence of the form $(C, A)^s (C)$, for some $s \in \mathbb{N}$, whose two preceding elements and two succeeding elements all equal A . For each such occurrence, in turn, we remove

6.5. Proof of Theorem 13

it, and concatenate the sequence $(C, C, A)(C, A)^{s-1}$ to the end of our current sequence. Note that after each such removal-and-concatenation the number of elements equal to A and the number of elements equal to C do not change, and by Proposition 11 and Part 2 of Proposition 8, the λ of the sequence does not decrease. We denote by $(\tilde{O}_i^{(j)})$ the resulting sequence. Then it follows from this discussion that $(C)^{q_j}(\tilde{O}_i^{(j)})$ and $(O_i^{(j)})$ have the same number of elements equal to A and the same number of elements equal to C . It further follows that

$$\lambda((\tilde{O}_i^{(j)})) \geq \lambda((O_i^{(j)})), \quad (6.23)$$

and

$$(\tilde{O}_i^{(j)}) = (R_i^{(j,1)})(R_i^{(j,2)}) \dots (R_i^{(j,w_j)}), \quad (6.24)$$

where $w_j \in \mathbb{N}$, and for each $1 \leq k \leq w_j$, the sequence $(R_i^{(j,k)})$ is either of the form

$$(C)^2(A)(C, A)^{t-1}, \quad \text{for some } t \geq 1, \quad (6.25)$$

or of the form

$$(C)^2(A, C)^s(A)^g(C, A)^t, \quad \text{for some } s, t \geq 0, g \geq 2. \quad (6.26)$$

Set $(\tilde{N}_i) = (C)^{\tilde{u}}(\tilde{O}_i^{(1)})(\tilde{O}_i^{(2)}) \dots (\tilde{O}_i^{(m)})$, where $\tilde{u} = u + \sum_{j=1}^m q_j$. Then $\varrho(\varepsilon((\tilde{N}_i))) = \varrho((M_i))$ and $|\varepsilon((\tilde{N}_i))| = |(M_i)|$. Additionally, by (6.23), Proposition 8 and Fact 4, we have $\lambda((N_i)) \leq \lambda((\tilde{N}_i))$. Now, for $j=1, \dots, m$ and $k=1, \dots, w_j$, denote by $(F_i^{(j,k)})$ the sequence $\varepsilon((R_i^{(j,k)}))$. To finish the proof, we claim, it's enough to show that for every such j and k ,

$$\frac{\log \lambda((F_i^{(j,k)}))}{|(F_i^{(j,k)})|} \leq h_d(\varrho((F_i^{(j,k)}))). \quad (6.27)$$

Indeed, assume that this holds. Then, noting that $h_d(\text{cap}_{\text{ind}}(S)) = 0$, we have

$$\begin{aligned} \frac{\log \lambda((M_i))}{|(M_i)|} &= \frac{\log \lambda((N_i))}{|(M_i)|} \leq \frac{\log \lambda((\tilde{N}_i))}{|\varepsilon((\tilde{N}_i))|} \\ &= \sum_{j=1}^m \sum_{k=1}^{w_j} \left(\frac{\log \lambda((F_i^{(j,k)}))}{|(F_i^{(j,k)})|} \frac{|(F_i^{(j,k)})|}{|\varepsilon((\tilde{N}_i))|} \right) \\ &\leq h_d(\text{cap}_{\text{ind}}(S)) \frac{|\varepsilon((C)^{\tilde{u}})|}{|\varepsilon((\tilde{N}_i))|} + \sum_{j=1}^m \sum_{k=1}^{w_j} \left(h_d(\varrho((F_i^{(j,k)}))) \frac{|(F_i^{(j,k)})|}{|\varepsilon((\tilde{N}_i))|} \right) \end{aligned} \quad (6.28)$$

$$\leq h_d \left(\text{cap}_{\text{ind}}(S) \frac{|\varepsilon((C)^{\tilde{u}})|}{|\varepsilon((\tilde{N}_i))|} + \sum_{j=1}^m \sum_{k=1}^{w_j} \left(\varrho((F_i^{(j,k)})) \frac{|(F_i^{(j,k)})|}{|\varepsilon((\tilde{N}_i))|} \right) \right) \quad (6.29)$$

$$= h_d(\varrho(\varepsilon((\tilde{N}_i)))) = h_d(\varrho((M_i))),$$

where (6.28) follows from our assumption, and (6.29) follows from the concavity of h_d asserted in Proposition 9.

So we proceed to show that (6.27) holds for every $j=1, \dots, m$, $k=1, \dots, w_j$. This follows from the next proposition and the fact that each $(R_i^{(j,k)})$ is either of the form (6.25) or (6.26).

Proposition 12. *The following statements hold.*

1. Let $t \geq 1$ be an integer and let $(F_i) = \varepsilon((C)^2(A)(C, A)^{t-1})$, then

$$\frac{\log \lambda((F_i))}{|(F_i)|} \leq h_d(\varrho((F_i))).$$

2. Let $s, t \geq 0$, $g \geq 2$ be integers and let $(F_i) = \varepsilon((C)^2(A, C)^s(A)^g(C, A)^t)$, then

$$\frac{\log \lambda((F_i))}{|(F_i)|} \leq h_d(\varrho((F_i))).$$

The proof is now completed. ■

6.5.2 Proof of propositions

Proof of Proposition 8. Part 1. Consider the matrix $MN = \mathbf{a}(\mathbf{b}N)$. Clearly, it has rank at most 1, and therefore the eigenvalue 0 has geometric multiplicity at least

$m-1$; hence $m-1$ of the eigenvalues are 0. The last eigenvalue must then equal the trace of the matrix, which in this case, is $(\mathbf{b}N)\mathbf{a}$. Obviously, it is a largest real eigenvalue.

Part 2. Using part 1 we have,

$$\begin{aligned}\lambda(MN_1MN_2) &= \lambda(M(N_1MN_2)) \\ &= \mathbf{b}(N_1MN_2)\mathbf{a} \\ &= \mathbf{b}N_1\mathbf{a}\mathbf{b}N_2\mathbf{a} \\ &= \lambda(MN_1)\lambda(MN_2).\end{aligned}\quad \blacksquare$$

Proof of Proposition 9. We make use of the following lemma.

Lemma 6. For all positive integers d ,

$$\mathcal{C}_d > \mathcal{C}_{d+1}.$$

Proof of Lemma 6. It is well known (cf. [18]) that $\mathcal{C}_d = \log \lambda_d$, where $\lambda_d > 0$ is the largest real root of the polynomial $P_d(x)$ given by

$$P_d(x) = x^{2d+3} - \sum_{i=0}^{d+2} x^i.$$

Let γ be any positive root of $P_d(x)$. Choose any $x > \gamma$, and write $x = (1+\delta)\gamma$ for $\delta > 0$. Then

$$\begin{aligned}P_d(x) &= (1+\delta)^{2d+3}\gamma^{2d+3} - \sum_{i=0}^{d+2} (1+\delta)^i\gamma^i \\ &> (1+\delta)^{2d+3}\gamma^{2d+3} - (1+\delta)^{2d+3} \sum_{i=0}^{d+2} \gamma^i \\ &= (1+\delta)^{2d+3}P_d(\gamma) \\ &= 0.\end{aligned}$$

It follows that λ_d is the only positive root of P_d , and that $P_d(x) > 0$ for all $x > \lambda_d$. Moreover, as P_d is continuous and $P_d(0) = -1 < 0$ it follows that for all $0 \leq x < \lambda_d$, $P_d(x) < 0$. Clearly, the above holds for $P_{d+1}(x)$ and λ_{d+1} as well; hence to show the claim it's enough to prove that $P_{d+1}(\lambda_d) > 0$. Now, since $P_d(\lambda_d) = 0$, we have

6.5. Proof of Theorem 13

$\lambda_d^{2d+3} = \sum_{i=0}^{d+2} \lambda_d^i$ and

$$\begin{aligned} P_{d+1}(\lambda_d) &= \lambda_d^{2d+5} - \sum_{i=0}^{d+3} \lambda_d^i \\ &= \sum_{i=2}^{d+4} \lambda_d^i - \sum_{i=0}^{d+3} \lambda_d^i \\ &= \lambda_d^{d+4} - \lambda_d - 1. \end{aligned} \tag{6.30}$$

Using $P_d(\lambda_d) = 0$ again, we get

$$\begin{aligned} \lambda_d^{d+4} &= \lambda_d^{-(d-1)} \lambda_d^{2d+3} \\ &= \sum_{i=-(d-1)}^3 \lambda_d^i > 1 + \lambda_d. \end{aligned}$$

Thus, from (6.30), we get $P_{d+1}(\lambda_d) > 0$ and the claim follows. \blacksquare

We now return to the proof of Proposition 9. We first note that for all $d \geq 1$

$$\mathcal{C}_d \leq \log(9/8) \iff d \geq 17. \tag{6.31}$$

Indeed it's a simple matter to verify that for $d=17$, the LHS holds, and for $d=16$ it does not; (6.31) then follows by applying Lemma 6.

We show that h_d is concave. Note that for a sequence $(\mathbf{x}_i) \subseteq \mathbb{R}^2$, with $|\mathbf{x}_i| = k$, $\mathbf{x}_i = (x_i, y_i)$ and $x_1 < \dots < x_k$, the function $\mathbb{L}_{(\mathbf{x}_i)}$ is concave iff the sequence of the slopes of the linear segments is non-increasing, namely,

$$\frac{y_i - y_{i-1}}{x_i - x_{i-1}} \geq \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \quad 2 \leq i \leq k-1.$$

We check this for h_d , when $d \leq 16$. In this case,

$$h_d = \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_4} = \mathbb{L}_{(0, \mathcal{C}_d), (\frac{2}{4d+5}, \frac{1}{4d+5}), (\frac{1}{2d+2}, 0)},$$

so one needs to verify that

$$\frac{1/(4d+5) - \mathcal{C}_d}{2/(4d+5)} \geq \frac{0 - 1/(4d+5)}{1/(2d+2) - 2/(4d+5)}.$$

Using simple algebraic manipulations this can be reduced to $\mathcal{C}_d \leq 1$, which certainly holds. As for the case $d \geq 17$, here,

$$\begin{aligned} h_d &= \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4} \\ &= \mathbb{L}_{(0, \mathcal{C}_d), (\frac{3}{6d+8}, \frac{\log 3}{6d+8}), (\frac{2}{4d+5}, \frac{1}{4d+5}), (\frac{1}{2d+2}, 0)}, \end{aligned}$$

so one needs to verify the following two inequalities:

$$\frac{\log 3/(6d+8) - \mathcal{C}_d}{3/(6d+8)} \geq \frac{1/(4d+5) - \log 3/(6d+8)}{2/(4d+5) - 3/(6d+8)} \quad (6.32)$$

$$\frac{1/(4d+5) - \log 3/(6d+8)}{2/(4d+5) - 3/(6d+8)} \geq \frac{0 - 1/(4d+5)}{1/(2d+2) - 2/(4d+5)}. \quad (6.33)$$

Again, using algebraic manipulations, (6.32) can be reduced to $\mathcal{C}_d \leq \log(9/8)$, which holds by (6.31) and our assumption on d , and (6.33) can be reduced to $2 \geq \log 3$, which is obviously true. Next, we verify that h_d satisfies the other properties listed in the proposition. Clearly, h_d satisfies Properties 2,4,5 with equality, and for $d \geq 17$, it satisfies Property 3 with equality as well. It remains to check Property 3 for $d \leq 16$, namely that

$$\mathbb{L}_{(0, \mathcal{C}_d), (\frac{2}{4d+5}, \frac{1}{4d+5})} (3/(6d+8)) \geq \frac{\log 3}{6d+8}.$$

The latter inequality can be reduced to $\mathcal{C}_d \geq \log(9/8)$, which holds by our assumption on d and (6.31). Thus, h_d is concave and satisfies Properties 2,3,4 and 5. It's easy to verify using the definition of h_d and concavity that it is the smallest such function. \blacksquare

Before we show the proofs of Propositions 10 and 11, we develop some tools for calculating $\lambda((X_i))$, where $(X_i) \subseteq \{A, C\}$ is a sequence of the form

$$(C)^2(A)^{g_1}(C)(A)^{g_2} \dots (C)(A)^{g_k}, \text{ for some } k \geq 1 \text{ and } g_1, \dots, g_k \geq 1. \quad (6.34)$$

To this end, we define the following sets. For a positive integer g , let $\Delta(g) \subseteq S_g$ be given by

$$\Delta(g) = \left\{ w \in S_g : \begin{array}{l} w \text{ does not begin with '0}^i\text{1', } 1 \leq i \leq d-1, \\ \text{and does not end with '00'} \end{array} \right\}.$$

For positive integers k and g_1, g_2, \dots, g_k , let $\prod_j \Delta(g_j)$ denote the cartesian product $\Delta(g_1) \times \dots \times \Delta(g_k)$; define $\Delta(g_1, \dots, g_k)$ by

$$\Delta(g_1, \dots, g_k) = \left\{ (w_1, \dots, w_k) \in \prod_j \Delta(g_j) : \begin{array}{l} \text{for } 1 \leq i < k, \text{ if } w_i \text{ ends with a '0'} \\ \text{then } w_{i+1} \text{ begins with a '0'} \end{array} \right\}.$$

Finally, for symbols $a, b \in \{0, 1\}$, and positive integers k, g_1, \dots, g_k , define the set $\Delta_{a \rightarrow b}(g_1, \dots, g_k)$ by

$$\Delta_{a \rightarrow b}(g_1, \dots, g_k) = \left\{ (w_1, \dots, w_k) \in \Delta(g_1, \dots, g_k) : \begin{array}{l} w_1 \text{ starts with } a \\ \text{and } w_k \text{ ends with } b \end{array} \right\}.$$

6.5. Proof of Theorem 13

So for example, $\Delta(1) = \{0, 1\}$, $\Delta(2) = \{01, 10\}$ if $d = 1$ and $\Delta(2) = \{10\}$ for $d \geq 2$, $\Delta(i) = \emptyset$ for $3 \leq i \leq d$, and $\Delta(d+1) = \{0^d 1\}$ for $d \geq 2$, and so on. Also note that $\Delta(1, 1, \dots, 1)$, where the number of 1's is some integer s , is given by

$$\begin{aligned} \Delta(1, 1, \dots, 1) = \{ & (0, 0, 0, \dots, 0), \\ & (1, 0, 0, \dots, 0), \\ & (1, 1, 0, \dots, 0), \\ & \vdots \\ & (1, 1, 1, \dots, 1)\}, \end{aligned} \tag{6.35}$$

and has $s+1$ elements. The following proposition shows how these sets can be used to compute $\lambda((X_i))$, with (X_i) of the form (6.34).

Proposition 13. *For all positive integers k, g_1, \dots, g_k*

$$\lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_k}) = |\Delta(g_1, \dots, g_k)|$$

Proof. For an integer n , let \mathcal{P}_n denote the set of all paths of length n in $\widehat{\mathcal{G}}_\star$. For a path $\gamma \in \mathcal{P}_n$ we denote its starting vertex (resp. terminating vertex) by $\sigma(\gamma)$ (resp. $\tau(\gamma)$). Let k and g_1, \dots, g_k be positive integers and let \mathbf{c}, \mathbf{r} be the vectors defined in Fact 5 so that $C^2 = \mathbf{c}\mathbf{r}$. By Proposition 8,

$$\lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_k}) = \mathbf{r} A^{g_1} C A^{g_2} \dots C A^{g_k} \mathbf{c}.$$

Let $(M_i) = \varepsilon((A)^{g_1}(C)(A)^{g_2} \dots (C)(A)^{g_k})$ and let Γ be the set

$$\Gamma = \{\gamma \in \mathcal{P}_{|(M_i)|} : \gamma \text{ matches } (M_i), \sigma(\gamma) \in \{a_0, a_{2d+2}\}, \tau(\gamma) \in \{a_0, a_1\}\}.$$

Then it follows that

$$\begin{aligned} \lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_k}) &= \mathbf{r} A^{g_1} C A^{g_2} \dots C A^{g_k} \mathbf{c} \\ &= \mathbf{r} \left(\prod_i M_i \right) \mathbf{c} \\ &= |\Gamma|. \end{aligned}$$

We will show that $|\Gamma| = |\Delta(g_1, \dots, g_k)|$ by exhibiting a bijection between these two sets. For $i=1, 2, \dots, k$ define $s_i = 1 + \sum_{j=1}^{i-1} (g_j + 2d + 2)$, and $t_i = s_i + g_i - 1$. Then $1 = s_1 < t_1 < s_2 < t_2 < \dots < s_k < t_k = |(M_i)|$ (s_i, t_i denote the start and end indices of the sequence $(A)^{g_i}$ in (M_i)). For a path $\gamma = (e_1, \dots, e_n) \in \mathcal{P}_n$ let $\widehat{\mathcal{L}}_\star(\gamma) = \widehat{\mathcal{L}}_\star(e_1) \widehat{\mathcal{L}}_\star(e_2) \dots \widehat{\mathcal{L}}_\star(e_n) \in \{0, 1, \square\}^n$, be the word generated by the path,

and for $1 \leq i \leq j \leq n$ denote by $\widehat{\mathcal{L}}_\star(\gamma)_i$ the symbol $\widehat{\mathcal{L}}_\star(e_i)$ and by $\widehat{\mathcal{L}}_\star(\gamma)_{i:j}$ the word $\widehat{\mathcal{L}}_\star(e_i)\widehat{\mathcal{L}}_\star(e_{i+1})\dots\widehat{\mathcal{L}}_\star(e_j) \in \{0, 1, \square\}^*$. Note that, as $\widehat{\mathcal{G}}_\star$ is the only non-trivial component of $\widehat{\mathcal{G}}$, a word $w \in \{0, 1, \square\}^*$ is generated by some path in $\widehat{\mathcal{G}}_\star$ iff for any nonnegative integer m , there exists words $z, y \in \{0, 1, \square\}^m$, such that $zwy \in \widehat{\mathcal{S}}$. Next, fix $\gamma \in \Gamma$. Since $M_j = B$ iff $j = t_i + d + 1$ for some $1 \leq i < k$, we have $\widehat{\mathcal{L}}_\star(\gamma)_{t_i+d+1} = \square$ for $1 \leq i < k$, and $\widehat{\mathcal{L}}_\star(\gamma)_j \in \{0, 1\}$ for all $j \in \{1, \dots, t_k\} \setminus \{t_1+d+1, t_2+d+1, \dots, t_{k-1}+d+1\}$. Also, observe that in any path $\delta \in \mathcal{P}_{2d+1}$ with $\widehat{\mathcal{L}}_\star(\delta)_{d+1} = \square$, it must hold that $\widehat{\mathcal{L}}_\star(\delta) = 0^d \square 0^d$. It follows that $\widehat{\mathcal{L}}_\star(\gamma)_{t_i+1:t_i+2d+1} = 0^d \square 0^d$ for all $1 \leq i < k$. We now define $\phi : \Gamma \rightarrow \Delta(g_1, \dots, g_k)$ by

$$\phi(\gamma) = (\widehat{\mathcal{L}}_\star(\gamma)_{s_1:t_1}, \widehat{\mathcal{L}}_\star(\gamma)_{s_2:t_2}, \dots, \widehat{\mathcal{L}}_\star(\gamma)_{s_k:t_k}), \gamma \in \Gamma,$$

and claim that it is a bijection. To show this, we need to verify the following statements:

1. ϕ is well-defined: for all $\gamma \in \Gamma$, $\phi(\gamma) \in \Delta(g_1, \dots, g_k)$
2. ϕ is one-to-one.
3. ϕ is onto $\Delta(g_1, \dots, g_k)$.

1. Let $\gamma \in \Gamma$ and $1 \leq i \leq k$. Since $\widehat{\mathcal{L}}_\star(\gamma)_{s_i:t_i} \in \{0, 1\}^*$ we have $\widehat{\mathcal{L}}_\star(\gamma)_{s_i:t_i} \in \widehat{\mathcal{S}}$. Since $\sigma(\gamma) \in \{a_0, a_{2d+2}\}$, $\widehat{\mathcal{L}}_\star(\gamma)$ does not begin with ‘0^r1’ for any $1 \leq r < d$ which implies $\widehat{\mathcal{L}}_\star(\gamma)_{s_1:t_1}$ does not begin with ‘0^r1’ for any $1 \leq r < d$, as well. For $2 \leq i \leq k$, $\widehat{\mathcal{L}}_\star(\gamma)_{s_i:t_i}$ does not begin with ‘0^r1’ for any $1 \leq r < d$, since otherwise $\widehat{\mathcal{L}}_\star(\gamma)_{t_{i-1}+1:s_i+r} = 0^d \square 0^d 00^r 1$ or $\widehat{\mathcal{L}}_\star(\gamma)_{t_{i-1}+1:s_i+r} = 0^d \square 0^d 10^r 1$, and both words are not in $\widehat{\mathcal{S}}$. Additionally, for $1 \leq i \leq k-1$, $\widehat{\mathcal{L}}_\star(\gamma)_{s_i:t_i}$ does not end with ‘00’ since otherwise $\widehat{\mathcal{L}}_\star(\gamma)_{t_{i-1}:t_i+2d+1} = 000^d \square 0^d$ which is not in $\widehat{\mathcal{S}}$. And, as $\tau(\gamma) \in \{a_0, a_1\}$, $\widehat{\mathcal{L}}_\star(\gamma)$ does not end in ‘00’, which implies that $\widehat{\mathcal{L}}_\star(\gamma)_{s_k:t_k}$ does not end with ‘00’, as well. This shows that for all $1 \leq i \leq k$, $\widehat{\mathcal{L}}_\star(\gamma)_{s_i:t_i} \in \Delta(g_i)$. Finally, let $1 \leq i \leq k-1$, and assume $\widehat{\mathcal{L}}_\star(\gamma)_{s_i:t_i}$ ends with a ‘0’. If $\widehat{\mathcal{L}}_\star(\gamma)_{s_{i+1}:t_{i+1}}$ begins with a ‘1’, then $\widehat{\mathcal{L}}_\star(\gamma)_{t_i:s_{i+1}} = 00^d \square 0^d 01$ or $\widehat{\mathcal{L}}_\star(\gamma)_{t_i:s_{i+1}} = 00^d \square 0^d 11$, and both are not in $\widehat{\mathcal{S}}$. Therefore, if $\widehat{\mathcal{L}}_\star(\gamma)_{s_i:t_i}$ ends with a ‘0’ then $\widehat{\mathcal{L}}_\star(\gamma)_{s_{i+1}:t_{i+1}}$ must begin with a ‘0’. It follows that $\phi(\gamma) \in \Delta(g_1, \dots, g_k)$.

2. Let $\gamma_1, \gamma_2 \in \Gamma$, such that $\phi(\gamma_1) = \phi(\gamma_2)$. Clearly, all nonempty paths in G starting at a_{2d+2} generate a word beginning with ‘1’ and all nonempty paths in $\widehat{\mathcal{G}}_\star$ starting at a_0 generate a word beginning with ‘0’. By our assumption, $\widehat{\mathcal{L}}_\star(\gamma_1)$ and $\widehat{\mathcal{L}}_\star(\gamma_2)$ begin with the same symbol; hence $\sigma(\gamma_1) = \sigma(\gamma_2)$. Now, observe that for any path $\delta \in \Gamma$, and $2 \leq i \leq k$, $\widehat{\mathcal{L}}_\star(\delta)_{s_{i-1}} = \widehat{\mathcal{L}}_\star(\delta)_{s_i}$, where, $\bar{0} = 1$ and $\bar{1} = 0$; otherwise, $\widehat{\mathcal{L}}_\star(\delta)_{t_{i-1}+1:s_i} = 0^d \square 0^d 00$ or $\widehat{\mathcal{L}}_\star(\delta)_{t_{i-1}+1:s_i} = 0^d \square 0^d 11$ and both words are not in

\widehat{S} . Thus for all $2 \leq i \leq k$,

$$\widehat{\mathcal{L}}_\star(\gamma_1)_{t_{i-1}+1:s_{i-1}} = 0^d \square 0^d \overline{\widehat{\mathcal{L}}_\star(\gamma_1)_{s_i}} = 0^d \square 0^d \overline{\widehat{\mathcal{L}}_\star(\gamma_2)_{s_i}} = \widehat{\mathcal{L}}_\star(\gamma_2)_{t_{i-1}+1:s_{i-1}}.$$

It follows that $\widehat{\mathcal{L}}_\star(\gamma_1) = \widehat{\mathcal{L}}_\star(\gamma_2)$ and since $\widehat{\mathcal{G}}_\star$ is deterministic, we have $\gamma_1 = \gamma_2$. Therefore ϕ is one-to-one.

3. Let $\mathbf{j} = (j_1, j_2, \dots, j_n)$ be an n -tuple of positive integers for some $n \geq 1$. Consider the set $\Delta(j_1, \dots, j_n) = \Delta(\mathbf{j})$. For a word $\mathbf{w} = (w^{(1)}, \dots, w^{(n)}) \in \Delta(\mathbf{j})$, with $w^{(i)} = w_1^{(i)} w_2^{(i)} \dots w_{j_i}^{(i)}$, $w_r^{(i)} \in \{0, 1\}$, $1 \leq r \leq j_i$, $1 \leq i \leq n$, we define the word $z(\mathbf{w}) \in \widehat{\Sigma}^*$ by

$$z(\mathbf{w}) = w^{(1)} 0^d \square 0^d \overline{w_1^{(2)}} w^{(2)} 0^d \square 0^d \overline{w_1^{(3)}} w^{(3)} \dots 0^d \square 0^d \overline{w_1^{(n)}} w^{(n)}.$$

It is not hard to verify that for all $\mathbf{w} \in \Delta(\mathbf{j})$, any filling of the ‘ \square ’s of $z(\mathbf{w})$ with symbols from $\{0, 1\}$ does not contain the pattern 0^{2d+3} nor any of the patterns $10^r 1$ for $0 \leq r < d$. It follows that $z(\mathbf{w}) \in \widehat{S}$. More is true; fix a $\mathbf{w} \in \Delta(\mathbf{j})$ and for an integer m , let $\mathbf{j}_m = (y_1, \dots, y_m, j_1, \dots, j_n, y_1, \dots, y_m)$ with $y_1 = y_2 = \dots = y_m = 1$, and $\mathbf{w}_m = (x_1, \dots, x_m, w^{(1)}, \dots, w^{(n)}, \overline{x_1}, \dots, \overline{x_m})$ with $x_1 = x_2 = \dots = x_m = 1$. Then, clearly $\mathbf{w}_m \in \Delta(\mathbf{j}_m)$ and by our previous argument $z(\mathbf{w}_m) \in \widehat{S}$. It thus follows that $z(\mathbf{w})$ can be extended indefinitely on both sides and is thus generated by a path in $\widehat{\mathcal{G}}_\star$.

Now, let $\mathbf{w} = (w^{(1)}, \dots, w^{(k)}) \in \Delta(g_1, \dots, g_k)$, and let γ be a path in $\widehat{\mathcal{G}}_\star$ generating $z(\mathbf{w})$. We show that there is a path $\gamma' \in \Gamma$ generating $z(\mathbf{w})$. This will conclude the proof, since obviously for such a path γ' , $\phi(\gamma') = \mathbf{w}$. If $k = g_1 = 1$, then $z(\mathbf{w}) \in \{0, 1\}$ and clearly there is a path $\gamma' \in \Gamma$ generating $z(\mathbf{w})$. So we assume $k > 1$ or $g > 1$. In this case, observe that the length of $z(\mathbf{w})$ is at least 2, and exactly one of the last two symbols of $z(\mathbf{w})$ must be a ‘1’. Therefore, $\tau(\gamma) \in \{a_0, a_1\}$. Additionally, either $z(\mathbf{w})$ has the prefix ‘1’, or $z(\mathbf{w})$ begins with a ‘0’ and either $g_1 > 1$, in which case $z(\mathbf{w})$ has the prefix $0^r 1$ with $r \geq d$, or $g_1 = 1$ and (since $k > 1$), $z(\mathbf{w})$ has the prefix $00^d \square 0^d 1$. It is easily verified from Figure 6.3, that in every case there is a path β in $\widehat{\mathcal{G}}_\star$ generating the corresponding prefix of $z(\mathbf{w})$ with $\sigma(\beta) \in \{a_0, a_{2d+2}\}$ and $\tau(\beta) = a_0$. By replacing the initial part of γ generating this prefix with β we obtain a path γ' generating $z(\mathbf{w})$ with $\sigma(\gamma') \in \{a_0, a_{2d+2}\}$ and $\tau(\gamma') = \tau(\gamma) \in \{a_0, a_1\}$. Clearly γ' matches (M_i) , and consequently $\gamma' \in \Gamma$. ■

To prove Propositions 10 and 11 we need the following lemma, which shows a relation between $\Delta(g_1, g_2)$ and $\Delta(g_1 + g_2)$ for $g_1, g_2 \geq 2$.

Lemma 7. *For all $g_1, g_2 \geq 2$, there exists a function $T: \Delta(g_1, g_2) \rightarrow \Delta(g_1 + g_2)$ such that*

1. T is one-to-one.
2. For any $(ax, yb) \in \Delta(g_1, g_2)$, if $T(ax, yb) = cwd$, where $a, b, c, d \in \{0, 1\}$ and $x, y, w \in \{0, 1\}^*$, then $d=b$ and $c \leq a$ (using the normal order of the integers in $\{0, 1\}$).

Proof. Let $g_1, g_2 \geq 2$. For $(x, y) \in \Delta(g_1, g_2)$, we define $T(x, y)$ as follows.

1. If $x = w10, y = 0z$, where $w, z \in \{0, 1\}^*$ and z doesn't start with 0^{2d+1} , then

$$T(x, y) = xy.$$

2. If $x = w0^{d+1}10, y = 0^{2d+2}1z$, where $w, z \in \{0, 1\}^*$, then

$$T(x, y) = w0^d10^{d+2}10^{d+1}1z.$$

3. If $x = w10, y = 0^{2d+2}1z$, where $w, z \in \{0, 1\}^*$ and w doesn't end with 0^{d+1} , then

$$T(x, y) = w0^{d+2}10^{d+1}1z.$$

4. If $x = w1, y = 0z$, where $w, z \in \{0, 1\}^*$, then

$$T(x, y) = xy.$$

5. If $x = w1, y = 1z$, where $w, z \in \{0, 1\}^*$ and w doesn't end with 0^{2d+2} , then

$$T(x, y) = w01z.$$

6. If $x = w0^{2d+2}1, y = 1z$, where $w, z \in \{0, 1\}^*$, then

$$T(x, y) = w0^{d+2}10^d1z.$$

For $i = 1, 2, \dots, 6$, let Λ_i denote the subset of pairs $(x, y) \in \Delta(g_1, g_2)$ satisfying the conditions of case i above. Then, one can verify that $\{\Lambda_1, \dots, \Lambda_6\}$ is a partition of $\Delta(g_1, g_2)$ and that in each case, $T(x, y) \in \Delta(g_1 + g_2)$, thus T is well-defined. Moreover, if $(ax', y'b) \in \Lambda_i$ and $T(ax', y'b) = cwd$, where $a, b, c, d \in \{0, 1\}$ and $x', y', w \in \{0, 1\}^*$, then it's easy to verify that $d=b$, and, unless $i=3$, then $c=a$; for $i = 3$, $c \leq a$. It remains to show that T is one-to-one. First, observe that T restricted to Λ_i is one-to-one for $i = 1, 2, \dots, 6$. Next, for $(x, y) \in \Delta(g_1, g_2)$, let $T(x, y)^{(1)}, T(x, y)^{(2)} \in \{0, 1\}^*$ be the sub-words of $T(x, y)$ given by

$$T(x, y) = T(x, y)^{(1)}T(x, y)^{(2)}, \quad |T(x, y)^{(1)}| = g_1, |T(x, y)^{(2)}| = g_2,$$

and consider the following table.

Conditions	$T(x, y)^{(1)}$	$T(x, y)^{(2)}$
$(x, y) \in \Lambda_1$	$v10$	$0w$
$(x, y) \in \Lambda_2$	$v100$	$0w$
$(x, y) \in \Lambda_3, g_1=2$	00	$0w$
$(x, y) \in \Lambda_3, g_1 \geq 3$	$v000$	$0w$
$(x, y) \in \Lambda_4$	$v1$	$0w$
$(x, y) \in \Lambda_5$	$v0^{d+1}$	$1w$
$(x, y) \in \Lambda_6$	$v10^d$	$1w$

Each entry in the leftmost column describes the conditions on x, y, g_1 under which $T(x, y)^{(1)}$ and $T(x, y)^{(2)}$ have the forms written in the corresponding entries under the rightmost two columns; here v, w denote arbitrary words over $\{0, 1\}$. For example, the first line claims that if $(x, y) \in \Lambda_1$ then $T(x, y)^{(1)}$ ends with ‘10’ and $T(x, y)^{(2)}$ begins with a ‘0’. This and the claims corresponding to the other lines can be easily verified from the definition of T . Clearly, for any $(x, y) \in \Lambda_i$ and $(s, t) \in \Lambda_j$ with $i \neq j$, we have $T(x, y) \neq T(s, t)$, and it follows that T is one-to-one. ■

We can now prove Propositions 10 and 11. For a vector $\mathbf{v} = (v_1, v_2, \dots, v_k)$, with positive integer entries, we use the abbreviation $\Delta(\mathbf{v})$ for $\Delta(v_1, v_2, \dots, v_k)$, and for a positive integer k we denote by $\mathbf{1}_k$ the vector in \mathbb{Z}^k with every entry equal to 1.

Proof of Proposition 10. Let $T : \Delta(g_i, g_{i+1}) \rightarrow \Delta(g_i + g_{i+1})$ be a function satisfying the properties listed in Lemma 7. Set $\mathbf{g} = (g_1, \dots, g_k)$ and $\mathbf{g}' = (g_1, \dots, g_{i-1}, g_i + g_{i+1}, g_{i+2}, \dots, g_k)$ and $U : \Delta(\mathbf{g}) \rightarrow \Delta(\mathbf{g}')$ be the function defined by

$$U(x_1, \dots, x_k) = (x_1, \dots, x_{i-1}, T(x_i, x_{i+1}), x_{i+2}, \dots, x_k), (x_1, \dots, x_k) \in \Delta(\mathbf{g}).$$

Since T satisfies Property (2) in Lemma 7, it follows that for every $z \in \Delta(\mathbf{g})$, $U(z) \in \Delta(\mathbf{g}')$; hence U is well-defined. Since T is one-to-one, so is U and therefore $|\Delta(\mathbf{g})| \leq |\Delta(\mathbf{g}')|$. The claim now follows from Proposition 13. ■

Proof of Proposition 11. Let $\mathbf{g}^{(1)} = (g_1, \dots, g_i) \in \mathbb{Z}^i$ and $\mathbf{g}^{(2)} = (g_{i+1}, \dots, g_k) \in \mathbb{Z}^{k-i}$, where every $g_j > 0$. Define the block-vectors

$$\mathbf{g} = \left(\mathbf{g}^{(1)} \mid \mathbf{1}_s \mid \mathbf{g}^{(2)} \right)$$

$$\mathbf{g}' = \left(\mathbf{g}^{(1)} \mid \mathbf{g}^{(2)} \right).$$

6.5. Proof of Theorem 13

Finally, let $U : \Delta(\mathbf{g}) \rightarrow \Delta(\mathbf{g}') \times \Delta(\mathbf{1}_s)$ be the function given by

$$U(x_1, \dots, x_i, y_1, \dots, y_s, x_{i+1}, \dots, x_k) = (x_1, \dots, x_k, y_1, \dots, y_s),$$

for all $(x_1, \dots, x_i, y_1, \dots, y_s, x_{i+1}, \dots, x_k) \in \Delta(\mathbf{g})$. Note that for such a vector, if x_i ends with a '0', then $y_j=0$, for $j = 1, 2, \dots, s$, and consequently x_{i+1} must begin with a '0'. Therefore $(x_1, \dots, x_i, x_{i+1}, \dots, x_k) \in \Delta(\mathbf{g}')$ and U is well-defined. Since U is obviously one-to-one, it follows that $|\Delta(\mathbf{g})| \leq |\Delta(\mathbf{g}')| |\Delta(\mathbf{1}_s)|$. Now, set $M = C^2 A^{g_1} C A^{g_2} \dots C A^{g_i} (C A)^s C A^{g_{i+1}} \dots C A^{g_k}$; then by Propositions 13, 10 and 8, we get

$$\begin{aligned} \lambda(M) &= |\Delta(\mathbf{g})| \leq |\Delta(\mathbf{g}')| |\Delta(\mathbf{1}_s)| \\ &= \lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_{i-1}} C A^{g_i} C A^{g_{i+1}} C A^{g_{i+2}} \dots C A^{g_k}) \lambda(C^2 A (C A)^{s-1}) \\ &\leq \lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_{i-1}} C A^{g_i+g_{i+1}} C A^{g_{i+2}} \dots C A^{g_k}) \lambda(C^2 A (C A)^{s-1}) \\ &= \lambda(C^2 A^{g_1} C A^{g_2} \dots C A^{g_{i-1}} C A^{g_i+g_{i+1}} C A^{g_{i+2}} \dots C A^{g_k} C^2 A (C A)^{s-1}). \end{aligned}$$

■

Proof of Proposition 12, part 1. For $t=1$, it can be easily verified using Proposition 13, that the conclusion holds with equality. So assume $t \geq 2$; in this case, setting $\rho = \rho((\mathbf{F}_i))$ we have,

$$\rho = \frac{t+1}{(t+1)(2d+2)+t} = \frac{1}{2d+3-1/(t+1)} \leq \frac{1}{2d+8/3} = \frac{3}{6d+8}.$$

By Proposition 9, it follows that $h_d(\rho) \geq \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}(\rho)$. So, it's enough to show

$$\frac{\log \lambda((F_i))}{|(F_i)|} \leq \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}(\rho)$$

Now, $\lambda((F_i)) = \lambda(C^2 A (C A)^{t-1})$, which by Proposition 13 is $|\Delta(\mathbf{1}_t)|$; so by (6.35), $\lambda((F_i)) = t+1$. Therefore, we need to show

$$\begin{aligned} &\frac{\log \lambda((F_i))}{|(F_i)|} \leq \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}(\rho) \\ \iff &\frac{\log \lambda((F_i))}{|(F_i)|} \leq \mathbb{L}_{(0, \mathcal{C}_d), (\frac{3}{6d+8}, \frac{\log 3}{6d+8})}(\rho) \\ \iff &\frac{\log(t+1)}{|(F_i)|} \leq \frac{\mathcal{C}_d - (\log 3)/(6d+8)}{-3/(6d+8)} \left(\rho - \frac{3}{6d+8} \right) + \frac{\log 3}{6d+8} \\ \iff &\frac{\log(t+1)}{|(F_i)|} \leq \mathcal{C}_d \left(1 - \frac{\rho}{3/(6d+8)} \right) + \frac{\log 3}{3} \rho \\ \iff &\frac{\log(t+1)}{|(F_i)|} \leq \mathcal{C}_d \left(1 - \frac{\rho}{3/(6d+8)} \right) + \frac{(t+1)(\log 3)/3}{|\mathbf{F}_i|}. \end{aligned}$$

6.5. Proof of Theorem 13

The last inequality holds, since $\rho \leq 3/(6d+8)$, $\mathcal{C}_d \geq 0$, and for any nonnegative integer t , $(t+1)(\log 3)/3 \geq \log(t+1)$. ■

It remains to prove part 2 of Proposition 12. For this, we require the following two lemmas. The first establishes some properties of $|\Delta(\cdot)|$ and $|\Delta_{a \rightarrow b}(\cdot)|$, for $a, b \in \{0, 1\}$.

Lemma 8. *The following statements hold.*

1. $|\Delta(\cdot)|$ and $|\Delta_{a \rightarrow b}(\cdot)|$ for all bits $a, b \in \{0, 1\}$ obey the RLL($d, 2d+2$) recursion. Namely, for all positive integers g ,

$$\begin{aligned} |\Delta(g+2d+3)| &= \sum_{i=0}^{d+2} |\Delta(g+i)| \\ |\Delta_{a \rightarrow b}(g+2d+3)| &= \sum_{i=0}^{d+2} |\Delta_{a \rightarrow b}(g+i)|. \end{aligned} \tag{6.36}$$

2. $|\Delta(g)| \leq |\Delta(g+1)|$ for all $g \geq 3$.
3. $\log(|\Delta(\cdot)|)$ is “eventually superadditive”: for all $g_1, g_2 \geq 3$,

$$|\Delta(g_1)| |\Delta(g_2)| \leq |\Delta(g_1+g_2)|$$

4. $\lim_{g \rightarrow \infty} \frac{\log |\Delta(g)|}{g} = \sup_{g \geq 2} \frac{\log |\Delta(g)|}{g} = \mathcal{C}_d$

5. $|\Delta_{0 \rightarrow 0}(g)| = |\Delta_{1 \rightarrow 1}(g)|$ for all positive integers g .
6. $|\Delta_{0 \rightarrow 1}(g)| \leq 2|\Delta_{0 \rightarrow 0}(g)|$ for $g \neq d+1$.
7. $|\Delta_{0 \rightarrow 1}(g)| \leq \frac{1}{2}|\Delta(g)|$ for all g if $d = 1$, and for $g \neq d+1$ if $d > 1$.

Proof. Part 1. Let $V' \subseteq V$ be the set of vertices $\{a_0, \dots, a_{2d+2}\}$, and let \mathcal{G}' be the subgraph of $\widehat{\mathcal{G}}_\star$ induced by V' (so \mathcal{G}' is the “conventional” deterministic presentation of RLL($d, 2d+2$)). For vertices $u, v \in V'$ and positive integer g , let $\mathcal{W}_{u \rightarrow v}(g) \in \{0, 1\}^g$ denote the set of words that are generated by paths of length g in \mathcal{G}' , starting at u and terminating at v . Since \mathcal{G}' is deterministic, clearly, the number of such paths is $|\mathcal{W}_{u \rightarrow v}(g)|$. Let g be a positive integer. It’s not hard to verify that

1. $\Delta_{0 \rightarrow 0}(g) = \mathcal{W}_{a_0 \rightarrow a_1}(g)$
2. $\Delta_{0 \rightarrow 1}(g) = \mathcal{W}_{a_0 \rightarrow a_0}(g)$

$$3. \Delta_{1 \rightarrow 0}(g) = \mathcal{W}_{a_{2d+2} \rightarrow a_1}(g)$$

$$4. \Delta_{1 \rightarrow 1}(g) = \mathcal{W}_{a_{2d+2} \rightarrow a_0}(g)$$

Let $A' = A(\mathcal{G}')$. It follows that each $|\Delta_{a \rightarrow b}(g)|$, for $a, b \in \{0, 1\}$, is equal to a single entry of $(A')^g$. Now, the characteristic polynomial of A' is (cf. [18])

$$x^{2d+3} - \sum_{i=0}^{d+2} x^i.$$

Invoking the Cayley-Hamilton Theorem, we get

$$(A')^{g+2d+3} = \sum_{i=0}^{d+2} (A')^{g+i}.$$

Thus, every $|\Delta_{a \rightarrow b}(g)|$ satisfies the required recursion and therefore also $|\Delta(g)| = \sum_{a,b} |\Delta_{a \rightarrow b}(g)|$.

Part 2. For $d=1$, any word in $\Delta(g)$ can be extended from the left to a word in $\Delta(g+1)$ and the claim follows. For $d \geq 2$, we use induction on g . Table 6.1, which can be easily verified, shows the sets $\Delta(g)$ for $g=1, 2, \dots, 2d+3$ along with $|\Delta(g)|$ for $g=1, 2, \dots, 2d+6$ and $d \geq 2$. Evidently, $|\Delta(g)| \leq |\Delta(g+1)|$ for $3 \leq g \leq 2d+5$. This shows the induction basis. The induction step follows from the recursion relation (6.36).

Part 3. Since the claim is symmetric in g_1, g_2 , it's enough to prove it only for $g_1 \leq g_2$. We use induction on g_2 . For the basis of the induction we verify the claim for all $3 \leq g_1 \leq g_2 \leq 2d+5$. For $1 \leq d \leq 7$ we verified the induction basis using a computer, so here we assume $d \geq 8$. Consider Table 6.1. If $3 \leq g_1 \leq d$, then $|\Delta(g_1)| = 0$ and the claim holds trivially. If $g_1 = d+1$, then $|\Delta(g_1)| = 1$ and the claim follows from the monotonicity of $|\Delta(g)|$ for $g \geq 3$ shown in part 2. If $g_1 = g_2 = d+2$, the claim holds since $|\Delta(d+2)|^2 = 9 \leq |\Delta(2d+4)| = 11$. If $g_1 = d+2$ and $d+3 \leq g_2 \leq 2d+1$ then $|\Delta(d+2)||\Delta(g_2)| = 12 < |\Delta(2d+5)| \leq |\Delta(d+2+g_2)|$, with the last inequality following from the monotonicity of $|\Delta(g)|$, for $g \geq 3$. If $g_1 = d+2$ and $2d+2 \leq g_2 \leq 2d+5$ then

$$\begin{aligned} |\Delta(d+2+g_2)| &\geq |\Delta(3d+4)| = \sum_{i=d+1}^{2d+3} |\Delta(i)| \\ &= 1+3+(d-1)4+5+8 \geq 45 \\ &> |\Delta(d+2)||\Delta(2d+5)| \geq |\Delta(d+2)||\Delta(g_2)|. \end{aligned}$$

If $d+3 \leq g_1 \leq 2d+1$ and $g_1 \leq g_2 \leq 2d+1$ then

$$|\Delta(g_1)||\Delta(g_2)| = 16 = |\Delta(2d+6)| \leq |\Delta(g_1+g_2)|.$$

6.5. Proof of Theorem 13

g	$\Delta(g)$	$ \Delta(g) $
1	$\{0, 1\}$	2
2	$\{10\}$	1
$3 \leq g \leq d$	\emptyset	0
$d+1$	$\{0^d 1\}$	1
$d+2$	$\{0^d 10, 0^{d+1} 1, 10^d 1\}$	3
$d+3 \leq g \leq 2d+1$	$\{0^{g-2} 10, 0^{g-1} 1, 10^{g-2} 1, 10^{g-3} 10\}$	4
$2d+2$	$\{0^{2d} 10, 0^d 10^d 1, 0^{2d+1} 1, 10^{2d-1} 10, 10^{2d} 1\}$	5
$2d+3$	$\{0^{2d+1} 10, 0^d 10^d 10, 0^{2d+2} 1, 0^{d+1} 10^d 1, 0^d 10^{d+1} 1, 10^{2d} 10, 10^{2d+1} 1, 10^d 10^d 1\}$	8
$2d+4$	$\{\dots\}$	11
$2d+5$	$\{\dots\}$	14, if $d=2$; 13, if $d \geq 3$.
$2d+6$	$\{\dots\}$	21, if $d=2$; 17, if $d=3$; 16, if $d \geq 4$.

Table 6.1: Values of $|\Delta(g)|$ for $1 \leq g \leq 2d+6$ and $d \geq 2$.

If $d+3 \leq g_1 \leq 2d+1$ and $2d+2 \leq g_2 \leq 2d+5$ then

$$\begin{aligned}
 |\Delta(g_1+g_2)| &\geq |\Delta(3d+5)| = \sum_{i=d+2}^{2d+4} |\Delta(i)| \\
 &= 3 + (d-1)4 + 5 + 8 + 11 \geq 55 \\
 &> |\Delta(g_1)| |\Delta(2d+5)| \geq |\Delta(g_1)| |\Delta(g_2)|.
 \end{aligned}$$

Finally, if $2d+2 \leq g_1 \leq g_2 \leq 2d+5$ then, since $d \geq 8$ and hence $2d+11 \leq 3d+3$, we

have

$$\begin{aligned}
 |\Delta(g_1+g_2)| &\geq |\Delta(4d+4)| = \sum_{i=2d+1}^{3d+3} |\Delta(i)| \geq \sum_{i=2d+1}^{2d+11} |\Delta(i)| \\
 &= \sum_{i=2d+1}^{2d+6} |\Delta(i)| + \sum_{i=2d+7}^{2d+11} |\Delta(i)| = 57 + \sum_{i=2d+7}^{2d+11} \sum_{j=d+1}^{2d+3} |\Delta(i-j)| \\
 &\geq 57 + \sum_{i=2d+7}^{2d+11} (1+3+4(i-(d+1)-(d+2))) = 197 \\
 &> |\Delta(2d+5)|^2 \geq |\Delta(g_1)| |\Delta(g_2)|.
 \end{aligned}$$

This shows the basis of the induction.

As for the induction step, let $k \geq 2d+5$ be an integer and assume the claim holds for all $3 \leq g_1 \leq g_2 \leq k$. We will prove that the claim holds for all $3 \leq g_1 \leq g_2 \leq k+1$. For $g_1=g_2=k+1$ we have

$$\begin{aligned}
 |\Delta(k+1)| |\Delta(k+1)| &= \sum_{i=d+1}^{2d+3} \sum_{j=d+1}^{2d+3} |\Delta(k+1-i)| |\Delta(k+1-j)| \\
 &\leq \sum_{i=d+1}^{2d+3} \sum_{j=d+1}^{2d+3} |\Delta(2k+2-i-j)| = \sum_{i=d+1}^{2d+3} |\Delta(2k+2-i)| \\
 &= |\Delta(2k+2)|.
 \end{aligned}$$

where the equalities follow from (6.36), and the inequality follows from the induction hypotheses (note that in the first double sum $k+1-i \geq 3$ and $k+1-j \geq 3$ due to our assumption on k). The case $g_2=k+1, 3 \leq g_1 \leq k$ is handled in a similar manner.

Part 4. Let $\theta : \{1, 2, \dots\} \rightarrow [0, \infty)$ be the function defined by

$$\theta(g) = \begin{cases} 0 & \text{if } g \leq d+1 \\ \log |\Delta(g)| & \text{otherwise.} \end{cases}, \quad g = 1, 2, \dots$$

Then it's easily verified, using parts 2 and 3, that this function is superadditive for all positive integers g . Hence, by Lemma 1, $\lim_{g \rightarrow \infty} \theta(g)/g$ exists and satisfies

$$\lim_{g \rightarrow \infty} \frac{\theta(g)}{g} = \sup_{g \geq 1} \frac{\theta(g)}{g}.$$

Since θ is nonnegative, the RHS is equal to $\sup_{g \geq d+2} (\theta(g)/g)$, thus

$$\lim_{g \rightarrow \infty} \frac{\theta(g)}{g} = \lim_{g \rightarrow \infty} \frac{\log |\Delta(g)|}{g} = \sup_{g \geq d+2} \frac{\theta(g)}{g} = \sup_{g \geq d+2} \frac{\log |\Delta(g)|}{g}.$$

Let A' be the matrix defined in the proof of part 1 above, and let $a, b \in \{0, 1\}$. Since A' is primitive, it holds that $\lim_{g \rightarrow \infty} ((A')^g / \lambda(A')^g)$ exists and is strictly positive in each entry. As $\Delta_{a \rightarrow b}(g)$ is equal to a single entry of $(A')^g$, it follows that there exists a positive real constant $c_{a,b}$, such that

$$\lim_{g \rightarrow \infty} \frac{|\Delta_{a \rightarrow b}(g)|}{\lambda(A')^g} = c_{a,b}.$$

Since $|\Delta(g)| = \sum_{a,b} |\Delta_{a \rightarrow b}(g)|$, this implies that

$$\lim_{g \rightarrow \infty} \frac{\log |\Delta(g)|}{g} = \log \lambda(A') = C_d.$$

It remains to check that $\sup_{g \geq 2} (\log |\Delta(g)|) / g = \sup_{g \geq d+2} (\log |\Delta(g)|) / g$. This holds since for $d \geq 2$ and $2 \leq g < d+2$, $(\log |\Delta(g)|) / g \leq 0 \leq C_d$, and for $d=1$, it can be verified that $(\log |\Delta(2)|) / 2 = 1/2 < C_1$.

Part 5. Obviously the claim holds for $g=1$, so assume $g \geq 2$. In this case, let $\psi: \Delta_{0 \rightarrow 0}(g) \rightarrow \Delta_{1 \rightarrow 1}(g)$ be given by

$$\psi(0w0) = 10w, \quad 0w0 \in \Delta_{0 \rightarrow 0}(g), w \in \{0, 1\}^*.$$

It's easy to verify that for any $z \in \Delta_{0 \rightarrow 0}(g)$, $\psi(z) \in \Delta_{1 \rightarrow 1}(g)$ and that ψ is one-to-one and onto $\Delta_{1 \rightarrow 1}(g)$. This shows the claim.

Part 6. Note, that for $1 \leq g \leq d$, $\Delta_{0 \rightarrow 1}(g) = \emptyset$, and the claim holds. To show the claim for $g \geq d+2$, we define a map $\phi: \Delta_{0 \rightarrow 1}(g) \rightarrow \Delta_{0 \rightarrow 0}(g)$. For a word $x \in \Delta_{0 \rightarrow 1}(g)$, $\phi(x)$ is defined as follows:

1. If $x = w0^{d+1}1$, where $w \in \{0, 1\}^*$, then

$$\phi(x) = w0^d10.$$

2. if $x = w0^{d+1}10^d1$, where $w \in \{0, 1\}^*$, then

$$\phi(x) = w0^d10^d10.$$

3. if $x = w10^d1$, and w does not end with 0^{d+1} , then

$$\phi(x) = w0^d10.$$

For $i = 1, 2, 3$, let $\Lambda_i \subseteq \Delta_{0 \rightarrow 1}(g)$ denote the set of words satisfying the conditions of case i above. Observe, that $\{\Lambda_1, \Lambda_2, \Lambda_3\}$ is a partition of $\Delta_{0 \rightarrow 1}(g)$, and that in each case $\phi(x) \in \Delta_{0 \rightarrow 0}(g)$; thus ϕ is well-defined. We claim that ϕ

6.5. Proof of Theorem 13

is at most “two-to-one”. More precisely, we claim that there are no 3 distinct words $x, y, z \in \Delta_{0 \rightarrow 1}(g)$ such that $\phi(x) = \phi(y) = \phi(z)$; otherwise, since, clearly, ϕ restricted to Λ_i is one-to-one, each of x, y, z must belong to a different Λ_i , say $x \in \Delta_1, y \in \Delta_2$ and $z \in \Delta_3$; but then $\phi(y)$ ends with $10^d 10$ while $\phi(z)$ ends with $0^{d+1} 10$. Thus ϕ is at most two-to-one, and it follows that $|\Delta_{0 \rightarrow 1}(g)| \leq 2|\Delta_{0 \rightarrow 0}(g)|$.

Part 7. Let $g \neq d+1$. Then by parts 5 and 6,

$$\begin{aligned} |\Delta(g)| &\geq |\Delta_{0 \rightarrow 0}(g)| + |\Delta_{1 \rightarrow 1}(g)| + |\Delta_{0 \rightarrow 1}(g)| \\ &= 2|\Delta_{0 \rightarrow 0}(g)| + |\Delta_{0 \rightarrow 1}(g)| \\ &\geq 2|\Delta_{0 \rightarrow 1}(g)|, \end{aligned}$$

and the claim follows. The case $d = 1$ and $g = d+1 = 2$, is easily verified. \blacksquare

We will use the next lemma in the proof of Proposition 12, part 2. It gives a bound on $\lambda((N_i))$, where $(N_i) \subseteq \{A, C\}$ is a sequence of the form (6.26).

Lemma 9. *Let $s, t \geq 0$ and $g \geq 2$ be integers. Set $\lambda = \lambda((C)^2(A, C)^s(A)^g(C, A)^t)$. If $g = d+1$ and $d > 1$ then*

$$\lambda = (s+1)(t+1),$$

otherwise,

$$\lambda \leq |\Delta(g)| \left(\frac{s+1}{4} + \frac{t+1}{4} + \frac{(s+1)(t+1)}{2} \right).$$

Proof. Let $\mathbf{g} \in \mathbb{Z}^{s+t+1}$ be the block vector given by $\mathbf{g} = (\mathbf{1}_s | g | \mathbf{1}_t)$. For $a, b \in \{0, 1\}$ let $\Gamma_{a \rightarrow b} \subseteq \Delta(\mathbf{g})$ be given by

$$\Gamma_{a \rightarrow b} = \{(x_1, \dots, x_s, y, z_1, \dots, z_t) \in \Delta(\mathbf{g}) : y \text{ begins with } a \text{ and ends with } b\}.$$

Clearly, $\{\Gamma_{a,b} : a, b \in \{0, 1\}\}$ is a partition of $\Delta(\mathbf{g})$. On the other hand, consider the following identities, when $s, t > 0$, which are easy to verify:

$$\begin{aligned} |\Gamma_{0 \rightarrow 0}| &= |\Delta(\mathbf{1}_s)| |\Delta_{0 \rightarrow 0}(g)| = (s+1) |\Delta_{0 \rightarrow 0}(g)| \\ |\Gamma_{0 \rightarrow 1}| &= |\Delta(\mathbf{1}_s)| |\Delta_{0 \rightarrow 1}(g)| |\Delta(\mathbf{1}_t)| = (s+1)(t+1) |\Delta_{0 \rightarrow 1}(g)| \\ |\Gamma_{1 \rightarrow 0}| &= |\Delta_{1 \rightarrow 0}(g)| \\ |\Gamma_{1 \rightarrow 1}| &= |\Delta_{1 \rightarrow 1}(g)| |\Delta(\mathbf{1}_t)| = (t+1) |\Delta_{1 \rightarrow 1}(g)| \end{aligned}$$

where we used (6.35). Note that these identities hold even when $s = 0$ or $t = 0$. By Proposition 13 it follows that

$$\begin{aligned} \lambda = |\Delta(\mathbf{g})| &= \sum_{a,b \in \{0,1\}} |\Gamma_{a \rightarrow b}| \\ &= (s+1) |\Delta_{0 \rightarrow 0}(g)| + (s+1)(t+1) |\Delta_{0 \rightarrow 1}(g)| \\ &\quad + |\Delta_{1 \rightarrow 0}(g)| + (t+1) |\Delta_{1 \rightarrow 1}(g)|. \end{aligned} \tag{6.37}$$

Now, if $g=d+1$ and $d>1$, then $\Delta(g)=\{0^d1\}$ and the claim follows from (6.37). It remains to show the case $g\neq d+1$ or $d=1$. If $|\Delta(g)| = 0$, then the claim readily follows from (6.37). Otherwise, rewriting (6.37), we obtain

$$\begin{aligned} \lambda &= |\Delta(g)| \left((s+1) \frac{|\Delta_{0\rightarrow 0}(g)|}{|\Delta(g)|} + (t+1) \frac{|\Delta_{1\rightarrow 1}(g)|}{|\Delta(g)|} \right. \\ &\quad \left. + \frac{|\Delta_{1\rightarrow 0}(g)|}{|\Delta(g)|} + (s+1)(t+1) \frac{|\Delta_{0\rightarrow 1}(g)|}{|\Delta(g)|} \right) \\ &\leq |\Delta(g)| \left((s+1) \frac{|\Delta_{0\rightarrow 0}(g)| + |\Delta_{1\rightarrow 0}(g)|/2}{|\Delta(g)|} \right. \\ &\quad \left. + (t+1) \frac{|\Delta_{1\rightarrow 1}(g)| + |\Delta_{1\rightarrow 0}(g)|/2}{|\Delta(g)|} + (s+1)(t+1) \frac{|\Delta_{0\rightarrow 1}(g)|}{|\Delta(g)|} \right). \end{aligned} \tag{6.38}$$

Now, by Lemma 8, part 5 we have

$$\begin{aligned} \frac{|\Delta_{0\rightarrow 0}(g)| + |\Delta_{1\rightarrow 0}(g)|/2}{|\Delta(g)|} &= \frac{|\Delta_{1\rightarrow 1}(g)| + |\Delta_{1\rightarrow 0}(g)|/2}{|\Delta(g)|} \\ &= \frac{1}{2} \left(1 - \frac{|\Delta_{0\rightarrow 1}(g)|}{|\Delta(g)|} \right). \end{aligned}$$

Substituting this into (6.38) and applying part 7 of Lemma 8, we obtain

$$\begin{aligned} \lambda &\leq |\Delta(g)| \left(\frac{s+1}{2} + \frac{t+1}{2} + \frac{|\Delta_{0\rightarrow 1}(g)|}{|\Delta(g)|} \left((s+1)(t+1) - \frac{s+1}{2} - \frac{t+1}{2} \right) \right) \\ &\leq |\Delta(g)| \left(\frac{s+1}{4} + \frac{t+1}{4} + \frac{(s+1)(t+1)}{2} \right). \end{aligned}$$

This completes the proof. ■

Proof of Proposition 12, part 2. Set $\rho = \varrho((F_i))$. Then since $g \geq 2$,

$$\rho = \frac{t+s+2}{(t+s+2)(2d+2)+t+s+g} = \frac{1}{(2d+3)+(g-2)/(t+s+2)} < \frac{3}{6d+8}.$$

Hence, by Proposition 9, it follows that $h_d(\rho) \geq \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}(\rho)$. So, it's enough to show

$$\frac{\log \lambda((F_i))}{|(F_i)|} \leq \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}(\rho)$$

6.5. Proof of Theorem 13

Set $\lambda = \lambda((F_i))$. We need to show

$$\begin{aligned}
 & \frac{\log \lambda}{|(F_i)|} \leq \mathbb{L}_{\mathbf{p}_1, \mathbf{p}_2}(\rho) \\
 \Leftrightarrow & \frac{\log \lambda}{|(F_i)|} \leq \mathbb{L}_{(0, \mathcal{C}_d), (\frac{3}{6d+8}, \frac{\log 3}{6d+8})}(\rho) \\
 \Leftrightarrow & \frac{\log \lambda}{|(F_i)|} \leq \frac{\mathcal{C}_d - (\log 3)/(6d+8)}{-3/(6d+8)} \left(\rho - \frac{3}{6d+8} \right) + \frac{\log 3}{6d+8} \\
 \Leftrightarrow & \frac{\log \lambda}{|(F_i)|} - \frac{\log 3}{3} \rho \leq \mathcal{C}_d \left(1 - \frac{\rho}{3/(6d+8)} \right) \\
 \Leftrightarrow & \frac{\log \lambda - (t+s+2)(\log 3)/3}{(1 - \frac{\rho}{3/(6d+8)})|(F_i)|} \leq \mathcal{C}_d \\
 \Leftrightarrow & \frac{\log \lambda - (t+s+2)(\log 3)/3}{g + (t+s-4)/3} \leq \mathcal{C}_d \tag{6.39}
 \end{aligned}$$

Now, if $g = d+1$, and $d > 1$, then by Lemma 9, the numerator of the LHS of (6.39) is $\log(s+1) - (s+1)(\log 3)/3 + \log(t+1) - (t+1)(\log 3)/3$ which is non-positive for any nonnegative integers s, t . Since the denominator of the LHS of (6.39) is always positive, it follows that the LHS of (6.39) is nonpositive and (6.39) holds in this case. So we assume that $d=1$ or $g \neq d+1$. In this case, by Lemma 9,

$$\begin{aligned}
 \frac{\log \lambda - (t+s+2)\frac{\log 3}{3}}{g + \frac{t+s-4}{3}} & \leq \frac{\log |\Delta(g)| + \log\left(\frac{s+1}{4} + \frac{t+1}{4} + \frac{(s+1)(t+1)}{2}\right) - (t+s+2)\frac{\log 3}{3}}{g + \frac{t+s-4}{3}} \\
 & = \frac{\log |\Delta(g)| - \alpha}{g + \beta},
 \end{aligned}$$

where we set $\alpha = (t+s+2)\frac{\log 3}{3} - \log\left(\frac{s+1}{4} + \frac{t+1}{4} + \frac{(s+1)(t+1)}{2}\right)$ and $\beta = \frac{t+s-4}{3}$. Observe that

$$\begin{aligned}
 \alpha & \geq (t+s+2)\frac{\log 3}{3} - \log((s+1)(t+1)) \\
 & = (t+1)\frac{\log 3}{3} - \log(t+1) + (s+1)\frac{\log 3}{3} - \log(s+1) \\
 & \geq 0.
 \end{aligned}$$

6.6. Open questions

Now, if $t + s \geq 4$ then $\beta \geq 0$ and

$$\begin{aligned} \frac{\log \lambda - (t + s + 2) \frac{\log 3}{3}}{g + \frac{t+s-4}{3}} &\leq \frac{\log |\Delta(g)| - \alpha}{g + \beta} \\ &\leq \frac{\log |\Delta(g)|}{g} \\ &\leq \mathcal{C}_d, \end{aligned}$$

where the last inequality follows from Lemma 8, part 4. Hence (6.39) holds in this case. Otherwise, if $t + s < 4$ then $\beta < 0$ and it can be verified that for all such t, s , we have $-\alpha \leq \beta \mathcal{C}_1$. By Lemma 6 this implies that

$$-\alpha \leq \beta \mathcal{C}_d \tag{6.40}$$

Again, by Lemma 8, part 4, we have $(\log |\Delta(g)|)/g \leq \mathcal{C}_d$, which implies

$$\log |\Delta(g)| \leq g \mathcal{C}_d \tag{6.41}$$

Summing equations (6.41) and (6.40) and dividing by $g + \beta$ (which is positive), we get

$$\frac{\log |\Delta(g)| - \alpha}{g + \beta} \leq \mathcal{C}_d.$$

Therefore (6.39) holds in this case as well. ■

6.6 Open questions

By (6.2), the independence capacity of $\text{RLL}(d, k)$ remains $1/(2d+2)$ for all $2d + 1 \leq k \leq 3d + 1$. Is it possible to generalize the derivation in the proof of Theorem 13 to obtain the tradeoff function for $\text{RLL}(d, k)$ for this range of d and k ?

The “reverse-concatenation” encoding scheme described in Section 6.1 is used in practice for certain digital storage systems where the relevant constraint is $\text{RLL}(0, k)$. Knowing the tradeoff function for this constraint is thus especially important. Unfortunately, currently, this function is only known exactly when $k = 1, 2$ [5], and for $k = 3$ for insertion rates in $[0, 1/4]$ [37].

We so far restricted ourselves to dealing with tradeoff functions of 1-dimensional and binary constraints. It remains a task for future work to generalize the definition to higher-dimensional and non-binary constraints.

Chapter 7

Bounds on capacity using probability

In this chapter we give some probabilistic inequalities that hold for certain 2-dimensional binary constraints. Using these, we obtain a lower bound, stated in Theorem 16, on the capacity of certain constraints of the form $S^{(\mathcal{V})} \otimes \text{RLL}(0, 1)$, where $S^{(\mathcal{V})}$ is a 1-dimensional constraint over $\{0, 1\}$. We suspect that this bound is usually inferior to the bounds we get using the method described in Chapter 3: for example, for the (bit-flipped) hard square constraint, Theorem 16 gives $\text{cap}(\text{RLL}(0, 1)^{\otimes 2}) \geq 0.49445718\dots$, whereas with the method of Chapter 3 one gets $\text{cap}(\text{RLL}(0, 1)^{\otimes 2}) \geq 0.58789116\dots$. However, the arguments presented here do not seem to require symmetry of either the horizontal or vertical strips of the constraint, and thus may prove to be generalizable to constraints for which the method of Chapter 3 cannot be used.

For the rest of this chapter, we assume $\Sigma = \{0, 1\}$.

7.1 Some correlation inequalities

The method described in this chapter uses correlation inequalities: specifically the FKG and Holley inequality. We summarize them here. A partially ordered set (Λ, \leq) is called a *lattice*, if any two elements $x, y \in \Lambda$ have a smallest upper bound denoted $x \vee y$ (i.e. $x \vee y \geq x$ and $x \vee y \geq y$ and for all $z \in \Lambda$ s.t. $z \geq x$ and $z \geq y$, we have $z \geq x \vee y$) and a greatest lower bound denoted $x \wedge y$. A lattice is *distributive* if it satisfies either of the following two equivalent conditions:

$$\begin{aligned} x \vee (y \wedge z) &= (x \vee y) \wedge (x \vee z) && \text{for all } x, y, z \in \Lambda, \\ x \wedge (y \vee z) &= (x \wedge y) \vee (x \wedge z) && \text{for all } x, y, z \in \Lambda. \end{aligned}$$

As an example consider the set $\Lambda = \Sigma^{m \times n}$ for some $m, n \in \mathbb{N}$. As usual we say that for Δ_1, Δ_2 in Λ , $\Delta_1 \leq \Delta_2$ if the inequality holds entry-wise. Then \leq is a partial order on Λ and (Λ, \leq) is a finite distributive lattice. For two elements $\Delta_1, \Delta_2 \in \Sigma^m$, the arrays $\Delta_1 \vee \Delta_2$ (resp. $\Delta_1 \wedge \Delta_2$) is the array formed by taking an entry-wise maximum (resp. minimum) of Δ_1, Δ_2 . Every lattice that we use in this

chapter will essentially be such a lattice and in fact, it can be shown that every finite distributive lattice is isomorphic to a sub-lattice of $\Sigma^{m \times 1}$ for some m [31, Chapter 14, Theorem 15]. Henceforth, we will regard $\Sigma^{m \times n}$ as the lattice $(\Sigma^{m \times n}, \leq)$.

Let (Λ, \leq) be a finite distributive lattice. A function $f : \Lambda \rightarrow \mathbb{R}$ is *increasing* (i.e. non-decreasing) if for all $x, y \in \Lambda$ with $x \leq y$, $f(x) \leq f(y)$. We call a subset $A \subseteq \Lambda$ *increasing* if its indicator function is *increasing* and a 1-dimensional constraint $S \subseteq \Sigma^*$, *increasing*, if S_n is increasing (w.r.t the lattice Σ^n) for all $n \in \mathbb{N}$. The notions of a decreasing function, subset and 1-dimensional constraint are defined in an analogous manner. A function $f : \Lambda \rightarrow \mathbb{R}$ is *monotone* if it is increasing or decreasing.

Now, let $\mu : \mathbb{P}(\Lambda) \rightarrow [0, 1]$ be a probability measure on Λ , where $\mathbb{P}(\Lambda)$ denotes the power set of Λ . For $x \in \Lambda$ and such a measure μ , we abuse notation and write $\mu(x)$ to mean $\mu(\{x\})$. For any real function f on Λ we use $\mathbb{E}_\mu(f) = \mathbb{E}(f)$ to denote the expectation of f w.r.t μ , namely

$$\mathbb{E}_\mu(f) = \sum_{x \in \Lambda} f(x)\mu(x).$$

The following is known as the Holley inequality.

Theorem 14 (Holley inequality). *Let (Λ, \leq) be a finite distributive lattice, and $\mu_1, \mu_2 : \mathbb{P}(\Lambda) \rightarrow [0, 1]$ be two probability measures on Λ such that*

$$\mu_1(x \wedge y)\mu_2(x \vee y) \geq \mu_1(x)\mu_2(y) \quad \text{For all } x, y \in \Lambda. \quad (7.1)$$

Then for any increasing function $f : \Lambda \rightarrow \mathbb{R}$,

$$\mathbb{E}_{\mu_1}(f) \leq \mathbb{E}_{\mu_2}(f).$$

See [15] for a proof.

A simple corollary of this inequality is the following lemma which we use in the next section.

Lemma 10. *Let W_0, \dots, W_{m-1} be m random variables—each taking values in Σ —with two probability distributions μ_1, μ_2 such that W_0, \dots, W_{m-1} are independent w.r.t both μ_1 and μ_2 . Assume that for all $t \in [m]$, $\mu_1(W_t = 1) \leq \mu_2(W_t = 1)$. Then for any increasing set $A \subseteq \Sigma^m$*

$$\mu_1((W_0, \dots, W_{m-1}) \in A) \leq \mu_2((W_0, \dots, W_{m-1}) \in A).$$

Proof. It's easy to verify that (7.1) holds for μ_1 and μ_2 . The result follows by applying the Holley inequality to the indicator function of A . ■

7.2. Bounds on capacity using probability

We will also make use of the following inequality, known as the FKG inequality ([8]). A probability measure $\mu : \mathbb{P}(\Lambda) \rightarrow [0, 1]$ on Λ is called *log supermodular* if for all $x, y \in \Lambda$.

$$\mu(x \wedge y)\mu(x \vee y) \geq \mu(x)\mu(y). \quad (7.2)$$

Theorem 15 (FKG inequality). *Let Λ be a finite distributive lattice, and let $\mu : \mathbb{P}(\Lambda) \rightarrow [0, 1]$ be a log supermodular probability measure on Λ . Then for any monotone functions $f, g : \Lambda \rightarrow \mathbb{R}$ the following statements hold:*

1. *If f, g are both increasing or both decreasing then they are positively correlated, namely $\mathbb{E}(f)\mathbb{E}(g) \leq \mathbb{E}(fg)$*
2. *If f is increasing and g is decreasing or vice versa then they are negatively correlated, namely $\mathbb{E}(f)\mathbb{E}(g) \geq \mathbb{E}(fg)$*

It is shown in [15] that this inequality is an easy corollary of Theorem 14. As the proof is short, we reproduce it here for completeness.

Proof. Observe that it is enough to prove the theorem for the case where both f and g are increasing—the other cases follow from this case by taking $\tilde{f} = -f$ or $\tilde{g} = -g$, as appropriate. Now, by adding a large enough constant to f , we may assume without loss of generality that f is strictly positive. For a subset $A \subseteq \Lambda$, set $\mu_1(A) = \mu(A)$ and $\mu_2(A) = \sum_{x \in A} (f(x)\mu(x)) / \mathbb{E}_\mu(f)$. Then it's easy to check that (7.1) is satisfied and therefore since g is increasing,

$$\begin{aligned} \mathbb{E}_{\mu_1}(g) &\leq \mathbb{E}_{\mu_2}(g) \\ \iff \mathbb{E}_\mu(g) &\leq \frac{\sum_{x \in \Lambda} f(x)\mu(x)g(x)}{\mathbb{E}_\mu(f)} \\ \iff \mathbb{E}_\mu(g)\mathbb{E}_\mu(f) &\leq \mathbb{E}_\mu(fg). \end{aligned}$$

■

7.2 Bounds on capacity using probability

Our goal in this section is to prove the following lower bound on the capacity of certain 2-dimensional constraints.

Theorem 16. *Let $S = S^{(\mathcal{V})} \otimes S^{(\mathcal{H})}$ be a 2-dimensional constraint with $S^{(\mathcal{H})} = \text{RLL}(0, 1)$ and $S^{(\mathcal{V})}$ an increasing 1-dimensional constraint over Σ . Let $\varphi = \frac{1+\sqrt{5}}{2}$ denote the golden mean and set $\mu_\infty(0) = \frac{1}{\varphi^2+1}$ and $\mu_\infty(1) = \frac{\varphi^2}{\varphi^2+1}$. Let $\mathcal{G}^{(\mathcal{V})} = ((V^{(\mathcal{V})}, E^{(\mathcal{V})}), \mathcal{L}^{(\mathcal{V})})$ be a lossless presentation of $S^{(\mathcal{V})}$, and $\mathcal{W}^{(\mathcal{V})} : E^{(\mathcal{V})} \rightarrow [0, 1]$*

7.2. Bounds on capacity using probability

be the edge-weighting function given by $\mathcal{W}^{(\mathcal{V})}(e) = \mu_\infty(\mathcal{L}^{(\mathcal{V})}(e))$. Let A be the $|V^{(\mathcal{V})}| \times |V^{(\mathcal{V})}|$ real matrix with entries indexed by $V^{(\mathcal{V})} \times V^{(\mathcal{V})}$ and given by

$$(A)_{(i,j)} = \sum_{\substack{e \in E^{(\mathcal{V})}, \\ \sigma(e)=i, \tau(e)=j}} \mathcal{W}^{(\mathcal{V})}(e).$$

Then

$$\text{cap}(S) \geq \log \varphi - \frac{1}{2} \left(1 - \text{cap}(S^{(\mathcal{V})}) - \log \lambda(A) \right),$$

Remark 1. If $S = \text{RLL}(1, \infty) \otimes S^{(\mathcal{V})}$ with $S^{(\mathcal{V})}$ a decreasing 1-dimensional constraint over Σ , then simultaneously changing ‘0’s to ‘1’s and ‘1’s to ‘0’s in every array of S would result in a constraint that satisfies the requirement of the above theorem and has the same capacity as S . Thus, we can also use the theorem to get a lower bound on $\text{cap}(S)$.

Proof. For a probability space $(\Omega, \mathbb{F}, \mu)$, and events $A, B \subseteq \Omega$ with $\mu(A) > 0$, we denote by $\mu(\cdot|A) : \mathbb{F} \rightarrow [0, 1]$ the conditional probability measure given by $\mu(B|A) = \mu(B \cap A)/\mu(A)$ for all $B \in \mathbb{F}$. For a word $w \in \Sigma^*$ of length m , $m \in \mathbb{N}$, we index its symbols by $[m]$ so that $w = w_0 w_1 \dots w_{m-1}$. For $m, n \in \mathbb{N}$, let $\mu_{m \times n} : \mathbb{P}(\Sigma^{m \times n}) \rightarrow [0, 1]$ be the uniform probability measure on $\Sigma^{m \times n}$. Clearly, $\mu_{m \times n}$ satisfies (7.2) with equality. For a subset $A \subseteq \Sigma^{m \times n}$, let $1_A : \Sigma^{m \times n} \rightarrow \{0, 1\}$ be the indicator function of A , and for integers $i \in [m]$ and $j \in [n]$, define the sets:

$$R_i = R_i^{(m,n)} = \{\Delta \in \Sigma^{m \times n} : \Delta_{i,*} \text{ satisfies } S^{(\mathcal{H})}\},$$

$$C_j = C_j^{(m,n)} = \{\Delta \in \Sigma^{m \times n} : \Delta_{*,j} \text{ satisfies } S^{(\mathcal{V})}\},$$

$$R = R^{(m,n)} = \bigcap_{i \in [m]} R_i^{(m,n)},$$

$$C_{\text{even}} = C_{\text{even}}^{(m,n)} = \bigcap_{\substack{j \in [n], \\ j \text{ even}}} C_j^{(m,n)}, \quad \text{and}$$

$$C_{\text{odd}} = C_{\text{odd}}^{(m,n)} = \bigcap_{\substack{j \in [n], \\ j \text{ odd}}} C_j^{(m,n)}.$$

Then since $S^{(\mathcal{H})}$ and $S^{(\mathcal{V})}$ are both increasing, it follows that, $1_{R \cap C_{\text{odd}}}$ and $1_{C_{\text{even}}}$

7.2. Bounds on capacity using probability

are increasing, and consequently, by the FKG inequality,

$$\begin{aligned}
 \mu_{m \times n}(S_{m \times n}) &= \mu_{m \times n}(R \cap C_{\text{odd}} \cap C_{\text{even}}) \\
 &= \mathbb{E}_{\mu_{m \times n}}(1_{R \cap C_{\text{odd}}} \cdot 1_{C_{\text{even}}}) \\
 &\geq \mathbb{E}_{\mu_{m \times n}}(1_{R \cap C_{\text{odd}}}) \mathbb{E}_{\mu_{m \times n}}(1_{C_{\text{even}}}) \\
 &= \mu_{m \times n}(R \cap C_{\text{odd}}) \mu_{m \times n}(C_{\text{even}}).
 \end{aligned}$$

It follows that

$$\begin{aligned}
 \text{cap}(S) &= \lim_{(m,n) \rightarrow \infty} \frac{\log |S_{m \times n}|}{mn} = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\log(2^{mn} \mu_{m \times n}(S_{m \times n}))}{mn} \\
 &\geq \limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{\log(2^{mn} \mu_{m \times n}(C_{\text{even}}) \mu_{m \times n}(R \cap C_{\text{odd}}))}{mn} \\
 &= \limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{\log(\mu_{m \times n}(C_{\text{even}}) \cdot 2^{mn} \cdot \mu_{m \times n}(R) \mu_{m \times n}(C_{\text{odd}} | R))}{mn} \\
 &= \limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \left(\frac{\log \left(\left(2^{-m} |S_m^{(\mathcal{V})}| \right)^{\lceil n/2 \rceil} |S_n^{(\mathcal{H})}|^m \right)}{mn} + \right. \\
 &\quad \left. \frac{\log \prod_{i=1}^{\lceil n/2 \rceil} \mu_{m \times n}(C_{2i-1} | R \cap \bigcap_{j=1}^{i-1} C_{2j-1})}{mn} \right) \\
 &= \text{cap}(S^{(\mathcal{H})}) + \frac{1}{2} \left(\text{cap}(S^{(\mathcal{V})}) - 1 \right) + \\
 &\quad \limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{\sum_{i=1}^{\lceil n/2 \rceil} \log \mu_{m \times n}(C_{2i-1} | R \cap \bigcap_{j=1}^{i-1} C_{2j-1})}{mn}. \quad (7.3)
 \end{aligned}$$

We claim that

$$\mu_{m \times n}(C_{2i-1} | R \cap \bigcap_{j=1}^{i-1} C_{2j-1}) \geq \mu_{m \times n}(C_{2i-1} | R), \quad (7.4)$$

for all $m, n \in \mathbb{N}$ and integer $1 \leq i \leq \lceil n/2 \rceil$. This follows from the right inequality given in the next Proposition.

Proposition 14. *For all $m, n \in \mathbb{N}$ and $i \in [n]$*

$$\mu_{m \times n}(C_i | R \cap \bigcap_{\substack{1 \leq j \leq i, \\ j \text{ odd}}} C_{i-j}) \leq \mu_{m \times n}(C_i | R) \leq \mu_{m \times n}(C_i | R \cap \bigcap_{\substack{1 \leq j \leq i, \\ j \text{ even}}} C_{i-j}).$$

7.2. Bounds on capacity using probability

The proof of Proposition 14 is given in Section 7.2.1. Using (7.4) in (7.3) and the well-known fact (c.f. [32]) that $\text{cap}(S^{(\mathcal{H})}) = \log \varphi$, we obtain

$$\begin{aligned} \text{cap}(S) &\geq \text{cap}(S^{(\mathcal{H})}) + \frac{1}{2} \left(\text{cap}(S^{(\mathcal{V})}) - 1 \right) + \\ &\quad \limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} \log \mu_{m \times n}(C_{2i-1} | R)}{mn} \\ &= \log \varphi - \frac{1}{2} (1 - \text{cap}(S^{(\mathcal{V})})) + \\ &\quad \limsup_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} \log \mu_{m \times n}(C_{2i-1} | R)}{mn} \end{aligned}$$

The theorem now follows from the next lemma whose proof is given in Section 7.2.2

Lemma 11. *The following statements hold:*

1. For all $m \in \mathbb{N}$, $\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} \log \mu_{m \times n}(C_{2i-1} | R)}{n} = \frac{1}{2} \log \sum_{w \in S_m^{(\mathcal{V})}} \prod_{j \in [m]} \mu_\infty(w_j)$.
2. $\lim_{m \rightarrow \infty} \frac{\log \sum_{w \in S_m^{(\mathcal{V})}} \prod_{j \in [m]} \mu_\infty(w_j)}{m} = \log \lambda(A)$.

7.2.1 Proof of Proposition 14

Proof. We prove the right inequality since it is the only one we use in the proof of Theorem 16 and the proof of the left inequality is similar. Let $m, n \in \mathbb{N}$ and $i \in [n]$. We again use the FKG inequality. Set $A = \{i - j : j \in \mathbb{N}, 1 \leq j \leq i, j \text{ is even}\}$, $\mathcal{S} = [m] \times A$ and let $\Lambda' = \Sigma^{\mathcal{S}}$, be the set of all binary-valued functions with domain \mathcal{S} . We think of a function in Λ' as a binary non-contiguous configuration on the ‘‘sites’’ of \mathcal{S} . For such a configuration $\Delta' \in \Lambda'$ and $j \in A$, we write $\Delta'_{*,j}$ to denote the 1-dimensional word $\Delta'(0, j) \Delta'(1, j) \dots \Delta'(m-1, j)$. As usual for $\Delta'_1, \Delta'_2 \in \Lambda'$ we write $\Delta'_1 \leq \Delta'_2$ if $\Delta'_1(\mathbf{j}) \leq \Delta'_2(\mathbf{j})$ for all $\mathbf{j} \in \mathcal{S}$. Then Λ' with this ordering is a finite distributive lattice. For a binary array $\Delta \in \Sigma^{m \times n}$, denote by $\pi(\Delta) \in \Lambda'$ the restriction of Δ to \mathcal{S} , namely the configuration $\Delta' \in \Lambda'$ given by $\Delta'(\mathbf{j}) = \Delta_{\mathbf{j}}$ for all $\mathbf{j} \in \mathcal{S}$. For a set $B \subseteq \Lambda'$, let $\pi^{-1}(B)$ denote the inverse image of B under π . Set $\mu = \mu_{m \times n}$ and let $\mu' : \mathbb{P}(\Lambda') \rightarrow [0, 1]$ be the push-forward probability measure of $\mu(\cdot | R)$, defined by $\mu'(B) = \mu(\pi^{-1}(B) | R)$. For $j \in A$, let $C'_j = \{\Delta' \in \Lambda' : \Delta'_{*,j} \in S^{(\mathcal{V})}\}$ and set $C' = \bigcap_{j \in A} C'_j$. Then $\pi^{-1}(C'_j) = C_j$ and $\pi^{-1}(C') = \bigcap_{j \in A} C_j$. Finally,

7.2. Bounds on capacity using probability

let $f, g : \Lambda' \rightarrow \mathbb{R}$ be the functions given by

$$f(\Delta') = 1_{C'}(\Delta') = \begin{cases} 1 & \text{If } \Delta' \in C'. \\ 0 & \text{otherwise.} \end{cases}, \quad \text{and}$$

$$g(\Delta') = \mu(C_i | R \cap \pi^{-1}(\{\Delta'\})),$$

for $\Delta' \in \Lambda'$. We claim that the following statements hold.

- μ' is log supermodular.
- f is increasing.
- g is increasing.

Assuming that these do hold, then by the FKG inequality, we have

$$\begin{aligned} \sum_{\Delta' \in \Lambda'} (f(\Delta')g(\Delta')\mu'(\Delta')) &\geq \sum_{\Delta' \in \Lambda'} (f(\Delta')\mu'(\Delta')) \sum_{\Delta' \in \Lambda'} (g(\Delta')\mu'(\Delta')) \\ \iff \sum_{\Delta' \in C'} (\mu(C_i | R \cap \pi^{-1}(\{\Delta'\}))\mu(\pi^{-1}(\{\Delta'\}) | R)) &\geq \\ \sum_{\Delta' \in C'} \mu(\pi^{-1}(\{\Delta'\}) | R) \sum_{\Delta' \in \Lambda'} (\mu(C_i | R \cap \pi^{-1}(\{\Delta'\}))\mu(\pi^{-1}(\{\Delta'\}) | R)) & \\ \iff \sum_{\Delta' \in C'} \mu(C_i \cap \pi^{-1}(\{\Delta'\}) | R) \geq & \\ \sum_{\Delta' \in C'} \mu(\pi^{-1}(\{\Delta'\}) | R) \sum_{\Delta' \in \Lambda'} \mu(C_i \cap \pi^{-1}(\{\Delta'\}) | R) & \\ \iff \mu(C_i \cap \bigcap_{j \in A} C_j | R) \geq \mu(\bigcap_{j \in A} C_j | R)\mu(C_i | R) & \\ \iff \mu(C_i | R \cap \bigcap_{j \in A} C_j) \geq \mu(C_i | R), & \end{aligned}$$

which is the desired inequality. So it remains to show that the above 3 claims hold.

We begin by showing that μ' is log supermodular. Let $\Delta' \in \Lambda'$. Then

$$\begin{aligned} \mu'(\Delta') &= \frac{\mu(\pi^{-1}(\{\Delta'\}) \cap R)}{\mu(R)} \\ &= \frac{2^{-mn} |\{\Delta \in \Sigma^{m \times n} : \forall t \in [m] \forall j \in A (\Delta_{t,j} = \Delta'(t,j) \text{ and } \Delta_{t,*} \in S^{(\mathcal{H})})\}|}{2^{-mn} |R|} \\ &= \prod_{t \in [m]} \frac{|\{w \in S_n^{(\mathcal{H})} : \forall j \in A w_j = \Delta'(t,j)\}|}{|S_n^{(\mathcal{H})}|} \end{aligned} \tag{7.5}$$

7.2. Bounds on capacity using probability

Let $j_0 < j_1 < \dots < j_{\ell-1} = i-2$ be the elements of A and for symbols $a, b \in \Sigma$ and integer $u \geq 0$, define $\alpha_u(a) = |\{w \in S_u^{(\mathcal{H})} : wa \in S^{(\mathcal{H})}\}|$, $\delta(a, b) = |\{x \in \Sigma : axb \in S^{(\mathcal{H})}\}|$ and $\omega_u(a) = |\{w \in S_u^{(\mathcal{H})} : aw \in S^{(\mathcal{H})}\}|$. Since the memory of $S^{(\mathcal{H})} = \text{RLL}(0, 1)$ is 1, it readily follows that

$$\left| \left\{ w \in S_n^{(\mathcal{H})} : \forall k \in [\ell] w_{j_k} = a_k \right\} \right| = \alpha_{j_0}(a_0) \left(\prod_{k \in [\ell-1]} \delta(a_k, a_{k+1}) \right) \omega_{n-i+1}(a_{\ell-1}).$$

Setting $r = n - i + 1$, we can rewrite (7.5) as

$$\mu'(\Delta') = \prod_{t \in [m]} \frac{\alpha_{j_0}(\Delta'(t, j_0)) \left(\prod_{k=0}^{\ell-2} \delta(\Delta'(t, j_k), \Delta'(t, j_{k+1})) \right) \omega_r(\Delta'(t, j_{\ell-1}))}{|S_n^{(\mathcal{H})}|}.$$

Now, pick $\Delta'_1, \Delta'_2 \in \Lambda'$, and set $\Gamma'_1 = \Delta'_1 \wedge \Delta'_2$ and $\Gamma'_2 = \Delta'_1 \vee \Delta'_2$.

$$\begin{aligned} \mu'(\Delta_1)\mu'(\Delta_2) &= |S_n^{(\mathcal{H})}|^{-2m} \left(\prod_{t \in [m]} \prod_{s=1}^2 \alpha_{j_0}(\Delta'_s(t, j_0)) \right) \\ &\quad \left(\prod_{t \in [m]} \prod_{k \in [\ell-1]} \prod_{s=1}^2 \delta(\Delta'_s(t, j_k), \Delta'_s(t, j_{k+1})) \right) \\ &\quad \left(\prod_{t \in [m]} \prod_{s=1}^2 \omega_r(\Delta'_s(t, j_{\ell-1})) \right) \end{aligned}$$

For all $a, b, c, d \in \Sigma$, it's easy to verify that $\delta(a, b)\delta(c, d) \leq \delta(a \wedge c, b \wedge d)\delta(a \vee c, b \vee d)$, and obviously $\alpha_{j_0}(a)\alpha_{j_0}(b) = \alpha_{j_0}(a \wedge b)\alpha_{j_0}(a \vee b)$ and $\omega_r(a)\omega_r(b) = \omega_r(a \wedge b)\omega_r(a \vee b)$. Thus,

$$\begin{aligned} \mu'(\Delta_1)\mu'(\Delta_2) &\leq |S_n^{(\mathcal{H})}|^{-2m} \left(\prod_{t \in [m]} \prod_{s=1}^2 \alpha_{j_0}(\Gamma'_s(t, j_0)) \right) \\ &\quad \left(\prod_{t \in [m]} \prod_{k \in [\ell-1]} \prod_{s=1}^2 \delta(\Gamma'_s(t, j_k), \Gamma'_s(t, j_{k+1})) \right) \\ &\quad \left(\prod_{t \in [m]} \prod_{s=1}^2 \omega_r(\Gamma'_s(t, j_{\ell-1})) \right) \\ &= \mu'(\Gamma'_1)\mu'(\Gamma'_2). \end{aligned}$$

7.2. Bounds on capacity using probability

Hence μ' is log supermodular.

The second claim that f is increasing follows immediately from the fact the $S^{(\mathcal{V})}$ is increasing. So it remains to show that g is increasing (this is where the proof of the left inequality of the proposition differs). Let $\Delta' \in \Lambda'$. Then

$$\begin{aligned}
 g(\Delta') &= \mu(C_i | R \cap \pi^{-1}(\Delta')) = \frac{\mu(C_i \cap R \cap \pi^{-1}(\Delta'))}{\mu(R \cap \pi^{-1}(\Delta'))} \\
 &= \frac{2^{-mn} \sum_{w \in S_m^{(\mathcal{V})}} \left| \left\{ \Delta \in \Sigma^{m \times n} : \begin{array}{l} \Delta_{*,i} = w, \forall t \in [m] \Delta_{t,*} \in S^{(\mathcal{H})}, \\ \forall t \in [m] \forall j \in A \Delta_{t,j} = \Delta'(t,j) \end{array} \right\} \right|}{2^{-mn} \left| \left\{ \Delta \in \Sigma^{m \times n} : \forall t \in [m] \forall j \in A (\Delta_{t,j} = \Delta'(t,j) \text{ and } \Delta_{t,*} \in S^{(\mathcal{H})}) \right\} \right|} \\
 &= \frac{\sum_{w \in S_m^{(\mathcal{V})}} \prod_{t \in [m]} \left| \left\{ v \in S_n^{(\mathcal{H})} : v_i = w_t, \forall j \in A v_j = \Delta'(t,j) \right\} \right|}{\prod_{t \in [m]} \left| \left\{ v \in S_n^{(\mathcal{H})} : \forall j \in A v_j = \Delta'(t,j) \right\} \right|}, \quad (7.6)
 \end{aligned}$$

For $t \in [m]$, let $q_t = |\{v \in S_n^{(\mathcal{H})} : \forall j \in A v_j = \Delta'(t,j)\}|$; then since $S^{(\mathcal{H})}$ has memory 1, it follows that for all $t \in [m]$, and $a \in \Sigma$ we have

$$\begin{aligned}
 \left| \left\{ v \in S_n^{(\mathcal{H})} : v_i = a, \forall j \in A v_j = \Delta'(t,j) \right\} \right| &= q_t \delta(\Delta'(t, i-2), a) \omega_{n-i-1}(a), \text{ and} \\
 \left| \left\{ v \in S_n^{(\mathcal{H})} : \forall j \in A v_j = \Delta'(t,j) \right\} \right| &= q_t \omega_{n-i+1}(\Delta'(t, i-2))
 \end{aligned}$$

Substituting this into (7.6), we obtain

$$g(\Delta') = \sum_{w \in S_m^{(\mathcal{V})}} \prod_{t \in [m]} \frac{\delta(\Delta'(t, i-2), w_t) \omega_{n-i-1}(w_t)}{\omega_{n-i+1}(\Delta'(t, i-2))} = \sum_{w \in S_m^{(\mathcal{V})}} \prod_{t \in [m]} \psi_{\Delta'(t, i-2)}(w_t),$$

where $\psi_a(b) = \frac{\delta(a,b) \omega_{n-i-1}(b)}{\omega_{n-i+1}(a)}$ for all $a, b \in \Sigma$. Observe that $\psi_a(0) + \psi_a(1) = 1$ and that $\psi_a(0), \psi_a(1) \in [0, 1]$ for all $a \in \Sigma$. It follows that $g(\Delta')$ is the probability that a randomly selected word consisting of m independent random bits, with the t th bit having a probability of $\psi_{\Delta'(t, i-2)}(1)$ to be 1, satisfies $S^{(\mathcal{V})}$. Now, let $\Delta'_1, \Delta'_2 \in \Lambda'$ satisfy $\Delta'_1 \leq \Delta'_2$. We claim that for each $t \in [m]$ $\psi_{\Delta'_1(t, i-2)}(1) \leq \psi_{\Delta'_2(t, i-2)}(1)$. Indeed, let $t \in [m]$ be such that $\Delta'_1(t, i-2) = 0$ and $\Delta'_2(t, i-2) = 1$. Then

$$\begin{aligned}
 \psi_{\Delta'_1(t, i-2)}(1) &= \frac{\delta(0, 1) \omega_{n-i-1}(1)}{\omega_{n-i+1}(0)} = \frac{\omega_{n-i-1}(1)}{\omega_{n-i}(1)}, \quad \text{and} \\
 \psi_{\Delta'_2(t, i-2)}(1) &= \frac{\delta(1, 1) \omega_{n-i-1}(1)}{\omega_{n-i+1}(1)} = \frac{2\omega_{n-i-1}(1)}{\omega_{n-i+1}(1)},
 \end{aligned}$$

and since $\omega_{n-i+1}(1) \leq 2|S_{n-i}^{(\mathcal{H})}| = 2\omega_{n-i}(1)$ it follows that $\psi_{\Delta'_1(t, i-2)}(1) \leq \psi_{\Delta'_2(t, i-2)}(1)$. As $S^{(\mathcal{V})}$ is increasing, Lemma 10 implies that $g(\Delta'_1) \leq g(\Delta'_2)$ and therefore g is increasing as claimed. \blacksquare

7.2.2 Proof of Lemma 11

Fix $m \in \mathbb{N}$. Observe that for any $n_1, n_2 \in \mathbb{N}$ we have

$$\begin{aligned}
 \mu_{m \times (n_1 + n_2 + 1)}(C_{n_1} | R) &= \frac{|C_{n_1} \cap R| 2^{-mn}}{|R| 2^{-mn}} \\
 &= \sum_{w \in S_m^{(\mathcal{V})}} \frac{\left| \left\{ \Delta \in \Sigma^{m \times (n_1 + n_2 + 1)} : \begin{array}{l} \Delta_{*, n_1} = w, \\ \forall t \in [m] \Delta_{t, *}, * \in S^{(\mathcal{H})} \end{array} \right\} \right|}{|R|} \\
 &= \sum_{w \in S_m^{(\mathcal{V})}} \prod_{j \in [m]} \frac{\left| \left\{ v \in S_n^{(\mathcal{H})} : v_{n_1} = w_j \right\} \right|}{\left| S_{n_1 + n_2 + 1}^{(\mathcal{H})} \right|} \\
 &= \sum_{w \in S_m^{(\mathcal{V})}} \prod_{j \in [m]} \mu_{n_1, n_2}(w_j), \tag{7.7}
 \end{aligned}$$

where we define $\mu_{n_1, n_2} : \Sigma \rightarrow [0, 1]$ to be the function given by $\mu_{n_1, n_2}(a) = \frac{|\{v \in S_{n_1 + n_2 + 1}^{(\mathcal{H})} : v_{n_1} = a\}|}{|S_{n_1 + n_2 + 1}^{(\mathcal{H})}|}$ for $a \in \Sigma$. Since $S^{(\mathcal{H})}$ is increasing, for any $v \in S_{n_1 + n_2 + 1}^{(\mathcal{H})}$ with $v_{n_1} = 0$, the word v' formed from v by changing the n_1 'th symbol to '1' is also in $S^{(\mathcal{H})}$. It follows that for all $n_1, n_2 \in \mathbb{N}$, $\mu_{n_1, n_2}(1) \geq \frac{1}{2}$, and thus by (7.7), one has

$$\mu_{m \times (n_1 + n_2 + 1)}(C_{n_1} | R) \geq (\mu_{n_1, n_2}(1))^m \geq 2^{-m}. \tag{7.8}$$

We next show that for all $a \in \Sigma$

$$\lim_{(n_1, n_2) \rightarrow \infty} \mu_{n_1, n_2}(a) = \mu_\infty(a). \tag{7.9}$$

In fact it can be shown that the above limit exists (and equals a different μ_∞) for all vertex constraints $S^{(\mathcal{H})}$ defined by a primitive graph. Let $f_n = |(\text{RLL}(0, 1))_n|$. It is well known (c.f. [32]) that f_n is the $(n + 2)$ Fibonacci number and is given by

$$f_n = \varphi^n \frac{\varphi^3}{1 + \varphi^2} + \bar{\varphi}^n \frac{\bar{\varphi}^3}{1 + \bar{\varphi}^2}, \tag{7.10}$$

where $\bar{\varphi} = 1 - \varphi$. Since $\sum_{a \in \Sigma} \mu_{n_1, n_2}(a) = \sum_{a \in \Sigma} \mu_\infty(a) = 1$, it's enough to

7.2. Bounds on capacity using probability

show that (7.9) holds for $a = 1$. In this case, clearly,

$$\begin{aligned}\mu_{n_1, n_2}(1) &= \frac{|\{v \in S_{n_1+n_2+1}^{(\mathcal{H})} : v_{n_1} = 1\}|}{f_{n_1+n_2+1}} \\ &= \frac{|\{x1y \in \Sigma^{n_1+n_2+1} : x \in S_{n_1}^{(\mathcal{H})}, y \in S_{n_2}^{(\mathcal{H})}\}|}{f_{n_1+n_2+1}} \\ &= \frac{f_{n_1} f_{n_2}}{f_{n_1+n_2+1}}.\end{aligned}$$

By substituting (7.10) into the last equality and taking the limit as $(n_1, n_2) \rightarrow \infty$ it can be verified that (7.9) holds for $a = 1$.

Now, set $L = \log \sum_{w \in S_m^{(\mathcal{V})}} \prod_{j \in [m]} \mu_\infty(w_j)$. It follows from (7.7) and (7.9) that $\lim_{(n_1, n_2) \rightarrow \infty} \log \mu_{m \times (n_1+n_2+1)}(C_{n_1} | R) = L$. Let $\epsilon > 0$, and choose $N \in \mathbb{N}$ such that for all $n_1, n_2 \geq N$, $|\log \mu_{m \times (n_1+n_2+1)}(C_{n_1} | R) - L| < \epsilon$. For $n \in \mathbb{N}$ set $A_n = \{j \in [n] : j \text{ odd}\}$, $A_{n,1} = \{j \in A_n : j < N \text{ or } j > n - 1 - N\}$ and $A_{n,2} = A_n \setminus A_{n,1}$. Then, for all $n \in \mathbb{N}$ with $n > 2N$, we have

$$\begin{aligned}\left| \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} \log \mu_{m \times n}(C_{2i-1} | R)}{\lfloor n/2 \rfloor} - L \right| &\leq \left| \frac{\sum_{i \in A_{n,1}} \log \mu_{m \times n}(C_i | R)}{|A_{n,1}|} \right| + \\ &\quad \left| \frac{\sum_{i \in A_{n,2}} (\log \mu_{m \times n}(C_i | R) - L) - |A_{n,1}|L}{|A_n|} \right|.\end{aligned}$$

Note that for all $i \in A_{n,2}$, it holds that $i \geq N$ and $n - i - 1 \geq N$; thus $|\log \mu_{m \times n}(C_i | R) - L| < \epsilon$. Also by (7.8), it follows that for all $i \in A_{n,1}$, $|\log \mu_{m \times n}(C_i | R)| \leq m$. Therefore we obtain

$$\begin{aligned}\left| \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} \log \mu_{m \times n}(C_{2i-1} | R)}{\lfloor n/2 \rfloor} - L \right| &\leq \frac{|A_{n,1}|}{|A_n|} m + \frac{\epsilon |A_{n,2}|}{|A_n|} + \frac{|A_{n,1}|}{|A_n|} L \\ &\leq \frac{2N(m+L)}{\lfloor n/2 \rfloor} + \epsilon,\end{aligned}$$

which is less than 2ϵ for large enough n . It follows that

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^{\lfloor n/2 \rfloor} \log \mu_{m \times n}(C_{2i-1} | R)}{\lfloor n/2 \rfloor} = \log \sum_{w \in S_m^{(\mathcal{V})}} \prod_{j \in [m]} \mu_\infty(w_j),$$

which obviously implies the first statement of the lemma.

7.3. Open questions

As for the second statement, for a path in $\mathcal{G}^{(\mathcal{V})}$, define its weight to be the product of the weights of its edges. Denote by \mathcal{W}_m the sum of all the weights of paths of length m in $\mathcal{G}^{(\mathcal{V})}$. Since $\mathcal{G}^{(\mathcal{V})}$ is a lossless presentation of $S^{(\mathcal{V})}$ we have

$$\frac{\mathcal{W}_m}{|V^{(\mathcal{V})}|^2} \leq \sum_{w \in S_m^{(\mathcal{V})}} \prod_{j \in [m]} \mu_\infty(w_j) \leq \mathcal{W}_m.$$

On the other hand, $\mathcal{W}_m = \mathbf{1}^\top A^m \mathbf{1}$ and, applying Perron-Frobenius theory, we have $\lim_{m \rightarrow \infty} \frac{\log \mathcal{W}_m}{m} = \log(\lambda(A))$. The second statement follows. ■

7.3 Open questions

Can Theorem 16 be generalized to other constraints $S^{(\mathcal{H})}$? Can one use similar probabilistic tools to obtain upper bounds on the capacity of axial products of monotone 1-dimensional constraints? Finally, can such tools be used to obtain bounds on the capacity of constraints in more than 2 dimensions?

Bibliography

- [1] R.J Baxter, *Hard hexagons: exact solution*, Journal of Physics A: Mathematical and General **13** (1980), L61–L70.
- [2] R. Berger, *The undecidability of the domino problem*, Memoirs of the American Mathematical Society (1966).
- [3] N.J. Calkin and H.S. Wilf, *The number of independent sets in a grid graph*, SIAM Journal of Discrete Math. **11** (1998), no. 1, 54–60.
- [4] M. Cohn, *On the channel capacity of read/write isolated memory*, Discrete Applied Mathematics **56** (1995).
- [5] J.C. de Souza, B.H. Marcus, R. New, and Wilson B.A., *Constrained systems with unconstrained positions*, IEEE Transactions on Information Theory **48** (2002), 866–879.
- [6] A. Desai, *Subsystem entropy for \mathbb{Z}^d sofic shifts*, Indagationes Mathematicae **17** (2006).
- [7] K. Engel, *On the Fibonacci number of an $m \times n$ lattice*, Fibonacci Quarterly **28** (1990), 72–78.
- [8] C.M. Fortuin, P.W. Kasteleyn, and J. Ginibre, *Correlation inequalities on some partially ordered sets*, Comm. Math. Phys. **22** (1971), no. 2, 89–103.
- [9] S. Friedland, *On the entropy of \mathbb{Z}^d subshifts of finite type*, Linear Algebra and its Applications **252** (1997), no. 1–3, 199–220.
- [10] S. Friedland, *Multi-dimensional capacity, pressure and Hausdorff dimension*, Mathematical System Theory in Biology, Communication, Computation and Finance (2003), 183–222.
- [11] S. Friedland and U.N. Peled, *The pressure, densities and first order phase transitions associated with multidimensional SOFT*, arXiv:0906.5176v3, <http://arxiv.org/abs/0906.5176v3>, submitted, 2009.

- [12] D.S. Gaunt and M.E. Fisher, *Hard-sphere lattice gases, I. plane-square lattice*, J. Chem. Phys. **43** (1965), 2840–2863.
- [13] M.J. Golin, X. Yong, Y. Zhang, and L. Sheng, *New upper and lower bounds on the channel capacity of read/write isolated memory*, Discrete Applied Mathematics **140** (2004).
- [14] S. Halevy and R. Roth, *Parallel constrained coding with application to two-dimensional constraints*, IEEE Transactions on Information Theory **48** (2002), 1009–1020.
- [15] R.H. Holley, *Remarks on the FKG inequalities*, Comm. Math. Phys. **36** (1974), no. 3, 227–231.
- [16] A. Horn and C. Johnson, *Matrix analysis*, Cambridge University Press, 1985.
- [17] R. Horst and N.V. Thoai, *DC programming: overview*, Journal of Optimization Theory and Applications **103** (1999), no. 1.
- [18] K.A.S. Immink, *Codes for mass data storage*, Shannon Foundation Publishers, Eindhoven, 2004.
- [19] K.A.S. Immink and K. Cai, *Simple classes of constrained systems with unconstrained positions that outperform the maxentropic bound*, IEEE International Symposium on Information Theory, 2008.
- [20] H. Ito, A. Kato, Zs. Nagy, and K. Zeger, *Zero capacity region of multidimensional run length constraints*, The Electronic Journal of Combinatorics **6** (1999).
- [21] H. Kamabe, *Insertion rate and optimization of redundancy of constrained systems with unconstrained positions*, Information Theory and Applications Workshop, 2009.
- [22] H. Kamabe, *Lower bounds of capacity of 2D constraints*, Information Theory and Applications Workshop, 2010.
- [23] P.W. Kasteleyn, *The statistics of dimers on a lattice*, Physica A **27** (1961), 1209–1225.
- [24] J.F.C. Kingman, *A convexity property of positive matrices*, The Quarterly Journal of Mathematics. Oxford. Second Series **12** (1963).
- [25] E.H. Lieb, *Residual entropy of square ice*, Phys. Rev. **162** (1967), no. 1, 162–172.

- [26] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, 1995 (reprinted 1999).
- [27] E. Loidor, *The tradeoff function for a class of RLL(d, k) constraints*, In review (SIAM Journal of Discrete Math).
- [28] E. Loidor and B.H. Marcus, *Improved lower bounds on capacities of symmetric 2-dimensional constraints using Rayleigh quotients*, IEEE Transactions on Information Theory **56** (2010), no. 4, 1624–1639.
- [29] E. Loidor, B.H. Marcus, and R. Pavlov, *Independence entropy of \mathbb{Z}^d shift spaces*, Submitted to Acta Applicandae Mathematicae.
- [30] E. Loidor, T.L. Poo, P. Chaichanavong, and B.H. Marcus, *Maximum insertion rate and capacity of multidimensional constraints*, IEEE International Symposium on Information Theory, 2008.
- [31] S. MacLane and G. Birkhoff, *Algebra*, second ed., Chelsea Pub Co, 1999.
- [32] B. Marcus, R. Roth, and P. Siegel, *Constrained systems and coding for recording channels*, Handbook of Coding Theory, Chapter 20, Elsevier Science.
- [33] M. Marcus and H. Minc, *A survey of matrix theory and matrix inequalities*, Allyn and Bacon, 1964.
- [34] N. Markley and M. Paul, *Maximal measures and entropy of z^{ν} subshifts of finite type*, Classical mechanics and dynamical systems **70** (1981), 135–157.
- [35] Z. Nagy and K. Zeger, *Capacity bounds for the three-dimensional (0,1) runlength limited channel*, IEEE Transactions on Information Theory **44** (2000), 1030–1033.
- [36] E. Ordentlich and R. Roth, *Independent sets in regular hypergraphs and multidimensional runlength-limited constraints*, SIAM J. Discret. Math. **17** (2004), no. 4, 615–623.
- [37] T.L. Poo, *Optimal code rates for constrained systems with unconstrained positions: An approach to combining error correction codes with modulation codes for digital storage systems*, Ph.D. thesis, Department of Electrical Engineering, Stanford, 2005.
- [38] T.L. Poo, P. Chaichanavong, and B.H. Marcus, *Trade-off functions for constrained systems with unconstrained positions*, IEEE Transactions on Information Theory **52** (2006), 1425–1449.

- [39] M. Schwartz and J. Bruck, *Constrained codes as networks of relations*, IEEE Transactions on Information Theory **54** (2008), 2179–2195.
- [40] A.J. van Wijngaarden and K.A.S. Immink, *Maximum runlength-limited codes with error control capabilities*, IEEE J. Select. Areas Commun. **19** (2001), 602–611.
- [41] K. Weber, *On the number of stable sets in an $m \times n$ lattice.*, Rostock. Math. Kolloq. **34** (1988), 28–36.
- [42] X. Yong and M.J. Golin, *New techniques for bounding the channel capacity of read/write isolated memory*, Data Compression Conference, 2002.