Recoverable Systems

Ohad Elishco Alexander Barg

***Abstract*—Motivated by the established notion of storage codes, we consider sets of infinite sequences over a finite alphabet such that every k-tuple of consecutive entries is uniquely recoverable from its l-neighborhood in the sequence. We address the problem of finding the maximum growth rate of the set, which we term capacity, as well as constructions of explicit families that approach the optimal rate. The techniques that we employ rely on the connection of this problem with constrained systems. In the second part of the paper we consider a modification of the problem wherein the entries in the sequence are viewed as random variables over a finite alphabet that follow some joint distribution, and the recovery condition requires that the Shannon entropy of the k-tuple conditioned on its l-neighborhood be bounded above by some $\epsilon > 0$. We study properties of measures on infinite sequences that maximize the metric entropy under the recoverability condition. Drawing on tools from ergodic theory, we prove some properties of entropy-maximizing measures. We also suggest a procedure of constructing an ϵ-recoverable measure from a corresponding deterministic system.**

I. Introduction

Consider the set of binary sequences $X_n \subset \{0,1\}^n$ with the property that every bit in the sequence is uniquely determined by its neighbors. What is the growth rate of the maximum size of X_n with this property as n increases without limit? Moreover, what can be said about the growth rate if we permit ourselves recovery with high probability rather than a deterministic decision? These questions form the starting point of this study which has its motivation in the recently established areas of storage coding and index coding, and draws on connections with constrained systems and entropy theory with the goal of establishing various "capacity" results for X_n and associated concepts.

Our main object is a *recoverable system* over a finite alphabet, which we proceed to define. We consider the set $Q^{\mathbb{Z}}$ of bi-infinite sequences $x = (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots)$ over a finite alphabet Q. A set $X \subset Q^{\mathbb{Z}}$ is called shift invariant if $TX = X$, where T is a shift by one position to the left. Let $[k] := \{0, 1, \dots, k-1\}$ and define the l-neighborhood $N_l([k])$ of the subset of coordinates $[k]$ as a set of all indices $j \in \mathbb{Z} \setminus [k]$ such that $|j - i| \leqslant l$ for some $i \in [k]$. The elements of the set $N_l([k])$ have a natural ordering, and we always think of the neighborhood as an ordered set.

Definition I.1 *Let Q be a finite alphabet and let $k, l \in \mathbb{N}$ be two integers. A set $X \subset Q^{\mathbb{Z}}$ is called a (k,l)-**recoverable system** if X is shift invariant and if there exists a deterministic function $f : Q^{2l} \to Q^k$ such that for every $x \in X$,*

$$f(x_{N_l([k])}) = x_{[k]}.$$

In other words, we look at sequences on $\mathbb{Z}$ (viewed as a graph) such that the value of a set of k consecutive vertices is recoverable from their l-neighborhood. The assumption of shift invariance is natural because the recovery property does not depend on the location of the k-tuple in x. For this reason it suffices to define the recovery function only for the set $x_{[k]}$.

A similar notion was studied earlier for general finite graphs under the name of *storage codes*, defined in [20], [21], and in [8], [13], [22] in the context of *guessing games* and dynamical systems on graphs. It has also been shown in [2], [20] that the capacity problem of storage codes on graphs is close (and for a sufficiently large alphabet is equivalent) to the minimum attainable rate of *index coding* as well as to the success probability in guessing games. In this context the concept of storage codes was introduced and studied in [3], [24].

An important difference between the storage codes and the above definition is that a storage code allows its own recovery function for every vertex while a recoverable system relies on a single repair rule for every vertex. Arguably, this restriction is appropriate for large graphs where it is difficult to account for many recovering functions, and it is also well suited for the analysis of storage capacity because it enables us to focus on shift invariant systems, and paves the way for using insights from symbolic dynamics, which we mention below.

Let Q be a finite alphabet and denote by Q^* the set of all finite words over Q. We say that a finite word w over Q is a **subword** in $x \in \mathbb{Q}^{\mathbb{Z}}$, denoted $w \prec x$, if w appears anywhere in x as a sequence of consecutive letters.

Definition I.2 *Let X be a (k,l)-recoverable system over a finite alphabet Q of size $|Q| = q$. Let $\mathcal{B}_n(X)$ denote the set of all length-n words each of which appears as a subword in some $x \in X$. The **capacity** of X is defined as*

$$\operatorname{cap}(X) = \lim_{n\to\infty} \frac{1}{n} \log_q |\mathcal{B}_n(X)|$$

where the limit exists due to the sub-additivity of $\log_q |\mathcal{B}_n(X)|$ and Fekete's lemma[1].

Recoverable systems are strongly related to the well-established notion of *constrained systems* [18], [19]. Informally speaking, a constrained system is a set of all sequences over a finite alphabet that do not include subwords from a set $\mathcal{F}$ of forbidden words. As explained below, a (k,l)-recoverable system is a sub-system of a constrained system. Indeed, for given $k, l \in \mathbb{N}$ the (k,l)-recoverability property gives rise to a

This work was presented in part at the 2021 IEEE International Symposium on Information Theory.

Ohad Elishco was with ISR, University of Maryland, College Park, MD 20742, USA. He is now with School of EE, Ben-Gurion University of the Negev, Beer Sheva, Israel. Email: elishco@gmail.com. Alexander Barg is with ISR and Department of ECE, University of Maryland, College Park, MD 20742, USA, and also with Inst. Inform. Trans. Probl., Russian Academy of Sciences, 127051 Moscow, Russia. Email abarg@umd.edu. This research was supported in part by the NSF grants CCF1814487, CCF2104489, as well as by NSF-BSF grant CCF-2110113.

[1] Fekete's lemma states that a sequence $\{a_n\}$ such that $a_{m+n} \leqslant a_m + a_n$ necessarily has a limit, and $\lim \frac{a_n}{n} = \inf_n \frac{a_n}{n}$ (see [10], [12]).

set of forbidden words $\mathcal{F}$ (by ruling out conflicts in recovery), and bi-infinite sequences without forbidden subwords form a subset in a constrained system.

The motivation to study (k,l)-recoverable systems is multifold. Indeed, they form a natural extension of the notion of storage codes to the infinite setting, and enable one to bring in methods from constrained systems such as de Bruijn graphs and Markov chains, which have not so far been used for storage codes. Recoverable systems also yield new insights into constructions of storage (index) codes for some classes of finite graphs, notably, circulant graphs. For such graphs we suggest a method of constructing a storage code over any given alphabet, yielding a lower bound on the maximum achievable code rate. We also show that capacity of recoverable systems (and related storage codes) can be increased by defining a probabilistic relaxation of the above definitions, which is a new concept in the area of erasure codes on graphs. This concept brings forth a link between recoverable systems and entropy of dynamical systems, enabling one to use tools such as metric entropy to derive bounds on the system capacity. It also suggests a group of new problems with a potential for connecting a broad array of methods from ergodic theory and discrete dynamical systems with the context of storing and recovering information on large graphs.

Our results are presented in the next three sections. We begin in Sec. II with establishing general properties related to the capacity of recoverable systems. In Sec. III we present two constructions of $(1,1)$-recoverable systems, one of which attains the maximum possible capacity whenever q is a whole square. Both constructions use techniques from constrained systems, namely the description of a system in terms of a Markov chain and de Bruijn graphs.

In Sec. IV we study a relaxation of the (k,l)-recoverability property wherein the recovered version of the contents is permitted to have a small residual entropy, resulting in a probabilistic version of the contents recovery. We call systems arising in this way (ϵ,k,l)-*recoverable*. In this part it is natural to switch to measure-theoretic language, making each vertex to carry a random variable X_i such that the collection of random variables $(X_i, i\in\mathbb{Z})$ follows some joint distribution. We provide a characterization of measures that maximize the entropy of such systems. We also show how to obtain a measure supported on an (ϵ,k,l)-recoverable system from a (k,l)-recoverable system with maximum capacity. The assumption of a single recovery function for every vertex on the line translates into a characterization of capacity as the topological entropy of the system. This allows us to use the variational principle which relates topological entropy and metric entropy, thus enabling us to borrow tools from entropy theory to evaluate the capacity.

Our results in this part are as follows. We prove that there exists an entropy-maximizing measure μ, which turns out to be a Markov measure whose conditional Shannon entropy equals ϵ. We propose a procedure of constructing an ϵ-recoverable measure from a deterministic recoverable system and bound below the increase of the system capacity due to the relaxation.

In the concluding Section V we present some details regarding the connection between (k,l)-recoverable systems and storage codes, and mention several problems for future research.

II. General properties of recoverable systems

We establish properties of (k,l)-recoverable systems by linking them to constrained systems, beginning with some notation. Similarly to the infinite case, if $w,u\in Q^*$ are finite words such that w appears in u as a consecutive subsequence, we say that w is a subword of u and denote this $w\prec u$. Let $|w|$ be the length of the word w. The concatenation of finite words w and u is denoted by wu and in particular, $w^n, n\geqslant 2$ is an n-fold concatenation of w with itself. Let T be a shift of the infinite word $x\in Q^{\mathbb{Z}}$ to the left, so for every $i\geqslant 1, j\in\mathbb{Z}$ and any $x\in Q^{\mathbb{Z}}$ we have $(T^i x)_j = x_{i+j}$.

The notion of recoverable systems is closely related to constrained systems. In what follows we will present some relevant notation and results from constrained systems (for a reference, see [19])

Let $G=(V,E,L)$ be a directed graph, where V is a finite set of vertices, E is the set of directed edges, and $L: E\to Q$ is a label function that assigns symbols from Q to edges, one symbol per edge. A path γ in G is a finite sequence of edges $\gamma=(e_0,\dots,e_{n-1})$. The label of the path $L(\gamma)$ is the sequence of symbols from Q obtained by reading off the labels of the edges in γ, i.e., $L(\gamma)=(L(e_i))_{i\in[n]}$. We allow multiple edges between a pair of vertices, but require that they carry different labels.

A **constrained system** $\mathscr{S}$ is the set of all finite words given by the labels of the paths in a directed labeled graph G. We call the graph G a *presentation* of $\mathscr{S}$, and we say that $\mathscr{S}$ is presented by G.

It is well known that a constrained system can also be defined by a (possibly infinite) set $\mathcal{F}\subseteq Q^*$ of finite words, called *forbidden words*, in the following way. Let $X_{\mathcal{F}}\subseteq Q^{\mathbb{Z}}$ be the set of all bi-infinite sequences such that no $x\in X_{\mathcal{F}}$ contains any word from $\mathcal{F}$ as a subword, i.e., there are no $i,j\in\mathbb{Z}$ for which $(T^i x)_{[j]}\in\mathcal{F}$. For every constrained system $\mathscr{S}$, it is possible to find a corresponding $\mathcal{F}$ such that $\mathscr{S}=\bigcup_{n\in\mathbb{N}}\mathcal{B}_n(X_{\mathcal{F}})$ (see [18]), where $\mathcal{B}_n(X_{\mathcal{F}})$ is the set of all length-n words that appear as a subword in some $x\in X_{\mathcal{F}}$. Therefore, both $\mathscr{S}$ or $X_{\mathcal{F}}$ will be referred to as a constrained system. Note however that they are objects of different kinds: while $\mathscr{S}$ is a set of finite words, $X_{\mathcal{F}}$ is a set of bi-infinite sequences.

The graph $G=(V,E,L)$ that presents a constrained system $X_{\mathcal{F}}$ can be described explicitly in the case that all the forbidden words have the same length, i.e., $\mathcal{F}\subseteq Q^n$ for some $n\in\mathbb{N}$. Namely, take $V=Q^{n-1}$ to be the set of all words of length $n-1$ over Q. Draw an edge from $u=(u_0,\dots,u_{n-2})$ to $v=(v_0,\dots,v_{n-2})$ if $v_i=u_{i+1},\ i\in[n-2]$ and $u_0\dots u_{n-2}v_{n-2}\notin\mathcal{F}$. The edge (u,v) carries the label $L((u,v))=v_{n-2}$. The sequences $x\in X_{\mathcal{F}}$ correspond to the sets of labels of the paths in G.

The capacity of a constrained system $\mathscr{S}$ is defined similarly to the capacity of a (k,l)-recoverable system. Denote by $\mathcal{B}_n(\mathscr{S})$ the set of all length-n words in $\mathscr{S}$, i.e., $\mathcal{B}_n(\mathscr{S}):=\mathscr{S}\cap Q^n$. Similarly, if the system is given by $X_{\mathcal{F}}$ then $\mathcal{B}_n(X_{\mathcal{F}})$ denotes the set of all length-n words that appear as subwords

in some $x \in X_{\mathcal{F}}$. The **capacity** of a constrained system $\mathscr{S}$ is defined as

$$\operatorname{cap}(\mathscr{S}) = \lim_{n\to\infty} \frac{1}{n} \log_q |\mathcal{B}_n(\mathscr{S})|.$$

In the language of symbolic dynamics this quantity is also known as the *topological entropy* of the system [26, Ch. 7]. We remark that we use the same notation for the capacity of constrained systems and recoverable systems. This usage is justified by Lemma II.2 below.

It is well known that the limit in the definition of $\operatorname{cap}(\mathscr{S})$ exists due to sub-additivity of the quantity $\log_q |\mathcal{B}_n(\mathscr{S})|$ and Fekete's lemma. Moreover,

$$\operatorname{cap}(\mathscr{S}) = \inf_{n\in\mathbb{N}} \frac{1}{n} \log_q |\mathcal{B}_n(\mathscr{S})|. \tag{1}$$

Assume that a constrained system $\mathscr{S}$ is presented by a graph $G = (V, E, L)$. The adjacency matrix A_G is a $|V|\times|V|$ matrix such that $(A_G)_{(u,v)}$ is the number of edges from u to v in G. If the graph G is strongly connected (there is a directed path between any two vertices), then the capacity $\operatorname{cap}(\mathscr{S})$ can be calculated using A_G. Indeed, the Perron-Frobenius theorem states that the largest eigenvalue of A_G, denoted λ, is real, simple, and positive with strictly positive left and right eigenvectors. Then the capacity of $\mathscr{S}$ equals $\operatorname{cap}(\mathscr{S}) = \log_q \lambda$ [19, Theorem 3.9]. In fact, $\operatorname{cap}(\mathscr{S}) = \log_q \lambda$ even if the graph G is not strongly connected (see [18, Theorem 4.4.4])

To analyze recoverable systems, we relate them to constrained systems using the following simple idea. Notice that if a word $w \in Q^k$ can be recovered from its l-neighborhood $u, v \in Q^l$, then no sequence $x \in X$ can contain a subword $uw'v$ for any $w' \in Q^k, w' \neq w$. This means that the words $uw'v$ are forbidden as subwords in some constrained system. This observation suggests the following definition.

Definition II.1 *Let $\mathcal{F} \subseteq Q^*$ be a set of finite words and let $k, l \in \mathbb{N}$. We say that $\mathcal{F}$ is a (k,l)-**admissible set** if for any $u, v \in Q^l$,*

$$\left|\left\{w \in Q^k \ : \ uwv \prec s \in \mathcal{F}\right\}\right| \geqslant q^k - 1.$$

It is immediate that a constrained system $X_{\mathcal{F}}$, where $\mathcal{F}$ is a (k,l)-admissible set, is a (k,l)-recoverable system.

Lemma II.2 *Any (k,l)-recoverable system $X \subseteq Q^{\mathbb{Z}}$ is a subset of a constrained system $X_{\mathcal{F}}$ for some (k,l)-admissible set $\mathcal{F}$. Conversely, any constrained system $X_{\mathcal{F}}$ with $\mathcal{F}$ a (k,l)-admissible set is a (k,l)-recoverable system.*

Proof: Let X be a (k,l)-recoverable system over the alphabet Q. Now let $\mathcal{F} = Q^{2l+k} \setminus \mathcal{B}_{2l+k}(X)$, then $\mathcal{F}$ is a (k,l)-admissible set and that $X \subseteq X_{\mathcal{F}}$. Indeed, if not, then there are two distinct words $w_1, w_2 \in Q^k$ such that $uw_1v \prec x$ and $uw_2v \prec y$, and by shift invariance we may assume that $x_{[k]} = w_1$ and $y_{[k]} = w_2$. This contradicts the recovery condition. Finally, the converse statement is obvious by Def. II.1. □

Let Y be a (k,l)-recoverable system over an alphabet Q of size q. As just argued, Y is a subset of some constrained system $X_{\mathcal{F}}$, where $\mathcal{F}$ is a (k,l)-admissible set, and thus the quantity

$$C_q(k,l) := \max_{\substack{Y\subseteq Q^{\mathbb{Z}} \\ Y \text{ is } (k,l)\text{-recoverable}}} \operatorname{cap}(Y). \tag{2}$$

is well defined. Moreover, $\operatorname{cap}(Y) \leqslant \operatorname{cap}(X_{\mathcal{F}})$, and since there is a finite number of possible (k,l)-admissible sets over Q, the maximum in (2) is attained. We say that X has **maximum capacity** if $\operatorname{cap}(X) = C_q(k,l)$ and note that the maximizing system is not necessarily unique.

The following properties of $C_q(k,l)$ are immediate: if $l_1 \geqslant l_2$ or $q_1 \geqslant q_2$, then

$$C_{q_1}(k,l_1) \geqslant (\log_{q_1} q_2) C_{q_2}(k,l_2) \tag{3}$$

To illustrate these definitions, let us consider the simplest instance of the concepts introduced above, namely binary $(1,1)$-recoverable systems.

Proposition II.3 *The capacity $C_2(1,1) \approx 0.4057$, and there is exactly one system X that attains it, up to symbol relabeling.*

Proof: The proof proceeds by examining all the possible binary $(1,1)$-recoverable systems. On account of Lemma II.2 this amounts to examining all the possible constrained systems $X_{\mathcal{F}}$, where $\mathcal{F}$ is a $(1,1)$-admissible set. The $(1,1)$-recoverability condition implies that a symbol is determined completely by the value of its neighbors. Thus, for instance, one of the sequences $000, 010$ is in $\mathcal{F}$. Overall, out of the eight possible triplets four are in $\mathcal{F}$, while the remaining four appear in the sequences in X.

If 010 or 101 are in $\mathcal{F}$, then $\operatorname{cap}(X) = 0$. This is because in either of these cases, X is formed of at most two sequences. Indeed, suppose that $010 \in \mathcal{F}$. Between the two triplets 111 and 101, one must be forbidden to allow recoverability. Say it is 111, then $X = \{(000)^{\mathbb{Z}}, (101)^{\mathbb{Z}}\}$ (the word $(0011)^{\mathbb{Z}}$ is forbidden to allow recoverability). Alternatively, if 101 is forbidden, then $X = \{(000)^{\mathbb{Z}}, (111)^{\mathbb{Z}}\}$. The same arguments apply to 101 upon the relabeling $0 \to 1, 1 \to 0$. Thus $010, 101 \notin \mathcal{F}$, which implies that $000, 111 \in \mathcal{F}$.

It remains to analyze the triplets 110 and 100. If $110 \in \mathcal{F}$, then by shift-invariance also $011 \in \mathcal{F}$, otherwise the capacity is 0. This yields that $000, 111, 110, 011 \in \mathcal{F}$ and the remaining four patterns are allowed. This system is known as a $(1,2)$-RLL constrained system, and its capacity is approximately 0.4057 [17, Ch. 4]. The pattern 100 yields the same result by relabeling. This concludes the proof. □

The following result enables us to claim that $C_q(k,l)$ is positive for general alphabets.

Corollary II.4 *Let $q \geqslant 3$ and let $2 \leqslant k \leqslant l-1$, then*

$$C_q(k,l) \geqslant C_q(k,k+1) \geqslant \frac{\log_q 2}{k+2}.$$

Proof: We explicitly describe a $(k,k+1)$-recoverable system X over $Q = \{0,1,2\}$ with positive capacity. By (3) this will imply the general claim.

Construct a set of sequences X by taking all bi-infinite sequences over $\{a,b\}$ and replacing each a with $2\,1^{k+1}$ (two followed by $k+1$ ones) and each b with $2\,0^{k+1}$. We claim that X is a $(k,k+1)$-recoverable system. Indeed, let $u = (u_0 \dots u_k), v = (v_0 \dots v_k) \in Q^{k+1}$, $w \in Q^k$ and let us show that w is recoverable from u and v. Let u_{-1} be the symbol right before u_0. Exactly one of the elements $u_{-1}, u_0, \dots, u_k$ is 2, and this gives rise to the following 4 cases.

(i) $u_{-1}=2$, then $w=2v_0^{k-1}$;
(ii) $u_0=2$, then $w=u_k 2v_0^{k-2}$;
(iii) $u_i=2, 1\leqslant i\leqslant k-1$, then $w=u_k^{i+1}2v_0^{k-i-2}$;
(iv) $u_k=2$, then $w=v_0^k$.

This shows that w can be recovered from its $(k+1)$-neighborhood.

To calculate $\mathrm{cap}(X)$, notice that $2^n = |\{a,b\}^n| = |\mathcal{B}_{n(k+2)}(X)|$. Thus,

$$\frac{1}{n(k+2)}\log_q|\mathcal{B}_{n(k+2)}(X)| = \frac{\log_q 2}{k+2}.$$

Since this equality is true for any n, this implies the claimed result. □

Our next result is an upper bound on the capacity of (k,l)-recoverable systems.

Lemma II.5 *For all $q\geqslant 2$ and $k\leqslant l$ we have*

$$C_q(k,l)\leqslant \frac{l}{k+l}.$$

Proof: From Lemma II.2 it suffices to show that $\mathrm{cap}(X_{\mathcal{F}})\leqslant \frac{l}{k+l}$ for every constrained system $X_{\mathcal{F}}$ with a (k,l)-admissible set $\mathcal{F}$. On account of (1) we need to prove that for every $\epsilon>0$, there exists $n\in\mathbb{N}$ for which

$$\frac{1}{n}\log_q|\mathcal{B}_n(X_{\mathcal{F}})|\leqslant (1+\epsilon)\frac{l}{k+l}.$$

Fix $\epsilon>0$ and take $m\in\mathbb{N}$ large enough such that $\frac{m+1}{m}<1+\epsilon$. Now take $n=m(k+l)+l$ and let F_n denote the set of indices given by

$$F_n=\{j\in[n]\,|\,j \bmod (k+l)\in[l]\}.$$

Consider $x\in X_{\mathcal{F}}$. If the symbols $x_j, j\in F_n$ are known, then the remaining symbols in $[n]\backslash F_n$ are determined uniquely because of the recoverability property. Since there are at most $q^{(m+1)l}$ different words that can appear in the locations F_n, we obtain

$$\begin{aligned}\frac{1}{n}\log_q\left(|\mathcal{B}_n(X_{\mathcal{F}})|\right)&\leqslant \frac{1}{n}\log_q q^{(m+1)l}\\&\leqslant \frac{(m+1)l}{m(k+l)}\\&\leqslant (1+\epsilon)\frac{l}{k+l}.\end{aligned}$$

□

An obvious observation, implied by this bound, is that many symbols cannot be recovered from a small-size neighborhood, i.e., $\lim_{k\to\infty}C_q(k,l)=0$.

For certain values of q we can use the well-known edge covering construction from the storage coding and index coding literature [5], [21] to attain the bound in this lemma.

Proposition II.6 *If q is a whole square, then $C_q(l,l)=\frac{1}{2}$. If $q=t^{k+1}$ for some integer t, then $C_q(k,1)=\frac{1}{k+1}$.*

Proof: Consider a graph $G(V,E)$ with $V=\mathbb{Z}$ and $(i,j)\in E$ if and only if $|i-j|=1$. Place a symbol of the alphabet Q on each edge and assign to a vertex the pair of symbols which appear on the edges adjacent to it, then the symbol $x_i, i\in\mathbb{Z}$ is determined once we know x_{i-1} and x_{i+1}. This gives rise to a $(1,1)$-recoverable system X over the alphabet of size q^2, where $q=|Q|$, and clearly $|\mathcal{B}_n(X)|\geqslant q^n$. If $q=t^2$, then we conclude that

$$C_q(1,1)\geqslant \lim_{n\to\infty}\frac{1}{n}\log_q t^n = 1/2.$$

Lemma II.5 implies that this is an equality.

For $l\geqslant 2$, assign the vertex x_i, $i\in\mathbb{Z}$ the pair of symbols that appear on the edges $(i-1,i)$ and $(i+l-1,i+l)$.

As before, this enables us to argue that $C_q(l,l)=\frac{1}{2}$. To prove the second claim of the lemma, construct a system X by assigning each vertex $i\in\mathbb{Z}$ the $k+1$-tuple that appears on the edges

$$(i-1,i),(i,i+1),\ldots,(i+k-1,i+k).$$

It is straightforward to check that the system X is $(k,1)$-recoverable and that its capacity is at least $1/(k+1)$. Now the claim follows from Lemma II.5. □

We can remove the assumptions on q in the limit of large alphabet.

Corollary II.7 *For $l,k\geqslant 1$ and $q\to\infty$ the limits of $C_q(l,l)$ and $C_q(k,1)$ exist, and*

$$\lim_{q\to\infty}C_q(l,l)=\frac{1}{2}$$
$$\lim_{q\to\infty}C_q(k,1)=\frac{1}{k+1}.$$

Proof: Letting $q=t^2+r$, we find

$$C_q(l,l)\geqslant\lim_{n\to\infty}\frac{1}{n}\log_{t^2+r}t^n=\log_{t^2+r}t,$$

which goes to $1/2$ as $q\to\infty$. Similarly, letting $q=t^{k+1}+r$, we find

$$\begin{aligned}C_q(k,1)&\geqslant\lim_{n\to\infty}\frac{1}{n}\log_{t^2+r}t^n\\&=\log_{t^{k+1}+r}t\to\frac{1}{k+1}.\end{aligned}$$

Now the claim follows from Lemma II.5. □

III. Constructions of $(1,1)$-recoverable systems

In this section we focus on the case $k=l=1$. For this case, the capacity is at most $1/2$ by Lemma II.5, and we have pointed out that it is attainable when q is a whole square. By extending the edge covering construction to all q and using (3) we can obtain lower bounds on $C_q(1,1)$ for general q. In this section we present two different constructions of recoverable systems that yield higher capacity values for all q except whole squares.

A. A recursive construction

In the first construction we relate $C_q(1,1)$ to $C_{q-2}(1,1)$. Let $G=(V,E)$ be a strongly connected graph and let $\Gamma_n(G)$ be the set of all directed paths of length $n\geqslant 1$ in G. Let p be a probability vector $p=(p_0,\ldots,p_{|V|-1})$ and let $\epsilon>0$. We denote by $\Gamma_n^{\epsilon}(G,p)$ the set of all directed paths $\gamma\in\Gamma_n(G)$ such that for every $v\in V$, γ traverses v at least $(p_v-\epsilon)n$ times. The following result is cited from [19]; we have adjusted it here to our notation.

Lemma III.1 ([19, Thm. 3.15]) *Let $G = (V, E)$ be a strongly connected graph. There exists a probability vector $p = (p_0, \dots, p_{|V|-1})$ such that for every $\epsilon > 0$, there exists an $n_\epsilon \in \mathbb{N}$ such that for all $n \geqslant n_\epsilon$,*

$$|\Gamma_n^\epsilon(G, p)| \geqslant (1-\epsilon)|\Gamma_n(G)|.$$

The idea of the construction, formalized in the next lemma, consists in adding two new symbols to the alphabet and adding a loop of length four to a frequently traversed vertex in the graph G that presents the system.

Lemma III.2 *We have*

$$C_{q+2}(1,1) \geqslant C_q(1,1)\log_{q+2} q + \frac{1}{q^2}\log_{q+2}\left(1+\frac{1}{q^2}\right).$$

Proof: Let X be a $(1,1)$-recoverable system with maximum capacity over Q. By Lemma II.2 we may assume that X is a constrained system $X_{\mathcal{F}}$ for some $(1,1)$-admissible set $\mathcal{F} \subseteq Q^3$ of forbidden words. Let $G_q = (V_q, E_q, L_q)$ be a graph that presents $X_{\mathcal{F}}$ and suppose that each vertex in G_q corresponds to a pair of symbols of Q. For two vertices $u = (u_1u_2), v = (v_1v_2)$, the edge $(u, v) \in E_q$ iff $u_2 = v_1$ and $u_1u_2v_2 \notin \mathcal{F}$. The label of the edge $L_q((u, v)) = v_2$. We may also assume that G_q is connected since the capacity can be calculated by considering only the connected components of G_q.

Fix $\epsilon > 0$ and use Lemma III.1 to find $n_\epsilon \in \mathbb{N}$ and a probability vector $p = (p_0, \dots, p_{|V_q|-1})$ such that for $n \geqslant n_\epsilon$

$$|\Gamma_n^\epsilon(G_q, p)| \geqslant (1-\epsilon)|\Gamma_n(G_q)|.$$

Since p is a probability vector, there exists a vertex $v \in V_q$ for which $p_v \geqslant \frac{1}{q^2}$. Assume that v corresponds to the pair $(ab) \in Q^2$.

Let us add symbols α and β to the alphabet Q. Construct a new graph $G_{q+2} = (V_{q+2}, E_{q+2}, L_{q+2})$ by adding the vertices $u_1 = (b\alpha), u_2 = (\alpha\beta), u_3 = (\beta a)$ with edges

$$\xi_1 = (ab, u_1), \xi_2 = (u_1, u_2),$$
$$\xi_3 = (u_2, u_3), \xi_4 = (u_3, ab)$$

and labels α, β, a, b, respectively. The graph G_{q+2} presents a system $X_{\mathcal{F}'}$ for some $\mathcal{F}'$ over the alphabet of size $q+2$.

Let us show that $X_{\mathcal{F}'}$ is a $(1,1)$-recoverable system. Consider a sequence $x \in X_{\mathcal{F}'}$. Note that if $x \notin X_{\mathcal{F}}$, it must contain at least one of the symbols α, β. By the construction of G_{q+2}, the words of length three that contain α or β and can appear as subwords in x, are

$$A := \{ab\alpha, b\alpha\beta, \alpha\beta a, \beta ab\}.$$

Therefore, the system $X_{\mathcal{F}'}$ is a constrained system that corresponds to the set of forbidden words

$$\mathcal{F}' = \left((Q \cup \{\alpha, \beta\})^3 \setminus (Q^3 \cup A)\right) \cup \mathcal{F}.$$

It is straightforward to check that $\mathcal{F}'$ is a $(1,1)$-admissible set, which implies that $X_{\mathcal{F}'}$ is a $(1,1)$-recoverable system.

Let us bound below the capacity of $X_{\mathcal{F}'}$. Assume first that $p_v < \frac{1}{4}$ and let $r \in \mathbb{N}$ be large enough such that $\lfloor (1-4p_v)r \rfloor \geqslant n_\epsilon$. Note that if $\gamma = (e_0, \dots, e_r)$ is a path in G_q, then it is also a path in G_{q+2}. Moreover, if for some $i \in [r]$, e_i terminates in v, then

$$\gamma' = (e_0, \dots, e_i, \xi_1, \xi_2, \xi_3, \xi_4, e_{i+1}, \dots, e_r)$$

is a path in G_{q+2} of length $|\gamma| + 4$. Let $\mathcal{B}_r^\epsilon(X_{\mathcal{F}})$ denote the set of words in $\mathcal{B}_r(X_{\mathcal{F}})$ with at least $M_v := \lfloor (p_v - \epsilon)r \rfloor$ appearances of ab. Each such word is obtained by reading off the labels of a length-n path in G_q. Then

$$\begin{aligned}
&|\mathcal{B}_r(X_{\mathcal{F}'})| \geqslant \\
&\sum_{i=0}^{M_v} \binom{M_v}{i} \left|\mathcal{B}_{r-4i}^\epsilon(X_{\mathcal{F}})\right| \\
&\overset{(a)}{\geqslant} \sum_{i=0}^{M_v} \binom{M_v}{i} (1-\epsilon) q^{(r-4i)\operatorname{cap}(X_{\mathcal{F}})} \\
&= q^{r\operatorname{cap}(X_{\mathcal{F}})}(1-\epsilon)\sum_{i=0}^{M_v}\binom{M_v}{i} q^{-4i\operatorname{cap}(X_{\mathcal{F}})} \\
&= q^{r\operatorname{cap}(X_{\mathcal{F}})}(1-\epsilon)\left(1+q^{-4\operatorname{cap}(X_{\mathcal{F}})}\right)^{M_v}
\end{aligned}$$

where to claim inequality (a) we used the estimate $|\mathcal{B}_{r-4i}^\epsilon(X_{\mathcal{F}})| \geqslant (1-\epsilon)|\mathcal{B}_{r-4i}(X_{\mathcal{F}})|$, valid for all $i \leqslant (p_v - \epsilon)r$, the assumption on r, and definition (1). We obtain

$$\begin{aligned}
\frac{1}{r}\log_{q+2}|\mathcal{B}_r(X_{\mathcal{F}'})| &\geqslant \operatorname{cap}(X_{\mathcal{F}})\log_{q+2} q \\
&+ \frac{(p_v-\epsilon)r-1}{r}\log_{q+2}(1+q^{-4\operatorname{cap}(X_{\mathcal{F}})}) \\
&+ \frac{1}{r}\log_{q+2}(1-\epsilon).
\end{aligned}$$

From Lemma II.5, $\operatorname{cap}(X_{\mathcal{F}}) \leqslant \frac{1}{2}$. Since $p_v \geqslant \frac{1}{q^2}$, taking $r \to \infty$ we obtain

$$\operatorname{cap}(X_{\mathcal{F}'}) \geqslant \operatorname{cap}(X_{\mathcal{F}})\log_{q+2} q + \left(\frac{1}{q^2} - \epsilon\right)\log_{q+2}\left(1+\frac{1}{q^2}\right).$$

Since ϵ is arbitrary,

$$\operatorname{cap}(X_{\mathcal{F}'}) \geqslant \operatorname{cap}(X_{\mathcal{F}})\log_{q+2} q + \frac{1}{q^2}\log_{q+2}\left(1+\frac{1}{q^2}\right).$$

The case $p_v \geqslant \frac{1}{4}$ is similar. We only consider the first $n/4$ visits to v, with r large enough such that $\lfloor \epsilon r \rfloor \geqslant n_\epsilon$, $M_v := \left\lfloor (\frac{1}{4} - \epsilon)r \right\rfloor$, and otherwise the proof is the same. □

For alphabets of size q not a whole square, i.e., $q = t^2 + r$ with $r > 0$, we have two options of constructing a recoverable system, namely relying on the alphabet of size t^2 (by the edge covering procedure, see the proof of Prop. II.7), or using the construction given in the proof of Lemma III.2. The following example shows that the latter approach can be superior to the former, yielding a system with larger capacity.

Example III.3 *Let X be a $(1,1)$-recoverable system over an alphabet of size $q = 6$, constructed by edge covering. This yields the estimate $\operatorname{cap}(X) = \log_6 2$. At the same time, from Lemma III.2 we obtain*

$$C_6(1,1) \geqslant \frac{1}{2}\log_6 4 + \frac{1}{16}\log_6\frac{17}{16} > \log_6 2.$$

B. Construction via de Bruijn graphs

We now provide a second construction of $(1,1)$-recoverable systems over an alphabet Q of size q, which relies on the de Bruijn graphs [9], [16]. The capacity analysis of the constructed systems will be carried out for the case of q close to the nearest whole square greater than it. For a finite alphabet Q and $d \in \mathbb{N}$, a **de Bruijn graph** of order d over Q is the directed labelled graph $\mathrm{DB}(Q,d) = (V,E,L)$, whose vertex set $V = Q^d$ corresponds to length-d words over Q. An edge $(u,v) \in E$ if $v = (v_0 \dots v_{d-1}), u = (u_0 \dots u_{d-1})$ and $v_i = u_{i+1}$ for all $i \in [d-1]$, and the label of $(u,v) \in E$ is $L((u,v)) = v_{d-1}$. In words, there is an edge from u to v if v is the $(d-1)$-tail of u appended by some symbol $v_{d-1} \in Q$ which serves as a label. The relevance of the de Bruijn graphs to (k,l)-recoverable systems is established in the next lemma.

Lemma III.4 *For a finite alphabet Q, $\mathrm{DB}(Q,d)$ yields a presentation of a $(k,1)$-recoverable system over Q^d for every $k \in [d]$.*

Proof: Consider a new labeling function L' such that for an edge $(u,v) \in \mathrm{DB}(Q,d)$, $L'((u,v)) = (u_0,\dots,u_{d-1}) \in Q^d$. By definition, $\mathrm{DB}(Q,d)$ with the labeling L' is a presentation of a constrained system $\mathscr{S}$ over Q^d.

To see that the system $\mathscr{S}$ is $(k,1)$-recoverable for every $k \in [d]$, notice that if x is a sequence obtained by reading off the labels of a path in $\mathrm{DB}(Q,d)$ (with L'), then $\boldsymbol{x}_{-1} = (x_{-1}, x_0, \dots, x_{d-2})$ and $\boldsymbol{x}_k = (x_k, x_{k+1}, \dots, x_{k+d-1})$. While $k \leqslant d-1$ we have that for every $i \in [k]$, the sequences $\boldsymbol{x}_i = (x_i, \dots, x_{i+d-1})$ can all be constructed from $\boldsymbol{x}_{-1}$ and $\boldsymbol{x}_k$, proving recoverability. □

Remark III.5 *The recovery procedure used in the proof relies on sharing the label between a pair of vertices across the edge that connects them, and is similar to the edge covering construction used in the proof of Proposition II.6. In fact, a de Bruijn graph $\mathrm{DB}(Q,d)$ can be though of as a presentation of a "d-block-encoder". In terms of Proposition II.6, if $G(V,E)$ is a graph with $V = \mathbb{Z}$, $(i,j) \in E$ if $|i-j| = 1$, and a symbol is placed on each edge, the d-block-encoding is the system obtained by assigning each vertex $i \in \mathbb{Z}$ the d-tuple that appears on the edges*

$$(i-d, i-d+1), \dots, (i-2, i-1), (i-1, i).$$

Thus, the capacity of the $(k,1)$-recoverable system presented by $\mathrm{DB}(Q,d)$ is $\frac{1}{d}$.

To conform with the definition of a de Bruijn graph, we fix the number of vertices to q^d. At the same time, any subgraph of $\mathrm{DB}(Q,d)$ corresponds to a $(k,1)$-recoverable systems with fewer than q^d vertices. We state this observation in the next corollary.

Corollary III.6 *Any subgraph of $\mathrm{DB}(Q,d)$ presents a $(k,1)$-recoverable system for any $k \in [d]$.*

By relabeling the vertices, we can obtain a $(k,1)$-recoverable systems over an alphabet of size strictly smaller than q^d. This is the key insight of our construction.

The construction makes use of de Bruijn graphs of order $d = 2$. Therefore, unless specified otherwise, for the rest of the section we assume that $d = 2$ and write $\mathrm{DB}(Q)$ instead of $\mathrm{DB}(Q,2)$. We will need some properties of the adjacency matrix A_D of $\mathrm{DB}(Q)$, assuming that the vertices are ordered according to the lexicographic order of the their labels. The following description of A_D is immediate.

Lemma III.7 *Let A_D be the adjacency matrix of $\mathrm{DB}(Q)$. Then for $i,j \in \left[q^2\right]$*

$$(A_D)_{(i,j)} = \begin{cases} 1 \text{ if } j \in \{(i \bmod q)\cdot q + [q]\} \\ 0 \text{ otherwise.} \end{cases}.$$

Proof: Interpret the words of length 2 over Q as base-q numbers from 0 to q^2-1. Clearly, a number u is connected to a number v if and only if $v \in \{(qu \bmod q^2) + [q]\}$. □

Example III.8 *The matrices A_D have a simple structure which we show in examples. This makes later proofs easier to follow. Let $A_{D_i}, i = 1,2$ be the adjacency matrices of $D_1 = \mathrm{DB}([2])$ and $D_2 = \mathrm{DB}([3])$, then*

$$A_{D_1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

and

$$A_{D_3} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

The following property of de Bruijn graphs is well known (e.g., [6, Exercise 1.5]). We include a short proof for completeness.

Lemma III.9 *Let A_D be the adjacency matrix of $\mathrm{DB}(Q)$. Then the eigenvalues of A_D are $\lambda_0 = 0$ with multiplicity q^2-1 and $\lambda_1 = q$ with multiplicity 1.*

Proof: In $\mathrm{DB}(Q)$ there is a unique path of length 2 from any $v = (v_0, v_1) \in V$ to any $u = (u_0, u_1) \in V$, namely, $\gamma = (v, w_1), (w_1, u)$ where $w_1 = (v_1 u_0)$. Thus, $(A_D)^2 = J$, the all ones matrix, which has a simple eigenvalue q^2, and eigenvalue 0 of multiplicity q^2-1. Thus, A_D has eigenvalues $\lambda_0 = 0$ and $\lambda_1 = q$ or $-q$. To argue that we should choose the $+$ sign note that $\mathrm{Tr} A_D = q$. □

We now construct a $(1,1)$-recoverable system X over $Q = [q]$ for $q \in \mathbb{N}$. According to Lemma III.4, $\mathrm{DB}([t])$ presents such a system, and if $q = t^2$ for some $t \in \mathbb{N}$, then by Lemma III.9 the capacity $\mathrm{cap}(X) = \log_q t = \frac{1}{2}$. Now suppose that q is not a whole square. As Corollary III.6 suggests, we can obtain a $(1,1)$-recoverable system by considering a subgraph of a de Bruijn graph. Let $t = \lceil \sqrt{q} \rceil$, $q = t^2 - r$, and consider the graph $\mathrm{DB}([t])$ with the adjacency matrix A_D. Taking a subgraph of $\mathrm{DB}([t])$ amounts to deleting rows and columns from A_D. The obtained matrix, which we call the **truncated matrix** A_Q, is the adjacency matrix of a de Bruijn subgraph that presents a $(1,1)$-recoverable system. This sequence of

operations, illustrated in the examples below, will be justified formally in the remainder of this section.

Before proceeding we make one remark. A simple way to construct the truncated matrix, which we largely use, is to delete the last $r = t^2 - q$ columns and rows in A_D. In what follows we restrict ourselves to alphabets Q of size $q = t^2 - r$, where $t = \lceil \sqrt{q} \rceil$ and $r \leqslant t$. Deleting the last r columns from A_D can be interpreted as deleting the outgoing edges of the vertices that correspond to these columns. If $r < t$, then upon deleting the last r columns, no all-zeros rows arise. At the same time, if $t = r$, deleting the last $r = t$ columns results in t all-zeros rows. These rows correspond to vertices with no outgoing edges, there is a complication that they are not the last t rows in the matrix. While such rows do not contribute to the capacity of the system, we cannot erase them since their incoming edges will point nowhere. Thus when $t = r$ we need a different method to delete rows and columns: instead of deleting the t last rows and columns, we erase the rows and columns in positions $\{it + (t-1), i = 0, 1, \ldots, t-1\}$.n

Example III.10 *In this example we construct a* $(1,1)$*-recoverable system over* $Q = [8]$*. One possible construction, according to Lemma III.4, is to generate* $\mathrm{DB}([2], 3)$ *and to use it as a* $(1,1)$*-recoverable system with capacity* $\frac{1}{3}$ *(in fact, it is also* $(1,2)$*-recoverable).*

Another way to obtain such a system is to use the procedure described above. Let $t = \lceil \sqrt{q} \rceil = 3$ *and consider the subgraph of the* 9×9 *matrix of* $\mathrm{DB}(3)$ *obtained after deleting the last column and last row. We have*

$$A_Q = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

Considering the graph presented by A_Q*, we obtain a* $(1,1)$*-recoverable system with capacity* $\log_8(1+\sqrt{3}) \approx 0.483 > \frac{1}{3}$*. Here* $1 + \sqrt{3}$ *is the largest eigenvalue of the matrix* A_Q*.*

Example III.11 *We now construct a* $(1,1)$*-recoverable system over* $Q = [6]$ *using the procedure described above. Let* $t = \lceil \sqrt{q} \rceil = 3$ *and consider the matrix* $\mathrm{DB}(3)$ *after deleting rows (and columns)* $2, 5, 8$*. We have*

$$A_Q = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

Considering the graph presented by A_Q*, we obtain a* $(1,1)$*-recoverable system with capacity* $\log_6(2) \approx 0.387$*. Notice that the last* $t - 1 = 2$ *columns in* A_Q *contain zeros. This implies that* A_Q *corresponds to a graph in which the last* $t-1$ *vertices have no incoming edges. Removing them does not affect the capacity but reduces the alphabet size. Indeed, if we remove those vertices we obtain* $\mathrm{DB}([2])$ *which has capacity* $\frac{1}{2}$*.*

The truncated matrix A_Q obtained from the above procedure has the following property.

Lemma III.12 *Let* Q *be a finite alphabet of size* $q = t^2 - r$ *with* $t = \lceil \sqrt{q} \rceil$*. If* $r < t$ *then the matrix* A_Q *is an adjacency matrix of a strongly connected graph* $G = (V, E, L)$*.*

Proof: We will prove a stronger statement, namely, that the diameter of the graph G with the adjacency matrix A_Q equals two. In other words, there is a path of length at most 2 between any two vertices.

Let $q = t^2 - r$ where $t = \lceil \sqrt{q} \rceil$ and $r < t$. According to Lemma III.7, every row in the adjacency matrix of $\mathrm{DB}(t)$ contains t consecutive ones. The truncated matrix A_Q of order q obtained after the deletion of at most r rows, contains two types of rows, rows with t consecutive ones, and rows with $t - r$ consecutive ones.

If a row contains t consecutive ones, they appear in positions $tj + [t] = \{tj, tj+1, \ldots, tj + (t-1)\}$ for some $j < t - 1$. More accurately, row j contains t consecutive ones in positions $t(j \bmod t) + [t]$ if $j \neq t - 1 \bmod t$. If $j \bmod t = t - 1$ then row j contains $t - r$ consecutive ones in positions $t(t-1) + [t-r]$. This implies that for every $j \in [q]$ and every set of t consecutive rows in A_Q, there is exactly one row from that set that has 1 in position j. This means that any set of vertices that corresponds to t consecutive rows in the matrix is a dominating set in G (i.e., it connects to all the other vertices in G).

A vertex v that corresponds to a row with t consecutive ones is connected to a set of vertices that correspond to t consecutive rows. Therefore, there is a directed path of length at most 2 from v to any other vertex in G. At the same time, every row that contains $t - r$ ones in the last positions, corresponds to a vertex v that is connected to the last $t - r$ vertices (rows). These $t - r$ vertices form a dominating set in G, giving rise to a path of length 2 from v to any other vertex in G. □

Lemma III.12 implies the following statement.

Proposition III.13 *Let* Q *be an alphabet of size* q *and let* $q = t^2 - r$*, where* $t = \lceil \sqrt{q} \rceil$*. Let* A_Q *be the truncated matrix. For* $r < t$*,* A_Q *has at most two non-zero eigenvalues (without accounting for multiplicity).*

Proof: Assume that $r < t$, and let us show that in this case $\mathrm{rk}((A_Q)^2) = 2$. In this case the truncation procedure described above does not give rise to any all-zero rows. The following statements are checked by inspection. The matrix A_Q contains two types of rows, those with t consecutive ones (located in positions $tj + [t]$ for $j < t - 1$) and those with $t - r$ consecutive ones (located in the last $t - r$ positions). The row j contains $t - r$ consecutive ones if $j \bmod t = t - 1$ and t consecutive ones otherwise. Because of this, there are also two different types of columns. A column of the first type contains t ones of which every two are separated by $t - 1$ zeros. A column of the second type contains $t - 1$ ones of which every two are separated by $t - 1$ zeros. In columns of the first type there is a single one that appears in the last $t - r$ positions, and in columns of the second type the last $t - r$ positions are all zeros. The columns of the second type are located in the last $r(t-1)$ positions. The arrangement of the rows and the columns is shown in Figure 1.

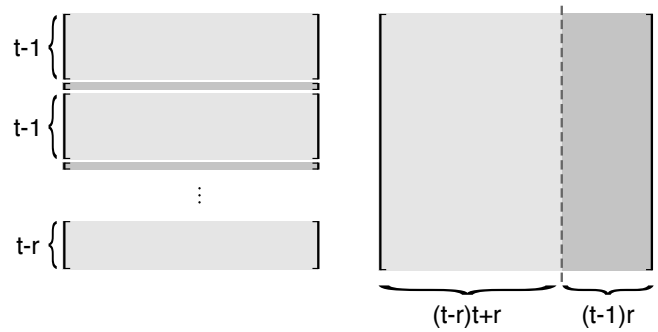


Figure 1: The structure of the matrix A_Q. The left figure details the arrangement of the rows: each block of size $t-1$ (light gray) contains rows of the first type, and is followed by a single row of the second type (dark gray). Overall there are $t-1$ rows of the second type. The last block is formed of $t-r$ rows of the first type. The right figure describes the structure of the columns. There are $(t-r)t+r$ columns of the first type (light gray) followed by $(t-1)r$ columns of the second type (dark gray).

The structure of the matrix implies that the product of any row with a first-type column is 1, the product of a first-type row with a second-type column is also 1, and the product of a second-type row with a second-type column is 0. Moreover, the matrix A_Q^2 contains rows of only two different kinds, those of all ones and those with $(t-r)t$ ones followed by $r(t-1)$ zeros. Hence $\mathrm{rk}(A_Q^2)=2$ and thus A_Q^2 has at most 2 non-zero eigenvalues, which may or may not be different. Note that the rank of A_Q can be greater than two; however, an eigenvector of A_Q is also an eigenvector of A_Q^2, therefore, the matrix A_Q has at most two eigenvectors with nonzero eigenvalues (an eigenvalue $\lambda \neq 0$ of A_Q gives rise to an eigenvalue λ^2 of A_Q^2). This proves the claim for $r<t$.

□

Next we deal with the case $r=t$.

Proposition III.14 *Let Q be a finite alphabet of size $q=t^2-t$ with $t=\lceil\sqrt{q}\rceil$. The Perron eigenvalue of the truncated matrix A_Q is $\lambda=t-1$.*

Proof: Recall that when $r=t$, the matrix A_Q is obtained from $\mathrm{DB}([t])$ by deleting rows (and columns) $\{it+(t-1) \,:\, i\in[t]\}$. The structure of $\mathrm{DB}([t])$ suggests the following deletion procedure. Every row of $\mathrm{DB}([t])$ contains exactly t ones, and we begin by eliminating all the rows in which the t ones appear at the end, i.e., in positions $(t-1)t+[t]$. There are exactly t such rows with numbers in the set $\{it+(t-1) \,:\, i\in[t]\}$. Next, delete every column that intersects the last one in some row. Exactly t columns are thus eliminated. Every row of the truncated matrix A_Q contains exactly $t-1$ ones, and so the all-ones vector is a positive eigenvector with eigenvalue $t-1$. For a nonnegative matrix, a positive eigenvector corresponds to the largest eigenvalue, and this finishes the proof. □

Remark III.15 *The construction of A_Q in the case $r=t$ generates a $(1,1)$-recoverable system with effective alphabet which is strictly smaller than $q=t^2-t$, namely it uses only $(t-1)^2$ letters out the q available letters in Q.*

We can now state and prove a lower bound on the maximum capacity of $(1,1)$-recoverable systems.

Theorem III.16 *Let Q be a finite alphabet of size $q=t^2-r$ with $t=\lceil\sqrt{q}\rceil$ and $r\leqslant t$. Then*

$$C_q(1,1)\geqslant \log_q\left(\frac{1}{2}\left(t-1+\sqrt{(t-1)^2+4(t-r)}\right)\right). \tag{4}$$

Proof: We prove the theorem by explicitly calculating the capacity of the system X presented by the graph $G=(V,E,L)$ with adjacency matrix A_Q. First observe that according to Corollary III.6, the system X is indeed a $(1,1)$-recoverable system.

For $r=t$, Proposition III.14 implies that $\lambda=t-1$ and $\mathrm{cap}(X)=\log_q(t-1)$. For $r<t$, Lemma III.12 states that A_Q is strongly connected, which implies that $\mathrm{cap}(X)=\log_q\lambda$, where λ is the maximal eigenvalue of A_Q. Define a vector $v\in\mathbb{R}^q$ as follows. For $i\in[q]$ and some real number ξ, to be chosen later, let

$$v_i=\begin{cases}\xi & \text{if } i \bmod t = t-1\\ 1 & \text{otherwise}\end{cases}.$$

Recall that A_Q has two types of rows and refer to Figure 1 for an illustration. We obtain

$$(A_Q v)_i=\begin{cases}t-r & \text{if } i \bmod t = t-1\\ t-1+\xi & \text{otherwise}\end{cases}.$$

Let us adjust $\xi\in\mathbb{R}$ so that v be an eigenvector of A_Q. If λ is an eigenvalue, then $A_Q v=\lambda v$ implies the following two relations

$$t-r=\lambda\xi$$
$$t-1+\xi=\lambda,$$

which are satisfied by

$$\xi_{1,2}=\frac{-(t-1)\pm\sqrt{(t-1)^2+4(t-r)}}{2},$$
$$\lambda_{1,2}=\frac{(t-1)\pm\sqrt{(t-1)^2+4(t-r)}}{2}.$$

From Proposition III.13, $\lambda_{1,2}$ are the only non-zero eigenvalues. Hence, λ_1 is the maximum eigenvalue which finishes the proof. □

Even though the de Bruijn graph construction yields $(1,1)$-recoverable systems for all values of q, we were able to compute the resulting capacity only for q that satisfy $t^2-t\leqslant q\leqslant t^2$ for all $t\geqslant 2$. A quick calculation from (4) yields that

$$C_q(1,1)\geqslant\log_q(t-1)\geqslant\frac{1}{2}-\frac{1}{(t-1)\ln q}$$

(we used a standard inequality $\ln(1+x)\geqslant x/(1+x)$ valid for $x>-1$). Together with the recursion in Sec. III-A this yields lower bounds on $C_q(1,1)$ for all values of $q\geqslant 4$. At the same time, the best bound is obtained if the recursion starts from the largest whole square smaller than q.

While the full comparison of the two bounds obtained in this section is cumbersome, in Fig. 2 we show some results for small values of q.

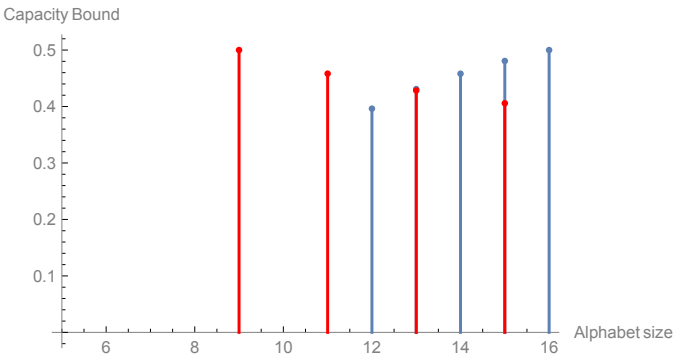


Figure 2: The bounds obtained in Sec. III for small values of q. The bound of Theorem III.16 is plotted for $q = 12, \ldots, 16$ and the bound of Lemma III.2 is plotted for $q = 9, 11, 13, 15$. To compute the bound of Lemma III.2 we started with $q = 9$ (with capacity $1/2$) and used it to find the capacity bounds for $q = 11, 13, 15$.

IV. ϵ-RECOVERABLE SYSTEMS

A natural relaxation of the recoverability requirement is to allow a small fraction of failures in the recovery of the window of size k in the sequence. In the first part of the section we formalize this concept using a measure-theoretic definition of recoverable systems.

A. Definition and general properties

Consider the alphabet Q together with the discrete σ-algebra and let $Q^{\mathbb{Z}}$ be the space of bi-infinite sequences endowed with the product Borel σ-algebra (under the product topology, $Q^{\mathbb{Z}}$ is compact and metrizable). For a sequence $x \in Q^{\mathbb{Z}}$ and a set $F \subseteq \mathbb{Z}$ we denote by x_F the projection of x on the coordinates in F, $x_F \in Q^F$. For a word $w \in Q^F$, denote by $[\![w]\!]_F$ the cylinder set

$$[\![w]\!]_F := \left\{x \in Q^{\mathbb{Z}} \ : \ x_F = w\right\}.$$

In particular, $[\![x]\!]_F$ denotes the cylinder set $[\![x_F]\!]_F$ of all sequences in $Q^{\mathbb{Z}}$ that coincide with x on F. For $F = [n]$ we will write $[\![w]\!]$ instead of $[\![w]\!]_{[n]}$.

For a set $X \subseteq Q^{\mathbb{Z}}$, we denote by $\mathcal{M}(X)$ the set of all probability measures on X. We will use the following subsets of $\mathcal{M}(Q^{\mathbb{Z}})$:

- (i) $\mathcal{M}_s(Q^{\mathbb{Z}})$ denotes the set of all shift invariant measures;
- (ii) $\mathcal{M}_r^{(k,l)}(Q^{\mathbb{Z}})$, all shift invariant (k,l)-recoverable measures (Def. IV.1);
- (iii) $\mathcal{M}_r^{(\epsilon,k,l)}(Q^{\mathbb{Z}})$, all (ϵ,k,l)-recoverable measures (Def. IV.4).

For a finite set $F \subset \mathbb{Z}$, we denote by μ_F a measure on Q^F defined as the F-marginal of μ. If $F = \{i\}$ is a singleton, we will write μ_i instead of $\mu_{\{i\}}$.

The measure-theoretic counterpart of capacity is called entropy and is defined next. Let $\mu \in \mathcal{M}_s(Q^{\mathbb{Z}})$ and let $F \subseteq \mathbb{Z}$. Recall that the **Shannon entropy**

$$H_\mu(F) = -\int \log_q \mu([\![x]\!]_F) d\mu(x).$$

As above, we will write $H_\mu(i)$ instead of $H_\mu(\{i\})$. The **entropy** of the measure is defined as

$$h(\mu) := \lim_{n\to\infty} \frac{1}{n} H_\mu([n]) \tag{5}$$

where the existence of the limit is a classic result in ergodic theory, e.g., [26, Thm. 4.22].

Let $\mathcal{A}$ be a μ-measurable σ-algebra. For a set $F \subseteq \mathbb{Z}$ we let

$$H_\mu(F|\mathcal{A}) = -\int \log_q \mu([\![x]\!]_F|\mathcal{A}) d\mu(x) \tag{6}$$

and define the conditional entropy with respect to $\mathcal{A}$ as

$$h(\mu|\mathcal{A}) := \lim_{n\to\infty} \frac{1}{n} H_\mu([n]|\mathcal{A}).$$

For a set of coordinates $F \subseteq \mathbb{Z}$, we denote by $\sigma(F)$ the σ-algebra generated by the coordinates in F (by all the cylinder sets of the form $[w]_F$).

Definition IV.1 *Let $\mu \in \mathcal{M}_s(Q^{\mathbb{Z}})$ be a shift invariant measure. For $k, l \in \mathbb{N}$, we say that μ is a (k,l)-recoverable measure if*

$$H_\mu([k]|\sigma(N_l([k]))) = 0.$$

The set of all shift invariant, (k,l)-recoverable measures over $Q^{\mathbb{Z}}$ is denoted by $\mathcal{M}_r^{(k,l)}(Q^{\mathbb{Z}})$.

For a measure $\mu \in \mathcal{M}_r^{(k,l)}$, shift-invariance implies that $H_\mu(j + [k] \mid \sigma(j + N_l([k]))) = 0$ for every $j \in \mathbb{Z}$, where $N_l([k])$ is the l-neighborhood of the set $[k]$. Since $X_{[k]}$ can take only a finite number of values (at most q^k), the recoverability property is equivalent to the property that $X_{[k]}$ is measurable with respect to $\sigma(N_l([k]))$.

Let $X_{\mathcal{F}} \subseteq Q^{\mathbb{Z}}$ be a constrained system. The capacity of $X_{\mathcal{F}}$ relates to the notion of entropy via the **variational principle** [26, Thm. 8.6] which states that

$$\mathrm{cap}(X_{\mathcal{F}}) = \sup_{\mu \in \mathcal{M}_s(X_{\mathcal{F}})} h(\mu).$$

To connect the definition of a recoverable system given earlier in the introduction with Definition IV.1, we recall the measure-theoretic approach to the capacity of constrained systems [19, Sec.3.2.3].

Construction IV.2 (Markov measure) *Let $X_{\mathcal{F}}$ be a constrained system presented by a graph $G = (V, E, L)$ and let A_G be its adjacency matrix. Suppose that G is strongly connected. The Perron-Frobenius theorem implies that the left and right eigenvectors with the largest eigenvalue of A_G are positive (have strictly positive coordinates). Denote these vectors by x and y and suppose w.l.o.g. that $x^T y = 1$. Define a matrix P by setting*

$$P_{u,v} = \frac{(A_G)_{(u,v)} y_v}{\lambda y_u}, \quad u, v \in V.$$

Define a Markov measure μ on $X_{\mathcal{F}}$ using P as the transition matrix and the vector $p := (x_v y_v, v \in V)$ as the starting distribution. It suffices to define μ on the cylinder sets $[\![w]\!]$. For every $m, n \in \mathbb{N}$, let $F = m + [n] \subseteq \mathbb{Z}$. For every path $\gamma = (e_0, \ldots, e_{n-1})$ in G with $w = L(\gamma)$ simply put

$$\mu([\![w]\!]) = p_{v_0} P_{v_0,v_1} \cdots P_{v_{n-1},v_n} \tag{7}$$

where $e_i = (v_i, v_{i+1})$. This completes the construction. □

The next theorem derives from results on the eigenvalues of the matrix A_G, proved in [19, Thm. 3.9] for primitive irreducible matrices and in [18], Exercise 4.5.14 without the primitivity assumption. The maximality of $h(\mu)$ is also discussed in [18, p. 443].

Theorem IV.3 *The measure μ is supported on the set $X_{\mathcal{F}}$ and is the unique measure for which $h(\mu) = \mathrm{cap}(X_{\mathcal{F}}) = \log \lambda$.*

This theorem together with the variational principle implies that for every $k, l \in \mathbb{N}$, there exists a shift invariant (k,l)-recoverable measure $\mu \in \mathcal{M}_r^{(k,l)}(Q^{\mathbb{Z}})$ with maximum entropy,

$$h_q^*(k,l) := \max_{\mu \in \mathcal{M}_r^{(k,l)}(Q^{\mathbb{Z}})} h(\mu).$$

In addition, any (k,l)-recoverable measure μ that maximizes the entropy is Markov with memory $2l+k-1$. Therefore, we will think of a system X as a discrete-time stochastic process $X = (\dots, X_{-1}, X_0, X_1, \dots)$. With this in mind, for a finite set $F = \{n_0, \dots, n_k\} \subseteq \mathbb{Z}$, we have

$$\begin{aligned} &\Pr(X_{n_0} = x_{n_0}, \dots, X_{n_k} = x_{n_k}) \\ &= \mu\left([\![x_{n_0} \dots x_{n_k}]\!]_F\right) \\ &= \mu_F\left(x_{n_0} \dots x_{n_k}\right) \end{aligned} \tag{8}$$

and the Shannon entropy takes the form

$$H_\mu(X_{n_0}, \dots, X_{n_k}) = -\int \log_q \mu_F(x)\, d\mu_F(x).$$

For a set $G = \{m_0, \dots, m_l\} \subseteq \mathbb{Z}$ we denote by $H_\mu(X_{n_0}, \dots, X_{n_k} \mid X_{m_0}, \dots, X_{m_l})$ the conditional Shannon entropy with respect to the σ-algebra generated by the pre-image of the random variables $X_{m_0}, \dots, X_{m_l}$.

Suppose that the random variables $\{X_j\}_j$ form the Bernoulli scheme, i.e., X_j are i.i.d with distribution $P(X = i) = p_i$, where $\boldsymbol{p} = (p_0, \dots, p_{q-1})$ is a probability vector. In this case, instead of $H_\mu(X_0)$, we will write

$$H_q(\boldsymbol{p}) = -\sum_{i \in [q]} p_i \log_q p_i$$

for the Shannon entropy. In the particular case that the vector $\boldsymbol{p}$ has only two nonzero entries, say p and $1-p$, we simplify the notation and write $H_q(p)$.

Let us define the ϵ-recoverability property.

Definition IV.4 *Let $\mu \in \mathcal{M}_s(Q^{\mathbb{Z}})$ and let $\epsilon > 0$. We say that μ is an (ϵ, k, l)-**recoverable measure** if for every $\alpha, \beta \in Q^l$,*

$$H_\mu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \leqslant \epsilon. \tag{9}$$

We say that μ is ϵ-recoverable if there exist some numbers k, l such that μ is (ϵ, k, l)-recoverable. The set of all (ϵ, k, l)-recoverable measures on $Q^{\mathbb{Z}}$ is denoted by $\mathcal{M}_r^{(\epsilon,k,l)}(Q^{\mathbb{Z}})$.

The problem that we consider in this section is to characterize (ϵ, k, l)-recoverable measures with maximum entropy. At this moment however, it is not clear that this problem is well posed. Indeed, the support of an ϵ-recoverable measure is not necessarily a constrained system since such a measure in principle can assign a positive probability to any word (cylinder set) $w \in Q^*$. At the same time, a constrained system is defined by a set of forbidden patterns $\mathcal{F} \subseteq Q^*$ which therefore should have probability zero. For this reason, it is not immediately clear if there exists an (ϵ, k, l)-recoverable measure μ that maximizes the entropy $h(\mu)$ in the class of such measures. In the following we show that such a measure indeed exists and describe some of its properties.

We observe the following.

1) For every finite set $F \subset \mathbb{Z}$, the coordinate restriction map π_F given by $\mu \mapsto \mu_F$ is continuous. Indeed, for every $w \in Q^F$ we have

$$\begin{aligned} &|\pi_F(\mu_1)(w) - \pi_F(\mu_2)(w)| \\ &= |\mu_1([\![w]\!]) - \mu_2([\![w]\!])| \\ &= (\mu_1 - \mu_2)([\![w]\!]). \end{aligned}$$

   Thus, $|\pi_F(\mu_1)(w) - \pi_F(\mu_2)(w)| \leqslant \|\mu_1 - \mu_2\|_{TV}$ for every cylinder set, which implies that $\|\pi_F(\mu_1) - \pi_F(\mu_2)\| \leqslant \|\mu_1 - \mu_2\|_{TV}$.
2) The Shannon entropy of a discrete distribution μ_F is a continuous function $H_q(\mu_F) \to [0,1]$.
3) The conditioning operation $\mu(\cdot) \mapsto \mu(\cdot \mid A)$ is continuous as a mapping of measures. Below we denote this operation by Con_A, where A is an event.

From 2), for every $\alpha, \beta \in Q^l$, the inverse image of the closed set $[0, \epsilon]$ under the Shannon entropy gives a closed set of (conditional) measures on k-tuples, i.e., the set

$$M_{\alpha,\beta} := H^{-1}_{X_{l+[k]} | X_{[l]} = \alpha, X_{l+k+[l]} = \beta}([0, \epsilon]) \subseteq \mathcal{M}(Q^k)$$

is closed. Let A be the event

$$\left\{ x \in Q^{[2l+k]} \ : \ x_{[l]} = \alpha, x_{l+k+[l]} = \beta \right\}.$$

From 3) we obtain that for every $\alpha, \beta \in Q^l$, the set $\mathsf{Con}_A^{-1}(M_{\alpha,\beta}) \subseteq \mathcal{M}(Q^{2l+k})$ is also closed. This implies that the finite intersection

$$M := \bigcap_{\alpha,\beta \in Q^l} \mathsf{Con}_A^{-1}(M_{\alpha,\beta}) \subseteq \mathcal{M}(Q^{2l+k})$$

is also closed. Upon setting $F = [2l+k]$, 1) implies that $\pi_F^{-1}(M) \subseteq \mathcal{M}(Q^{\mathbb{Z}})$ is a closed set. Since $\mathcal{M}_s(Q^{\mathbb{Z}})$ is closed and compact, taking the intersection $\mathcal{M}_s(Q^{\mathbb{Z}}) \cap \pi_F^{-1}(M)$ we obtain a closed (and hence compact) set. This is exactly the set of all (ϵ, k, l)-recoverable measures. The shift operator T is continuous, which implies that the function $\mu \to h(\mu)$ is affine [26, Thm. 8.1]. Finally, an affine function on a compact set attains its supremum.

We conclude that the problem stated above after Definition IV.4 is well posed. Therefore, we introduce the following notation: let

$$h_q^*(\epsilon, k, l) := \max_{\mu \in \mathcal{M}_r^{(\epsilon,k,l)}(Q^{\mathbb{Z}})} h(\mu).$$

The next lemma argues that this maximum is attainable by a Markov measure, which also implies that the entropy $h_q^*(\epsilon, k, l)$ is finite [26, Thm. 4.27].

Lemma IV.5 *Let $\epsilon > 0$ and let μ be an (ϵ, k, l)-recoverable measure with $h(\mu) = h_q^*(\epsilon, k, l)$. Then there exists a Markov measure ν with memory $2l+k-1$ such that $h(\nu) = h_q^*(\epsilon, k, l)$.*

Proof: We start by explaining the idea of the proof. To shorten the notation, we limit ourselves to the case $k = l = 1$, but the general proof follows the same arguments. The proof follows the known construction of measure extensions [15], [23]. The key point is that the (ϵ, k, l)-recoverable property is completely determined by the marginal distribution $\mu_{[2l+k]}$. We will construct a memory-2 Markov measure ν with the same distribution over Q^3 as μ, sometimes termed the 2nd order Markov approximation of μ. Since $\nu_{[3]} = \mu_{[3]}$ we can claim that ν is an $(\epsilon, 1, 1)$-recoverable measure, and show that ν maximizes the entropy.

For every $m \in \mathbb{N}$, define a measure ν_m over Q^m as follows. For $1 \leqslant m \leqslant 3$ and for every $w \in Q^m$ define $\nu_m([\![w]\!]) = \mu([\![w]\!])$. For $m = 4$ and every $w = (w_0, w_1, w_2, w_3) \in Q^m$ define

$$\nu_m([\![w]\!]) = \mu([\![w_0 w_1 w_2]\!])\mu([\![w_1 w_2 w_3]\!]|[\![w_1 w_2]\!])$$

and for $w = (w_0, \ldots, w_{m-1}) \in Q^m$, $m > 4$ define ν_m recursively in a similar fashion:

$$\begin{aligned}\nu_m([\![w]\!]) = \nu_{m-1}([\![w_{[m-1]}]\!])\cdot \\ \mu([\![w_{m-3}w_{m-2}w_{m-1}]\!]|[\![w_{m-3}w_{m-2}]\!]).\end{aligned} \tag{10}$$

Since ν_m is defined for every $m \in \mathbb{N}$, by the Kolmogorov extension theorem there exists a (unique) measure ν over $Q^{\mathbb{Z}}$ with ν_m as its marginals. Specifically, $\nu_{[3]} = \mu_{[3]}$, which implies that ν is $(\epsilon, 1, 1)$-recoverable. Since μ is shift invariant, so is ν, and by construction ν is Markov with memory 2. To finish the proof, we show that ν maximizes the entropy over all shift invariant measures that coincide with μ on Q^3. First notice that for $m \leqslant 3$ we have $H_\mu([m]) = H_\nu([m])$, where $H_\mu([m]) = H_\mu([\![x_{[m]}]\!])$, for instance, is the Shannon entropy of the cylinder set. For $m > 3$ we have

$$\begin{aligned}H_\nu([m]) &= H_{\nu_m}([m]) \\ &= H_{\nu_{m-1}}([m-1]) + H_\mu([3] \mid [2]) \\ &= H_{\nu_{m-1}}([m-1]) + H_\mu([3]) - H_\mu([2])\end{aligned}$$

where the first equality follows from (10) and shift-invariance of μ, and the second uses the chain rule for entropies $H_\mu([3]) = H_\mu([2]) + H_\mu([3] \mid [2])$. Here the notation $H_\mu([3] \mid [2])$ means $H_\mu([3] \mid \sigma([2]))$, i.e., conditioning on the σ-algebra generated by the coordinates $0, 1$ in the sequence w (see (6)). We obtain

$$H_\nu([m]) - H_{\nu_{m-1}}([m-1]) = H_\mu([3]) - H_\mu([2]). \tag{11}$$

Now consider $H_\mu([m])$ for some $m \in \mathbb{N}$. Using again the chain rule, we obtain

$$H_\mu([m]) = H_\mu([m-1]) + H_\mu(\{m-1\}|[m-1])$$

and since $H_\mu(\{m-1\}|[m-1]) \leqslant H_\mu(\{m-1\}|\{m-3, m-2\})$ (conditioning on a sub-σ-algebra reduces entropy), we have

$$\begin{aligned}&H_\mu([m]) \leqslant \\ &H_\mu([m-1]) + H_\mu(\{m-1\}|\{m-3, m-2\}).\end{aligned}$$

Since μ is shift invariant, we obtain

$$H_\mu([m]) \leqslant H_\mu([m-1]) + H_\mu(\{2\}|[2]). \tag{12}$$

We also have

$$H_\mu([3]) = H_\mu([2]) + H_\mu(\{2\}|[2]).$$

Subtracting this equality from (12), we obtain

$$H_\mu([m]) - H_\mu([m-1]) \leqslant H_\mu([3]) - H_\mu([2]).$$

Taken together with (11), this inequality implies that ν maximizes the entropy over all shift invariant measures whose length-3 marginals agree with μ. □

Next we show that $h_q^*(\epsilon, k, l)$ is obtained when the relaxation is exploited to its full extent. This fact is formally stated in the next theorem, whose proof draws in part on the ideas of [7]. We need to introduce additional elements of notation. Let $\mathbb{Z}^-$ denote the set of all negative integers and let $\mathcal{C}$ ($\mathcal{C}^-$) denote the σ-algebra generated by $\mathbb{Z}$ ($\mathbb{Z}^-$), i.e., by cylinders $[\![w]\!]_F$ with $F \subseteq \mathbb{Z}$ or $F \subseteq \mathbb{Z}^-$, respectively. Since we consider only shift invariant measures, we have $h(\mu) = H_\mu(0 \mid \mathcal{C}^-)$ [14, Eq. (15.19)]. Using shift invariance and the chain rule, we obtain that for every $n \in \mathbb{N}$,

$$h(\mu) = \frac{1}{n} H_\mu([n] \mid \mathcal{C}^-). \tag{13}$$

We remark that this general result applies to all shift invariant measures, whereas we use it for Markov measures. We could limit the conditioning in (13) to $2l + k - 1$ coordinates in the past, but the arguments below do not depend on this.

Theorem IV.6 *Fix $\epsilon > 0$ and let μ be an (ϵ, k, l)-recoverable measure with $h(\mu) = h_q^*(\epsilon, k, l)$. Then there exist $\alpha, \beta \in Q^l$, such that*

$$H_\mu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) = \epsilon.$$

Proof: Assume toward a contradiction that for every $\alpha, \beta \in Q^l$, $H_\mu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) < \epsilon$. Let α, β be words such that $\sum_{w \in Q^k} \mu_{[2l+k]}(\alpha w \beta) > 0$ and that $H_\mu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta)$ is maximal. We will show that $H_\mu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) = \epsilon$.

Let $p = (p_0, \ldots, p_{q^k-1})$ be a probability distribution on Q^k with $H_q(p) = \epsilon$. Using μ and p, we will construct a measure ν that has higher entropy than μ. Define ν independently on segments of length m in $\mathbb{Z}$, where the value of m will be specified later. To sample a sequence $x \in Q^{\mathbb{Z}}$ from ν, choose $x \in Q^{\mathbb{Z}}$ according to μ and then, if $x_{[l]} = \alpha, x_{l+k+[l]} = \beta$ then replace $x_{l+[k]}$ with a word chosen according to p. Repeat this procedure independently for every $j \in \mathbb{Z}$: if $x_{jm+[l]} = \alpha, x_{jm+l+k+[l]} = \beta$, then replace $x_{jm+l+[k]}$ with a word chosen according to p. The resulting measure ν is invariant under T^m but not necessarily shift invariant. To make it shift invariant, consider $\nu_s = \frac{1}{m}\sum_{i \in [m]} T^i(\nu)$ where $T(\nu)$ is the push-forward of ν, defined by $T(\nu)(\cdot) = \nu\left(T^{-1}(\cdot)\right)$.

Let us show that ν_s is an (ϵ, k, l)-recoverable measure. It suffices to show that for every $w_1, w_2 \in Q^l$, $H_{\nu_s}(X_{l+[k]} \mid X_{[l]}X_{l+k+[l]} = w_1 w_2) \leqslant \epsilon$. Note that for $l + k \leqslant i \leqslant m - 2l - k$ we have

$$\begin{aligned}&\nu(X_{i+l+[k]} \mid X_{i+[l]}X_{i+l+k+[l]} = w_1 w_2) = \\ &\mu(X_{i+l+[k]} \mid X_{i+[l]}X_{i+l+k+[l]} = w_1 w_2).\end{aligned}$$

By definition, we have

$$T^i(\nu)(X_{i+l+[k]} \mid X_{i+[l]}X_{i+l+k+[l]} = w_1w_2) = \nu\left(X_{l+[k]} \mid X_{[l]}X_{l+k+[l]} = w_1w_2\right).$$

Since the conditional entropy is continuous and since for every $w_1, w_2 \in Q^l$, $H_\mu(X_{l+[k]} \mid X_{[l]}X_{l+k+[l]} = w_1w_2) < \epsilon$, we can choose m large enough such that $H_{\nu_s}(X_{l+[k]} \mid X_{[l]}X_{l+k+[l]} = w_1w_2) \leqslant \epsilon$. Generally, the value of m depends on w_1, w_2, but there are finitely many possibilities to choose them, so it is possible to choose a large enough, but finite, m such that ν_s is an (ϵ, k, l)-recoverable measure.

Since μ maximizes the entropy, we have that

$$h(\nu_s) \leqslant h(\mu). \tag{14}$$

Next we define a partition $\mathcal{A}^m = \mathcal{A}_1^m \cup \mathcal{A}_2^m$ of $Q^{\mathbb{Z}}$ according to the values of x in the coordinates in $[m]$. Let $r = m - 2l - k$ and note that $[m] = \cup_{i=1}^4 J_i$, where $J_1 = [l], J_2 = \{l + [k]\}, J_3 = \{l + k + [l]\}, J_4 = \{2l + k + [r]\}$. The class $\mathcal{A}_1^m$ is formed of cylinder sets $[\![x_{[m]}]\!]$ such that $x_{J_1} \neq \alpha$ or $x_{J_3} \neq \beta$. The class $\mathcal{A}_2^m$ is formed of cylinder sets $[\![x_{[m]}]\!]$ such that $x_{J_1} = \alpha, x_{J_3} = \beta$ and $x_{J_4} = w$, where $w \in Q^r$. Formally,

$$\mathcal{A}_1^m = \bigcup_{u \in Q^{[m]},\ u_{J_1}u_{J_3} \neq \alpha\beta} [\![u]\!]$$

$$\mathcal{A}_2^m = \bigcup_{w \in Q^r} A_2^m(w), \quad A_2^m(w) = \bigcup_{v \in Q^k} [\![\alpha v \beta w]\!].$$

Thinking of the system as a sequence of random variables (8), let us rewrite (13) as

$$h(\mu) = \frac{1}{m} H_\mu(X_{[m]} \mid \mathcal{C}^-). \tag{15}$$

Now consider a function f that takes $x_{[m]}$ to its corresponding part in $\mathcal{A}^m$. Namely, if $x_{J_1} \neq \alpha$ or $x_{J_3} \neq \beta$, then f takes $x_{[m]}$ to itself. Otherwise, f takes $x_{[m]}$ to the part defined by $x_{J_1}, x_{J_3}, x_{J_4}$, i.e., f does not distinguish between m-words for which $x_{J_1} = \alpha, x_{J_3} = \beta, x_{J_4} = w$.

Clearly, we have

$$H_\mu(f(X_{[m]}) \mid \mathcal{C}^-, X_{[m]}) = 0,$$

which together with the chain rule implies that

$$H_\mu(X_{[m]}, f(X_{[m]}) \mid \mathcal{C}^-) = H_\mu(X_{[m]} \mid \mathcal{C}^-).$$

Therefore,

$$\begin{aligned} &H_\mu(X_{[m]} \mid \mathcal{C}^-) = H_\mu(X_{[m]}, f(X_{[m]}) \mid \mathcal{C}^-) \\ &= H_\mu(f(X_{[m]}) \mid \mathcal{C}^-) + H_\mu(X_{[m]} \mid \mathcal{C}^-, f(X_{[m]})). \end{aligned} \tag{16}$$

For two partitions $\mathcal{G}_1, \mathcal{G}_2$ of a set, their refinement is defined as $\mathcal{G}_1 \vee \mathcal{G}_2 = \{g_1 \cap g_2 : g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2\}$. Consider the partition $\vee_{j \in \mathbb{Z}} T^{jm}(\mathcal{A}^m)$ and let $\mathcal{A}$ be the σ-algebra generated by it. Notice that $\mathcal{A}$ is a sub-σ-algebra of $\mathcal{C}$ under which $X_{[m]}$ and $f(X_{[m]})$ cannot be distinguished. Further, let $\mathcal{A}^-$ be the sub-σ-algebra of $\mathcal{A}$ obtained by the restriction of $\mathcal{A}$ to $\mathbb{Z}^-$. Noticing that $\mathcal{C}^-$ is a refinement of $\mathcal{A}^-$, we obtain

$$\begin{aligned} H_\mu(f(X_{[m]}) \mid \mathcal{C}^-) &\leqslant H_\mu(f(X_{[m]}) \mid \mathcal{A}^-) \\ &= H_\nu(f(X_{[m]}) \mid \mathcal{A}^-) \\ &= H_\nu(f(X_{[m]}) \mid \mathcal{C}^-), \end{aligned} \tag{17}$$

where on the second line we use the fact that the distribution of $f(X_{[m]})$ is the same under μ and ν. To justify (17), we notice that $\mathcal{A}$ is the Borel σ-algebra of the system after applying f and it is generated by $\mathcal{A}^m$. Therefore, according to the Kolmogorov-Sinai theorem [25], [26, Thm. 4.17], the entropy of $f(X_{[m]})$ does not change if instead of $\mathcal{A}^-$ we condition on $\mathcal{C}^-$. Intuitively, the last equality follows since $\mathcal{A}$ contains all the information needed to describe $f(X_{[m]})$ explicitly.

Substituting (17) into (16), we obtain

$$\begin{aligned} &H_\mu(X_{[m]} \mid \mathcal{C}^-) \\ &\leqslant H_\nu(f(X_{[m]}) \mid \mathcal{C}^-) + H_\mu(X_{[m]} \mid \mathcal{C}^-, f(X_{[m]})) \\ &\leqslant H_\nu(f(X_{[m]}) \mid \mathcal{C}^-) + H_\mu(X_{[m]} \mid f(X_{[m]})). \end{aligned}$$

By the construction of ν and the assumption about p we have

$$\begin{aligned} &H_\mu(X_{[m]} \mid f(X_{[m]})) \leqslant H_\nu(X_{[m]} \mid f(X_{[m]})) \\ &= H_\nu(X_{[m]} \mid \mathcal{C}^-, f(X_{[m]})) \end{aligned} \tag{18}$$

(the past is independent of $X_{l+[k]}$ conditional on $f(X_{[m]})$), which implies

$$\begin{aligned} &H_\mu(X_{[m]} \mid \mathcal{C}^-) \leqslant \\ &H_\nu(f(X_{[m]}) \mid \mathcal{C}^-) + H_\nu(X_{[m]} \mid \mathcal{C}^-, f(X_{[m]})). \end{aligned}$$

Using (15), we obtain that

$$\begin{aligned} &h(\mu) \leqslant \\ &\frac{1}{m}(H_\nu(f(X_{[m]}) \mid \mathcal{C}^-) + H_\nu(X_{[m]} \mid \mathcal{C}^-, f(X_{[m]}))). \end{aligned} \tag{19}$$

Since ν is invariant with respect to T^m (i.e., invariant under a subgroup of $\mathbb{Z}$), monotonicity of the entropy implies that the limit in the definition of $h(\nu)$ exists (see (5)). Since ν_s is shift invariant with respect to $\mathbb{Z}$, we conclude that $h(\nu_s) = h(\nu)$. Using the chain rule in (13), we obtain that

$$\begin{aligned} &h(\nu_s) = \\ &\frac{1}{m} H_\nu(f(X_{[m]}) \mid \mathcal{C}^-) + \frac{1}{m} H_\nu(X_{[m]} \mid \mathcal{C}^-, f(X_{[m]})). \end{aligned}$$

Therefore, on account of (14), all the inequalities in (17)-(19) are equalities. Noticing that if $X_{[l]}X_{l+k+[l]} \neq \alpha\beta$, then $H_\mu(X_{[m]} \mid f(X_{[m]})) = 0$, we conclude from (18) that

$$\begin{aligned} &H_\mu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \\ &= H_\nu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \\ &= \epsilon. \end{aligned}$$

This contradicts the assumption on μ. □

Remark IV.7 *An informal justification of the claim of Theorem IV.6 is as follows. Let μ be a measure on $Q^{\mathbb{Z}}$ such that for every $\alpha, \beta \in Q^l$, the conditional entropy $H_\mu(X_{l+[k]} | X_{[l]} = \alpha, X_{l+k+[l]} = \beta) < \epsilon$. Now consider a measure ν that is a mixture of μ with an i.i.d. uniform distribution on $Q^{\mathbb{Z}}$, viz., $\nu = (1-r)\mu + r\eta$ where $r \in [0, 1]$ and η is a measure on $Q^{\mathbb{Z}}$ such that the entries are distributed i.i.d uniform. The value of r can be chosen such that $\nu_{[2l+k]}$ is arbitrarily close to $\mu_{[2l+k]}$ in the total variation distance. By continuity of entropy we can choose $r > 0$ such that*

$$\max_{\alpha,\beta \in Q^l} H_\nu(X_{l+[k]} | X_{[l]} = \alpha, X_{l+k+[l]} = \beta) < \epsilon.$$

Since the entropy is strictly concave, $h(\nu) > h(\mu)$, contradicting the maximality of μ. A similar approach is used below to generate $(\epsilon,1,1)$-recoverable systems by increasing the entropy of $(1,1)$-recoverable systems.

The following proposition implies that measures with a more relaxed recovery requirement have higher metric entropy.

Proposition IV.8 *Let $k,l \geqslant 1$ and $\epsilon_1 > \epsilon_2 > 0$, then*

$$h_q^*(\epsilon_1,k,l) > h_q^*(\epsilon_2,k,l).$$

Proof: Let $\mu_i, i=1,2$ be such that $h(\mu_i) = h_q^*(\epsilon_i,k,l)$. Clearly, $h(\mu_1) \geqslant h(\mu_2)$ because μ_2 is (ϵ_1,k,l)-recoverable. If this inequality holds with equality, then by Theorem IV.6 there are $\alpha,\beta \in Q^l$ such that

$$H_{\mu_2}(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) = \epsilon_1 > \epsilon_2,$$

a contradiction. □

The actual recovery procedure of a k-tuple from its l-neighborhood is rather straightforward and reduces to taking the k-word that has the largest conditional probability with respect to the measure μ. This statement will be made more formal once we prove Lemma IV.10 in the next section.

B. Constructing (ϵ,k,l)-recoverable measures

To construct recoverable measures we will rely on (k,l)-recoverable systems. It is intuitive that for small ϵ, the capacity $h^*(\epsilon,k,l)$ should approach the capacity of recoverable systems $C_q(k,l)$. In this section we formalize this intuition, transforming (k,l)-recoverable systems into ϵ-recoverable measures. This approach has its limitations in the sense that it works only for small values of ϵ.

We start with a well-known bound on the binary entropy $H_2(x)$.

Lemma IV.9 *For all $x \in [0,1]$*

$$H_2(x) \geqslant 4x(1-x).$$

Proof: Consider the function $f(x) = \frac{H_2(x)}{x(1-x)}$. It is immediate that $f'(x) \leqslant 0$ for $x \in [0,1/2]$, and so $f(x) \geqslant f(1/2) = 4$, or $H_2(x) \geqslant 4x(1-x)$. By symmetry around 1/2 we can claim this inequality also for $1/2 < x \leqslant 1$. □

In the next lemma we state a simple property of recoverable measures, namely that small conditional entropy implies that the symbols can be recovered with high probability. In the statement, $[\![\alpha w \beta]\!]$ refers to a cylinder set with $\alpha,\beta \in Q^l$ and $w \in Q^k$.

Lemma IV.10 *Let X be distributed according to a measure $\mu \in \mathcal{M}_s(Q^{\mathbb{Z}})$ that is (ϵ,k,l)-recoverable, and suppose that $0 < \epsilon < \frac{2}{q^k}$. Then for every $\alpha,\beta \in Q^l$ with $\sum_{w' \in Q^k} \mu([\![\alpha w' \beta]\!]) > 0$, there exists a unique $w \in Q^k$ such that $\Pr(X_{l+[k]} = w \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \geqslant 1 - \frac{\epsilon}{2}$.*

Proof: We begin with the case $k = l = 1$ and $q = 2$. Since μ is $(\epsilon,1,1)$-recoverable, we have that X_1 conditioned on $X_0 = \alpha, X_2 = \beta$ has a Bernoulli(p) distribution for some $p \in [0,1]$, such that $H_\mu(X_1 \mid X_0 = \alpha, X_2 = \beta) = H_2(p) \leqslant \epsilon$. By Lemma IV.9 we have

$$\epsilon \geqslant H_2(p) \geqslant 4p(1-p).$$

which implies that

$$p \geqslant \frac{1+\sqrt{1-\epsilon}}{2} \qquad \text{or} \qquad p \leqslant \frac{1-\sqrt{1-\epsilon}}{2}.$$

Since $\sqrt{1-\epsilon} \geqslant 1-\epsilon$ we have

$$p \geqslant 1 - \frac{\epsilon}{2} \qquad \text{or} \qquad p \leqslant \frac{\epsilon}{2}.$$

This implies that there is a symbol among $\{0,1\}$ with probability $\geqslant 1 - \epsilon/2$.

Now let $q > 2$ and $k,l \in \mathbb{N}$ and let us reduce this general case to the binary case. Fix $\alpha,\beta \in Q^l$ and define the mapping $f : Q^k \to \{0,1\}$ as follows. Let $w \in Q^k$ be such that

$$\begin{aligned} &\Pr(X_{l+[k]} = w | X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \\ &\geqslant \Pr(X_{l+[k]} = w' | X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \\ &> 0 \end{aligned} \tag{20}$$

for every $w' \neq w$. Notice that such a word w exists since $\sum_{w' \in Q^k} \mu([\![\alpha w' \beta]\!]) > 0$. Let $w' \in Q^k$ and define

$$f(w') = \begin{cases} 0, & \text{if } w' = w \\ 1, & \text{otherwise.} \end{cases}$$

So $f(X_{l+[k]})$ can be regarded as a binary RV. Thus, conditional on the event that $X_{[l]} = \alpha, X_{l+k+[l]} = \beta$, the RV $f(X_{l+[k]})$ is distributed according to Bernoulli(p) for some $p \in [0,1]$. Therefore,

$$\begin{aligned} &H_\mu(f(X_{l+[k]}) \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \\ &\leqslant H_\mu(X_{l+[k]} \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \end{aligned}$$

which implies

$$H_\mu(f(X_{l+[k]}) \mid X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \leqslant \epsilon.$$

From the binary case analysis we obtain that

$$p \geqslant 1 - \frac{\epsilon}{2} \qquad \text{or} \qquad p \leqslant \frac{\epsilon}{2}.$$

Recall that w is chosen according to (20). Since

$$\begin{aligned} 1 = \sum_{w \in Q^k} &\Pr(X_{l+[k]} = w | X_{[l]} \\ &= \alpha, X_{l+k+[l]} = \beta) \leqslant q^k p \end{aligned}$$

This relation together with the assumption that $\epsilon < \frac{2}{q^k}$ implies that $p \geqslant 1 - (\epsilon/2)$. □

Remark IV.11 1) Lemma IV.10 implies that for every $\alpha,\beta \in Q^l$, there is a unique $w_{\alpha,\beta} \in Q^k$ such that $\mu([\![\alpha w_{\alpha,\beta} \beta]\!])$ is close to 1. Observe that the small entropy constraint translates into a bound on the error probability of recovery. Define

$$\begin{aligned} &p_e(X) := \\ &\max_{\alpha,\beta \in Q^l} P(X_{l+[k]} = w_{\alpha,\beta} | X_{[l]} = \alpha, X_{l+k+[l]} = \beta) \end{aligned}$$

to be the maximum error probability, then the condition of (ϵ,k,l)-recoverability clearly implies that $p_e \leqslant \epsilon/2$.

2) Define the set

$$\mathcal{F} = \bigcup_{\alpha,\beta \in Q^l} \{\alpha w \beta \ : \ w \neq w_{\alpha,\beta}\}.$$

The constrained system $X_{\mathcal{F}}$ with the set of forbidden words $\mathcal{F}$ is (k,l)-recoverable, and moreover, if μ is a measure for which $h(\mu) = h_q^*(\epsilon,k,l)$ then $\mathrm{cap}(X_{\mathcal{F}}) = C_q(k,l)$. Indeed, if not, then there is a constrained system Y with $\mathrm{cap}(Y) > \mathrm{cap}(X_{\mathcal{F}})$. For a sufficiently small ϵ the construction of $\mathcal{F}$ and continuity of entropy imply that $h(\mu)$ is arbitrarily close to $\mathrm{cap}(X_{\mathcal{F}})$. Since μ is such that $h(\mu)$ attains the value $h_q^*(\epsilon,k,l)$, and since Y gives rise to a measure ν with $h(\nu) = \mathrm{cap}(Y)$, we obtain that $h(\mu) \geqslant \mathrm{cap}(Y)$. This yields a contradiction, proving our claim.

Remark IV.11.2 raises the question whether this relation also applies in reverse direction, i.e., whether it is possible to obtain an (ϵ,k,l)-recoverable measure with maximum entropy from a (k,l)-recoverable system with maximum capacity. In the remainder of this section we establish a partial result toward the resolution of this question. Namely, given a (k,l)-recoverable system and $\epsilon > 0$, we construct an (ϵ,k,l)-recoverable measure and calculate its entropy, obtaining a lower bound on $h_q^*(\epsilon,k,l)$.

We begin with an informal description of the construction. The (ϵ,k,l)-recoverable measure will be constructed by perturbing a Markov measure associated with a maximum-capacity deterministic (k,l)-recoverable system. We start with a graph that presents a (k,l)-recoverable system $X_{\mathcal{F}}$ with maximum capacity $C_q(k,l)$. Since the recoverability property is defined by words of length $2l+k$, we use a presentation $G = (V,E,L)$ of $X_{\mathcal{F}}$ such that vertices in V correspond to words of length $2l+k$ over Q. This system is equivalently described by a Markov measure defined in (7). Denoting it by ρ, let P^ρ and p^ρ be the matrix of transition probabilities and the stationary distribution of ρ, respectively.

To describe the perturbation, it will be convenient to switch to a different Markov measure, which we proceed to define. Let Y be the system obtained from $X_{\mathcal{F}}$ by considering non-overlapping subwords of length $2l+k$ (also called the $(2l+k)$th higher power of $X_{\mathcal{F}}$ [18]). We can view Y as a system over the alphabet Q^{2l+k}. Define a Markov measure μ on $Y \subseteq (Q^{2l+k})^{\mathbb{Z}}$ obtained from ρ as follows. The state set of μ is the same as the states of ρ. The matrix of transition probabilities of ρ is obtained as $P^\mu = (P^\rho)^{2l+k}$, and the stationary distribution is unchanged, i.e., $p^\mu = p^\rho$. In other words, μ is obtained from the $(2l+k)$ power of the graph G (the graph G^{2l+k} on V whose edges correspond to paths of length $2l+k$ in G). Consider a system $Z \subseteq (Q^{2l+k})^{\mathbb{Z}}$ obtained by passing Y through a conditional distribution, given by a stochastic matrix of order $2l+k$ with entries

$$W(w_0aw_2|w_0w_1w_2) = \frac{\delta}{q^k-1}\mathbb{1}(a \in Q^k\backslash\{w_1\}) + \bar{\delta}\mathbb{1}(a = w_1),$$

where $\delta > 0$ and $\bar{\delta} := 1-\delta$. This matrix can be also viewed as a memoryless communication channel on Q^{2l+k}. As a result of this operation, we obtain a new system, which we denote by Z. This system is presented by a graph D whose set of vertices is V together with the new vertices of the form $(w_0aw_2), a \in Q^k\backslash\{w_1\}$. The distribution on the sequences of the system Z is obtained as $\nu([\![w]\!]) = \mathbb{E}_\mu(W([\![x]\!]|\cdot))$, where $w \in Q^{2l+k}$, which can be extended to the cylinder sets of $(Q^{2l+k})^{\mathbb{Z}}$ by independence. The construction of ν is explained in detail below.

We finally use ν to construct an (ϵ,k,l)-recoverable measure η over $Q^{\mathbb{Z}}$ similarly to the procedure that appears in the proof of Lemma IV.5.

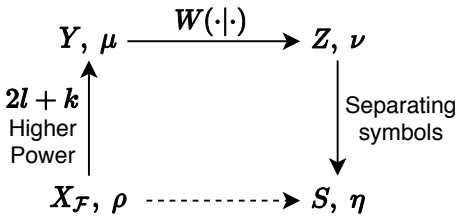


Figure 3: The procedure of constructing $\eta \in \mathcal{M}_s(Q^{\mathbb{Z}})$. From $X_{\mathcal{F}}$ (with the Markov measure ρ) we construct Y with a measure $\mu \in \mathcal{M}_s((Q^{2l+k})^{\mathbb{Z}})$ using the $(2l+k)$th higher power of $X_{\mathcal{F}}$. We then obtain Z, distributed according to ν, by passing the states of Y through a memoryless channel. In the last step we obtain the system S over the alphabet Q, together with the measure $\eta \in \mathcal{M}_s(Q^{\mathbb{Z}})$.

We proceed to implement this plan. The following sequence of steps transforms a graph that represents a (k,l)-recoverable system to a graph that supports ϵ-recoverability.

Construction IV.12 *1) Let $X_{\mathcal{F}}$ be a (k,l)-recoverable system with $\mathrm{cap}(X_{\mathcal{F}}) = C_q(k,l)$, presented by a graph G'. The vertices of G' correspond to $(2l+k-1)$-tuples of symbols in Q.*

2) Consider the set

$$\mathcal{F}' := \left\{w \in Q^{2l+k+1} : \exists u \in \mathcal{F} \text{ s.t. } u \prec w\right\}$$

and take $G = (V,E,L)$ to be the graph that presents $X_{\mathcal{F}'}$, so that its vertices correspond to words of length $2l+k$ over Q. Note that $X_{\mathcal{F}'} = X_{\mathcal{F}}$, so the graph G gives another presentation of the system $X_{\mathcal{F}}$.

3) Consider the graph $G^{2l+k} = (V(G^{2l+k}), E(G^{2l+k}), L(G^{2l+k}))$ with the same set of vertices as G. Two vertices in G^{2l+k} are connected by an edge if and only if there is a path of length $2l+k$ connecting them in G (including self-loops). The label of an edge $(v_1,v_2) \in E(G^{2l+k})$ is given by the label of the vertex v_2.

4) Construct a graph $D = (V(D), E(D), L(D))$. Given a vertex $u = (u_0u_1u_2)$, $u_0,u_2 \in Q^l$, $u_1 \in Q^k$ we write $u_a := (u_0au_2)$, where $a \in Q^k$ is some k-tuple. The set of vertices $V(D)$ is formed of $V(G^{2l+k})$ and all the vertices of the form $u_a = (u_0au_2), a \in Q^k\backslash\{u_1\}$ where $u = (u_0u_1u_2) \in V(G^{2l+k})$. The set $E(D)$ contains all the edges of the form (u_a, v_b), where $(u,v) \in E(G^{2l+k})$. □

The details of the construction are illustrated in Example A.1 in the Appendix, and the sequence of steps of the construction is shown schematically in Figure 3. To add details to Step 4, note that if a vertex $u \in V(G^{2l+k})$ corresponds to a triple $u = (u_0u_1u_2)$, $u_0,u_2 \in Q^l$, $u_1 \in Q^k$, then no other vertex in $V(G^{2l+k})$ is of the form (u_0au_2), where $a \neq u_1$.

In constructing $V(D)$ we add all the vertices of this form to $V(G^{2l+k})$, denoting the set of vertices that arise from $u \in V(G^{2l+k})$ by $(u)^*$; see Fig. 4.

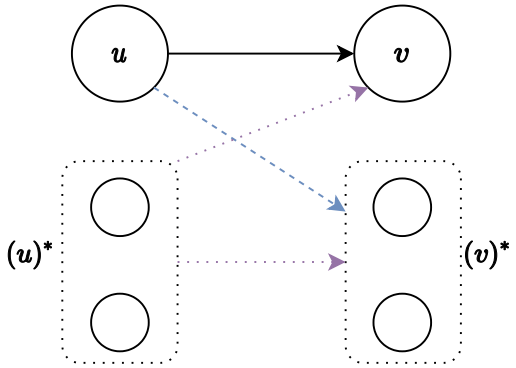


Figure 4: Construction of the graph D

The graph G^{2l+k} presents a constrained system over Q^{2l+k} which we denote by Y. It is evident that Y can be obtained from $X_{\mathcal{F}}$ by reading non-overlapping subwords of length $2l+k$ in the words $x \in X_{\mathcal{F}}$, and $X_{\mathcal{F}}$ can be obtained from Y by inverting this operation. It is a known fact that $\mathrm{cap}(Y) = \mathrm{cap}(X_{\mathcal{F}})$ (see [19, Prop. 3.13] or [18, Exercise 4.1.5], where this equality uses a different logarithm base and has a slightly different form). Accordingly, the system Z, presented by the graph D, is also over the alphabet Q^{2l+k}, and it will be used to construct an (ϵ, k, l)-recoverable system S over Q.

Let us define transition probabilities for the graph D. Theorem IV.3 implies that there exists a shift invariant Markov measure μ supported on Y and such that $h(\mu) = \mathrm{cap}(Y)$. Denote by P^μ the matrix of transition probabilities on $V(G^{2l+k})$ derived from μ. By shift-invariance, the marginal distribution of μ on any one coordinate equals the stationary distribution of P^μ. Denote this stationary distribution by p^μ.

Let $\epsilon > 0$ and let $\delta \in [0, (q-1)/q]$ be such that

$$H_q(\delta) + \delta \log_q(q^k - 1) = \epsilon. \tag{21}$$

Given a pair of vertices $(u, v) \in V(D)^2$, let

$$P_{uv} = \begin{cases} \bar{\delta} P^\mu_{uv} & \text{if } (u,v) \in E(G^{2l+k}) \\ \frac{\delta}{q^k-1} P^\mu_{u\bar{v}} & \text{if } u, \bar{v} \in V(G^{2l+k}); v \in (\bar{v})^* \\ \bar{\delta} P^\mu_{\bar{u}v} & \text{if } u \in (\bar{u})^*; \bar{u}, v \in V(G^{2l+k}) \\ \frac{\delta}{q^k-1} P^\mu_{\bar{u}\bar{v}} & \text{if } \bar{u}, \bar{v} \in V(G^{2l+k}); \\ & \quad u \in (\bar{u})^*, v \in (\bar{v})^* \end{cases} \tag{22}$$

(recall that $\bar{\delta} = 1 - \delta$). Referring to the example in the appendix (Fig. 5b), the solid edges correspond to line 1, the dashed edges to line 2, and the dotted edges to lines 3,4 in (22), respectively.

By construction, P is a stochastic matrix. Indeed, viewing the system Z presented by the graph D as the output of a memoryless channel we conclude that P is a stochastic matrix. Moreover, since P is irreducible, the stationary distribution exists and is unique. Together with the fact that for vertices $u \neq v$ in $V(G^{2l+k})$, the sets $(v)^* \cap (u)^*$ are disjoint, this implies the following result.

Lemma IV.13 *Define a vector p^ν of dimension $q|V(G^{2l+k})|$ by*

$$p^\nu_u = \begin{cases} \bar{\delta} p^\mu_u & \text{if } u \in V(G^{2l+k}) \\ \frac{\delta}{q^k-1} p^\mu_{\bar{v}} & \text{otherwise,} \end{cases} \tag{23}$$

where $\bar{v} \in V(G^{2l+k})$ is the unique vertex such that $u \in (\bar{v})^*$. Then p^ν is a probability vector and $p^\nu = p^\nu P$.

The stationary distribution p^ν together with P gives rise to a shift invariant measure ν on a set $Z \subseteq (Q^{2l+k})^{\mathbb{Z}}$ of bi-infinite sequences over Q^{2l+k}. For symbols $w_i \in Q^{2l+k}$, $i \in [n]$, the probability of the cylinder set $[\![w_0 \dots w_{n-1}]\!]$ is given by $p^\nu_{w_0} P_{w_0 w_1} \cdots P_{w_{n-2} w_{n-1}}$.

Our next goal is to construct an (ϵ, k, l)-recoverable measure η from ν. Let $Z = (\dots, Z_{-1}, Z_0, Z_1, \dots)$ be a system distributed according to ν. As the first step, let us transform sequences of $(2l+k)$-tuples $z \in (Q^{2l+k})^{\mathbb{Z}}$ into sequences $s \in Q^{\mathbb{Z}}$. Next, we construct a Markov approximation measure η to the measure ν using the procedure similar to the one used in the proof of Lemma IV.5, and such that $h(\eta) \geqslant h(\nu)$. Recall that ν is a 1-step Markov measure that corresponds to a system over the alphabet Q^{2l+k}. For small ϵ, there is an equality $h(\eta) = h(\nu)$ which is shown in Theorem IV.17.

Lemma IV.14 *The measure $\eta \in \mathcal{M}(Q^{\mathbb{Z}})$ obtained from ν is shift invariant.*

Proof: The fact that η is shift invariant follows from the shift invariance of ρ, and hence of μ. As explained in (10), η is obtained from the stationary distribution ν on Q^{2l+k} by extending ν to finite sequences. Since η is a $2l+k$ order Markov approximation of ν, we may take the state space of η to be $(2l+k)$-tuples. Since η has the same $(2l+k)$-tuples distribution as ν, both η and ν have the same stationary distribution. Any Markov process with initial stationary distribution is stationary, i.e., shift invariant. □

Let us show that η is an (ϵ, k, l)-recoverable measure. We will in fact show more, namely that the entropy condition is satisfied with equality.

Proposition IV.15 *Let η be a measure on $Q^{\mathbb{Z}}$ obtained from ν and denote by $S = (\dots, S_{-1}, S_0, S_1, \dots)$ the system that is distributed according to η. For every $\alpha, \beta \in Q^l$, we have*

$$H_\eta(S_{l+[k]} \mid S_{[l]} = \alpha, S_{l+k+[l]} = \beta) = \epsilon.$$

Proof: By the definition of η, for every $s = (s_0 \dots s_{2l+k-1}) \in Q^{2l+k}$, we have $\nu([\![s]\!]) = \eta([\![s]\!])$. Note that since ν is constructed from the shift invariant measure μ, if $s \in Q^{2l+k}$ with $\eta([\![s]\!]) > 0$, then $\nu([\![s]\!]) > 0$. Let $\alpha, \beta \in Q^l$ and define a probability vector $\boldsymbol{p} = (p_a)$ by setting

$$p_a = \frac{\nu([\![\alpha a \beta]\!])}{\sum_{b \in Q^k} \nu([\![\alpha b \beta]\!])}, \quad a \in Q^k.$$

Let us show that $H_q(\boldsymbol{p}) = \epsilon$. Let $c \in Q^k$ such that $\alpha c \beta \in V(G^{2l+k})$, then $\sum_{b \in Q^k} \nu([\![\alpha b \beta]\!]) = \mu([\![\alpha c \beta]\!])$ by Lemma IV.13. We obtain that $p_c = \bar{\delta}$, and for $a \neq c$, $p_a = \frac{\delta}{q^k - 1}$. From (21) we conclude that $H_q(\boldsymbol{p}) = \epsilon$, which was to be shown. □

Our next goal is to calculate the entropy $h(\nu)$, where we recall that ν is constructed using a (k, l)-recoverable measure

μ over $(Q^{2k+l})^{\mathbb{Z}}$. In the proof we use a standard expression for the entropy of a Markov chain Z with transition probabilities P and stationary distribution p^ν:

$$h(\nu) = -\sum_{u\in V} p_u^\nu \sum_{v\in V} P_{uv}\log P_{uv}$$

(see [26, Thm. 4.27]).

Proposition IV.16 *We have*

$$h(\eta) \geqslant h(\nu) = h(\mu) + \frac{1}{2l+k}\epsilon.$$

Proof: The inequality follows since η is the $(2l+k)$th Markov approximation of ν, so we only need to compute $h(\nu)$. Let (Z,ν) be the system obtained by Construction IV.12. We compute logarithms to the base q^{2l+k} and omit the base below in the proof. Using (23) and (22) and recalling that $V(D)\subset V(G^{2l+k})$, we split the sum on u,v into four parts as follows:

$$\begin{aligned}
-h(\nu) &= \sum_{u\in V(D)} p_u^\nu \sum_{v\in V(D)} P_{uv}\log P_{uv}\\
&= \sum_{u\in V(G^{2l+k})} p_u^\nu \sum_{v\in V(G^{2l+k})} P_{uv}\log P_{uv}\\
&\quad+\sum_{u\in V(G^{2l+k})} p_u^\nu \sum_{v\notin V(G^{2l+k})} P_{uv}\log P_{uv}\\
&\quad+\sum_{u\notin V(G^{2l+k})} p_u^\nu \sum_{v\in V(G^{2l+k})} P_{uv}\log P_{uv}\\
&\quad+\sum_{u\notin V(G^{2l+k})} p_u^\nu \sum_{v\notin V(G^{2l+k})} P_{uv}\log P_{uv}.
\end{aligned}$$

Evaluating the first sum, we obtain

$$\begin{aligned}
&\sum_{u\in V(G^{2l+k})} p_u^\nu \sum_{v\in V(G^{2l+k})} P_{uv}\log(P_{uv})\\
&= \sum_{u\in V(G^{2l+k})} \bar\delta p_u^\mu \sum_{v\in V(G^{2l+k})} \bar\delta P_{uv}^\mu \log(\bar\delta P_{uv}^\mu)\\
&= \bar\delta^2 \sum_{u\in V(G^{2l+k})} p_u^\mu \sum_{v\in V(G^{2l+k})} P_{uv}^\mu \log(\bar\delta P_{uv}^\mu)\\
&= -\bar\delta^2 h(\mu) + \bar\delta^2 \sum_{u\in V(G^{2l+k})} p_u^\mu \sum_{v\in V(G^{2l+k})} P_{uv}^\mu \log\bar\delta\\
&= -\bar\delta^2 h(\mu) + \bar\delta^2\log\bar\delta.
\end{aligned}$$

Similarly, the second sum evaluates to

$$\begin{aligned}
&\sum_{u\in V(G^{2l+k})} p_u^\nu \sum_{v\notin V(G^{2l+k})} P_{uv}\log(P_{uv})\\
&= \sum_{u\in V(G^{2l+k})} \bar\delta p_u^\mu \left(\sum_{v\in V(G^{2l+k})} \frac{(q^k-1)\delta}{q^k-1}\cdot\right.\\
&\qquad\left. P_{uv}^\mu \log\left(\frac{\delta}{q^k-1}P_{uv}^\mu\right)\right)\\
&= \sum_{u\in V(G^{2l+k})} \bar\delta p_u^\mu \sum_{v\in V(G^{2l+k})} \delta P_{uv}^\mu \log\left(\frac{\delta}{q^k-1}P_{uv}^\mu\right)\\
&= -\bar\delta\delta h(\mu) + \bar\delta\delta\log\frac{\delta}{q^k-1},
\end{aligned}$$

where the first equality follows from (22) and the fact that there are q^k-1 additional vertices in $V(D)$ for every vertex in $V(G^{2l+k})$.

The third and fourth sum are easily found to be

$$\begin{aligned}
&\sum_{u\notin V(G^{2l+k})} p_u^\nu \sum_{v\in V(G^{2l+k})} P_{uv}\log(P_{uv})\\
&= -\delta\bar\delta h(\mu) + \delta\bar\delta\log\bar\delta,
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{u\notin V(G^{2l+k})} p_u^\nu \sum_{v\notin V(G^{2l+k})} P_{uv}\log P_{uv}\\
&= -\delta^2 h(\mu) + \delta^2\log\frac{\delta}{q^k-1}.
\end{aligned}$$

Collecting the terms and using $\bar\delta = 1-\delta$, we obtain

$$\begin{aligned}
h(\nu) &= \left(\bar\delta^2 + 2\bar\delta\delta + \delta^2\right)h(\mu) - (\bar\delta^2+\delta\bar\delta)\log\bar\delta\\
&\quad - (\delta^2+\delta\bar\delta)\log\delta + (\delta+\bar\delta)\delta\log\left(q^k-1\right)\\
&= (\delta+\bar\delta)^2 h(\mu) - \bar\delta\log\bar\delta - \delta\log\delta\\
&\quad + \delta\log\left(q^k-1\right)\\
&= h(\mu) + H_{q^{2l+k}}(\delta) + \delta\log\left(q^k-1\right).
\end{aligned}$$

Using (21) we obtain

$$\begin{aligned}
h(\nu) &= h(\mu) + \frac{1}{2l+k}\left(H_q(\delta) + \delta\log_q\left(q^k-1\right)\right)\\
&= h(\mu) + \frac{\epsilon}{2l+k},
\end{aligned}$$

which finishes the proof. □

The next theorem forms one of the main results of this section and is an immediate consequence of Theorem IV.16 .

Theorem IV.17 *Let Q be a finite alphabet and let $\epsilon > 0$. Then*

$$h_q^*(\epsilon,k,l) \geqslant C_q(k,l) + \frac{1}{2l+k}\epsilon. \tag{24}$$

V. Concluding remarks

A. Relation to storage codes

In this section we discuss the connection between recoverable systems and storage coding. As noted above, storage codes in graphs were defined in several independent papers [20], [22], [24]. We will use the definition of [20], which is phrased in terms close to our work. Let $G=(V,E)$ be a finite graph and assume that $V=\{v_0,v_1,\ldots,v_{n-1}\}$. The neighborhood of a vertex $v\in V$ is an ordered collection of vertices $N(v)\subseteq V$ such that $(v,u)\in E$ if and only if $u\in N(v)$. A storage code $\mathcal{C}_G$ with recovery graph $G=(V,E)$ and $V=[n]$ is a subset of Q^n together with n deterministic recovery functions $f_v: Q^{N(v)}\to Q$ such that $c_v = f_v(c_u, u\in N(v))$ for every $c=(c_0,\ldots,c_{n-1})\in\mathcal{C}_G$ and every vertex $v\in V$. In words, a storage code is a set of vectors of length n such that the symbol in location i can be recovered from the symbols in locations specified by the neighbors of i in G.

The main problem studied in the literature is the capacity, or largest cardinality of a storage code for a given graph. Formally, the *capacity* of a storage code is defined as $M_q(G) = \frac{1}{n}\log_q|\mathcal{C}_G|$, and the absolute capacity is defined

as $M(G) = \sup_{q \geqslant 2} M_q(G)$. The definitions in earlier works usually do not include the normalization $1/n$, which we have added to better relate it to the notation adopted in this paper. The capacity is known for odd cycles C_n (and equals 1/2) [5], [21], and there are multiple bounds in the literature for other types of graphs [4], [5], [8], [21].

Before we describe the relation between storage codes and recoverable systems, let us make the following observation. Assume that $X_{\mathcal{F}}$ is a (k, l)-recoverable system. A sequence $x \in X_{\mathcal{F}}$ is said to have period r for some $r \in \mathbb{N}$ if $x_i = x_{i+r}$. Let $\mathcal{P}_n(X_{\mathcal{F}})$ denote the set of words in $X_{\mathcal{F}}$ with period n. It is known that the growth rate of the number of periodic sequences approaches the capacity of the system, namely:

$$\limsup_{n\to\infty} \frac{1}{n} \log_q |\mathcal{P}_n(X_{\mathcal{F}})| = \mathrm{cap}(X_{\mathcal{F}}).$$

[18, Thm. 4.3.6]. This implies a construction of storage codes for a family of graphs, which we illustrate for the cycle C_n.

Suppose that $X_{\mathcal{F}}$ is a $(1, 1)$-recoverable system, and note that a periodic sequence corresponds to the labels of a cycle in the graph G that presents it. The collection of n-words obtained from the cycles forms a storage code for C_n. We state this fact in the next obvious proposition.

Proposition V.1 *Let $X_{\mathcal{F}}$ be a $(1, 1)$-recoverable system. Then $\mathcal{P}_n(X_{\mathcal{F}})$ is a storage code for C_n.*

The set $\mathcal{P}_n(X_{\mathcal{F}})$ is shift invariant, i.e., if a word $w = (w_0, \ldots, w_{n-1}) \in \mathcal{P}_n(X_{\mathcal{F}})$ then also the cyclic shift $(w_1, \ldots, w_{n-1}, w_0) \in \mathcal{P}_n(X_{\mathcal{F}})$. This means that there is a single recovery function f such that every symbol w_i can be obtained by applying f with the arguments w_{i-1}, w_{i+1} (with indices mod n).

For the case $q = 2$ it is possible to obtain an explicit formula for the size $|\mathcal{P}_n(X_{\mathcal{F}})|$.

Proposition V.2 *Let $X_{\mathcal{F}}$ be a binary $(1, 1)$-recoverable system with maximum capacity constructed in Proposition II.3. Then*

$$|\mathcal{P}_n(X_{\mathcal{F}})| = \lambda_1^n + \lambda_2^n + \lambda_3^n$$

where $\lambda_1, \lambda_2, \lambda_3$ are the roots of $x^3 - x - 1$.

Proof: Let A be the adjacency matrix of the graph G given in Figure 6, i.e.,

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

The number of length-n cycles in G equals the trace of A^n, $|\mathcal{P}_n(X_{\mathcal{F}})| = \mathrm{Tr}(A^n)$. For convenience we denote $a_n = (A^n)_{1,1}$, $b_n = (A^n)_{2,2}$, $c_n = (A^n)_{3,3}$, thus $|\mathcal{P}_n(X_{\mathcal{F}})| = a_n + b_n + c_n$. Clearly, we have the following recursion: $a_n = a_{n-2} + a_{n-3}$ with initial values $a_0 = 1, a_1 = 0, a_2 = 1$. Further, from the form of the matrix A it is readily seen that $b_n = (A^{n-1})_{2,1}$, and we obtain a recursion $b_n = b_{n-2} + b_{n-3}$ with initial values $b_0 = 1, b_1 = 0, b_2 = 0$. Similarly, $c_n = c_{n-2} + c_{n-3}$ with initial values $c_0 = 1, c_1 = 0, c_2 = 1$.

Altogether, the sum $a_n + b_n + c_n$ satisfies a recursion $z_n = z_{n-2} + z_{n-3}$ with initial conditions $z_0 = 3, z_1 = 0, z_2 = 2$. The sequence $(z_n)_n$ is known as the Perrin sequence [1] in which the nth number is given by $z_n = \lambda_1^n + \lambda_2^n + \lambda_3^n$ where $\lambda_1, \lambda_2, \lambda_3$ are the roots of $x^3 - x - 1$ (sequence A001608 in OEIS; see http://oeis.org/A001608) . □

Notice that $\lambda_1 = \mathrm{cap}(X_{\mathcal{F}}) \approx 0.4057$ and $|\lambda_2|, |\lambda_3| < 1$, hence for large n we obtain $|\mathcal{P}_n(X_{\mathcal{F}})| \approx 0.4507^n$.

Let us stress again the differences between this result and the results in earlier works such as [5], [8], [21]: we analyze the size of the storage code for C_n for a given alphabet q, and we require the same recovery function for each vertex v, while the earlier works consider the supremum $\sup_q$ and allow different functions f_v.

The property of C_n that makes this construction possible is cyclic automorphism of the graph C_n. A more general family of graphs with a cyclic automorphism is *circulant graphs*, i.e., cycles with chords. A *circulant graph* is a graph $G = (V, E)$ with $V = (v_0, \ldots, v_{n-1})$ such that the graph TG obtained by relabeling the vertices $v_i \mapsto v_{i+1}$ (modulo n) is isomorphic to G. The bound in Proposition V.2 extends to this case without difficulty. Namely, if $X_{\mathcal{F}}$ is a (k, l)-recoverable system, where l is large enough to account for all the neighbors of the vertex, then $\mathcal{P}_n(X_{\mathcal{F}})$ provides a storage code for G.

B. Open problems

The concept of recoverable systems gives rise to a group of open questions. First, the family of circulant graphs is a subfamily of a larger class, namely, *transitive graphs* (when the automorphism group acts transitively on the set of vertices). Constructing recoverable systems that yield a bound on the capacity of storage codes for such graphs is an interesting open problem. Next, it is of interest to extend the construction of deterministic recoverable systems to the case $k, l > 1$, yielding bounds on the capacity for the more general case than the $(1, 1)$-recoverability considered above.

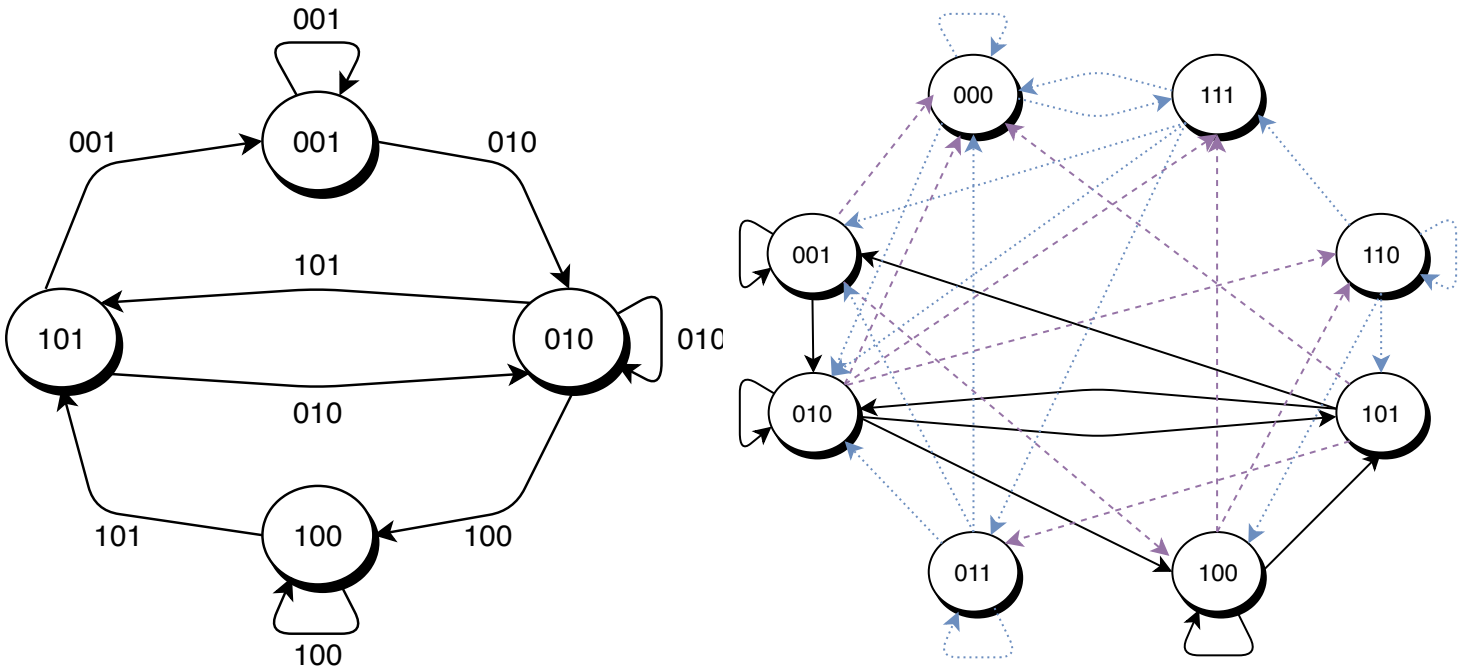
Several open questions arise for (ϵ, k, l)-recoverable systems. Under our definition, the residual entropy of the group of symbols is constrained by ϵ for *every realization* of the neighborhood; see (9). It is of interest to analyze the capacity of recoverable systems when this requirement holds only on average, i.e., when the entropy of the symbol group is conditioned on the random variables $X_{[l]}, X_{l+k+[l]}$ that form their neighborhood. Another open question is providing characterization of the entropy maximizing measures without restricting the values of ϵ.

Acknowledgments:

We are grateful to the reviewers of this submission for useful comments and corrections that have improved the presentation of the results. We also acknowledge a comment made by an anonymous reviewer of our conference paper [11] regarding a possible approach to the proof of Theorem IV.6, summarized in the remark in the main text.

APPENDIX

Example A.1 *(Refer to the discussion after Construction IV.12.)* We start with the binary $(1, 1)$-recoverable system given in Proposition II.3. Let Q be the binary alphabet, let $k = l = 1$ and let $\epsilon = 0.286$ which yields $\delta = 0.05, \bar{\delta} = 0.95$.



(a) (b)

Figure 5: An example for Steps 3 and 4 of Construction IV.12 for $q = 2$ and $k = l = 1$. Figure (a) shows the graph G^3 obtained in Step 3 of Construction IV.12. Figure (b) shows the graph D. The solid edges correspond to the edges in G^3, the dashed and dotted edges are added in Step 4 of Construction IV.12. The tails of the dashed edges are in $V(G^3)$ and the tails of the dotted edges are in $V(D) \setminus V(G^3)$.

The following example shows the construction of the graph D for the binary $(1, 1)$-recoverable system with maximum capacity given in Prop II.3. The graph that presents this system is shown in Fig. 6.

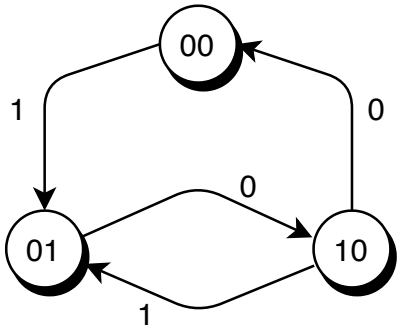


Figure 6: The graph G' presenting a binary $(1, 1)$-recoverable system with maximum capacity

Applying Steps $1, 2$ of Construction IV.12, we obtain the graph G^3 shown in Figure 5a. Now following Construction IV.2, we obtain $\mu = (0.177, 0.411, 0.177, 0.235)$ and

$$P^{\mu} = \begin{bmatrix} 0.43 & 0.57 & 0 & 0 \\ 0 & 0.43 & 0.245 & 0.325 \\ 0 & 0 & 0.43 & 0.57 \\ 0.43 & 0.57 & 0 & 0 \end{bmatrix},$$

where we have ordered the vertices of G^3 by the increase of their numerical value. The graph D is presented in Figure 5b. The matrix P is given in (25) at the top of the page, where the vertices are again ordered by the increase of their numerical value. The stationary distribution p^{ν} is given in (26) at the top of the page. For instance, we have

$$p_0 = \frac{\nu([001])}{\nu([001]) + \nu([011])} = \bar{\delta}$$

and

$$p_1 = \frac{\nu([001])}{\nu([001]) + \nu([011])} = \delta.$$

□

REFERENCES

[1] W. Adams and D. Shanks, "Strong primality tests that are not sufficient," *Mathematics of Computation*, vol. 39, no. 159, pp. 255–300, 1982.

[2] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, "Broadcasting with side information," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2008, pp. 823–832.

[3] F. Arbabjolfaei and Y.-H. Kim, "Three stories on a two-sided coin: Index coding, locally recoverable distributed storage, and guessing games on graphs," in *Proc. 53rd Ann. Allerton Conf. Commun. Contol Comput.*, Monticello, IL, 2015, pp. 843–850.

[4] A. Barg and G. Zémor, "High-rate storage codes on triangle-free graphs," preprint arXiv:2110.02378, October 2021.

[5] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Broadcasting with side information: Bounding and approximating the broadcast rate," *IEEE Trans. Inform. Theory*, vol. 59, no. 9, pp. 5811–5823, 2013.

[6] A. E. Brouwer and W. H. Haemers, *Spectra of Graphs*. New York e.a.: Springer, 2012.

[7] R. Burton and J. E. Steif, "Non-uniqueness of measures of maximal entropy for subshifts of finite type," *Ergodic Theory and Dynamical Systems*, vol. 14, no. 2, pp. 213–235, 1994.

$$P = \begin{bmatrix} 0.43\delta & 0 & 0.43\bar{\delta} & 0 & 0.245\bar{\delta} & 0.325\bar{\delta} & 0.245\delta & 0.325\delta \\ 0.57\delta & 0.43\bar{\delta} & 0.57\bar{\delta} & 0.43\delta & 0 & 0 & 0 & 0 \\ 0.43\delta & 0 & 0.43\bar{\delta} & 0 & 0.245\bar{\delta} & 0.325\bar{\delta} & 0.245\delta & 0.325\delta \\ 0.57\delta & 0.43\bar{\delta} & 0.57\bar{\delta} & 0.43\delta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.43\bar{\delta} & 0.57\bar{\delta} & 0.43\delta & 0.57\delta \\ 0.57\delta & 0.43\bar{\delta} & 0.57\bar{\delta} & 0.43\delta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.43\bar{\delta} & 0.57\bar{\delta} & 0.43\delta & 0.57\delta \\ 0.57\delta & 0.43\bar{\delta} & 0.57\bar{\delta} & 0.43\delta & 0 & 0 & 0 & 0 \end{bmatrix} \tag{25}$$

$$\begin{aligned} p^{\nu} &= (0.411\delta, 0.177\bar{\delta}, 0.411\bar{\delta}, 0.177\delta, 0.177\bar{\delta}, 0.235\bar{\delta}, 0.177\delta, 0.235\delta) \\ &\approx (0.02, 0.168, 0.391, 0.009, 0.168, 0.223, 0.009, 0.012) \end{aligned} \tag{26}$$

[8] P. J. Cameron, A. N. Dang, and S. Riis, "Guessing games on triangle-free graphs," *Electron. J. Combin.*, vol. 23, no. 1, pp. Paper 1.48, 15, 2016.

[9] N. G. de Bruijn, "A combinatorial problem," *Koninklijke Nederlandse Akademie v. Wetenschappen*, vol. 49, pp. 758–764, 1946.

[10] N. G. de Bruijn and P. Erdös, "Some linear and some quadratic recursion formulas II," *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen: Series A: Mathematical Sciences*, vol. 14, pp. 152–163, 1952.

[11] O. Elishco and A. Barg, "Capacity and construction of recoverable systems," in *Proc. 2021 IEEE Int. Sympos. Inf. Theory*, 2021.

[12] M. Fekete, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten," *Mathematische Zeitschrift*, vol. 17, no. 1, pp. 228–249, 1923.

[13] M. Gadouleau, "Finite dynamical systems, hat games, and coding theory," *SIAM J. Discrete Math.*, vol. 32, no. 3, pp. 1922–1945, 2018.

[14] H. O. Georgii, *Gibbs Measures and Phase Transitions*. W. de Gruyter, 1988.

[15] S. Goldstein, T. Kuna, J. L. Lebowitz, and E. R. Speer, "Translation invariant extensions of finite volume measures," *Journal of Statistical Physics*, vol. 166, no. 3-4, pp. 765–782, 2017.

[16] S. W. Golomb, *Shift Register Sequences*, 3rd ed. Holden-Day, San Francisco, 2017.

[17] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Shannon Foundation Publishers, 2004.

[18] D. Lind and B. H. Marcus, *An Introduction to Symbolic Dynamics and Coding*, 2nd ed. Cambridge University Press, 2021.

[19] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained systems and coding for recording channels," in *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, Eds. Elsevier, Amsterdam, 1998, vol. II, pp. 1635–1764.

[20] A. Mazumdar, "Storage capacity of repairable networks," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 5810–5821, 2015.

[21] A. Mazumdar, A. McGregor, and S. Vorotnikova, "Storage capacity as an information-theoretic vertex cover and the index coding rate," *IEEE Transactions on Information Theory*, vol. 65, no. 9, pp. 5580–5591, 2019.

[22] S. Riis, "Information flows, graphs, and their guessing numbers," *Electron. J. Comb.*, vol. 14, 2007, paper #R44.

[23] A. G. Schlijper, "On some variational approximations in two-dimensional classical lattice systems," *Journal of Statistical Physics*, vol. 40, no. 1-2, pp. 1–27, 1985.

[24] K. Shanmugam and A. Dimakis, "Bounding multiple unicasts through index coding and locally repairable codes," in *Proc.2014 IEEE Int. Sympos. Inf. Theory*, 2014, pp. 296–300.

[25] J. Sinai, "On the concept of entropy of a dynamical system," *Dokl. Akad. Nauk SSSR*, vol. 124, pp. 768–771, 1959, reprinted in *Ergodic and Information Theory*, Ed. by R.M. Gray and L.D. Davisson, Dowden, Hutchinson & Ross, Inc., 1977, pp. 243–246.

[26] P. Walters, *An Introduction to Ergodic Theory*. New York, NY: Springer-Verlag, 1982.