Dual Capacity Upper Bounds for Noisy Runlength Constrained Channels

Andrew Thangaraj

Abstract

Binary-input memoryless channels with a runlength constrained input are considered. Upper bounds to the capacity of such noisy runlength constrained channels are derived using the dual capacity method with Markov test distributions satisfying the Karush-Kuhn-Tucker (KKT) conditions for the capacity-achieving output distribution. Simplified algebraic characterizations of the bounds are presented for the binary erasure channel (BEC) and the binary symmetric channel (BSC). These upper bounds are very close to achievable rates, and improve upon previously known feedback-based bounds for a large range of channel parameters. For the binary-input Additive White Gaussian Noise (AWGN) channel, the upper bound is simplified to a small-scale numerical optimization problem. These results provide some of the simplest upper bounds for an open capacity problem that has theoretical and practical relevance.

I. Introduction

Runlength constrained input is commonly used in data storage applications, and channels with runlength constraints have been widely studied in information theory [1]. Characterizing the capacity of noisy channels with runlength-constrained input has been an open problem for quite some time now. Simulation-based methods for approximately computing lower and upper bounds to the noisy constrained capacity are well-known [2][3]. Numerical methods have been proposed for computing the capacity in some cases [4]. More recently, the feedback capacity, which serves as an upper bound to the non-feedback case, has been characterized as a computable optimization problem [5]. See references in [5] for a more complete bibliography of this area.

In this work, we derive upper bounds for noisy channels with runlength-constrained input. The main idea is the use of the dual capacity upper bound [6][7][8] and tuning it to the scenario of input-constrained noisy channels. The dual capacity bound has been used by several authors in applications such as optical channels [9], MIMO channels [10], phase noise channels [11] and peakpower-limited Gaussian channels [12]. In [3], the dual bound was used for runlength-constrained channels.

In the dual bound, a critical choice is that of the test distribution on the output alphabet. A Markov test distribution on the output of a runlength-constrained channel was used in [3]. The main innovation in this work is enforcing the Karush-Kuhn-Tucker (KKT) conditions for the capacity-achieving output distribution on the test distribution. This is done by equating suitably defined metrics of cycles in the state diagram of the constraint. As shown, this results in tight bounds and interesting and simple algebraic characterizations of the upper bound in several examples such as the runlength-constrained binary erasure channel and binary symmetric channel. For the binary-input Additive White Gaussian Noise (AWGN) channel, the bound can be computed by a small-scale optimization problem.

When compared to the feedback-based bound in [5], the dual capacity method in this paper is more direct and easily applicable to general (d, k) constraints and channels with continuous output. The resulting bounds are tighter and simpler algebraic characterizations in many cases. When compared to [3], the use of KKT conditions in test distributions is novel.

The rest of this paper is organized as follows. Section II introduces the notation and problem setup. Section III provides the main results. The proofs and computations of bounds are shown in Section IV, and concluding remarks are made in Section V.

II. Notation and Definitions

For integers a, b, the notation $[a : b]$ denotes the set of integers $\{a, a + 1, \ldots, b\}$. The set $[a : b]$ is empty if $a > b$. Given a sequence $v = (v_1, v_2, \ldots)$, and positive integers a, b the notation $v_{[a:b]}$ denotes the sub-sequence $(v_i : i \in [a : b])$. The sequence $v_{[1:N]}$ is also denoted by v^N. For a fraction x, we denote $\overline{x} = 1 - x$.

A directed graph $G = (V, E)$ consists of a set of vertices V and a set of directed edges $E \subseteq V \times V$. An edge $e = (v_1, v_2)$ is directed from v_1 to v_2. We will use the terms edges or arcs interchangeably for the directed edges of a directed graph. We will consider directed graphs that may have self-loops, i.e., edges of the form (v, v), but we will not consider graphs with multiple parallel edges between the same pair of vertices in the same direction. A walk of length k in a directed graph $G = (V, E)$ is an alternating sequence of vertices and edges $(v_1, e_1, v_2, e_2, v_3, \ldots, e_k, v_{k+1})$ such that $e_i = (v_i, v_{i+1}) \in E$. Often a walk is denoted simply by the sequence of vertices $(v_1, v_2, \ldots)$ or the sequences of edges $(e_1, e_2, \ldots)$. A walk with distinct vertices is called a path. A walk of length k with distinct v_1, v_2, $\ldots$, v_k and $v_1 = v_{k+1}$ is called a cycle of length k.

A. Runlength-constrained channels

For non-negative integers d and k with $k > d$, a (d, k)-constrained binary sequence is a sequence of bits for which there are at least d zeros and at most k zeros (k can be infinity) between any two 1s. A (d, k)-constrained sequence is usually represented as a walk on a state diagram. The state diagram for finite k and infinte k are shown in Fig. 1. As seen, the state
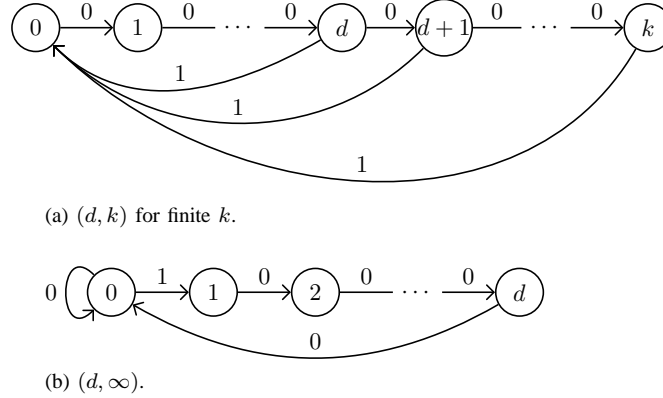

(a) (d, k) for finite k.

(b) (d, ∞).

Fig. 1. Directed graph $G_{d,k}$: State diagram for (d, k)-constrained sequences.

diagram for (d, k)-constrained sequences is a directed graph with edge labels, which we will denote $G_{d,k}$.

Let $\mathcal{X}_{d,k}^N$ denote the set of all (d, k)-constrained binary sequences of length N. The capacity of (d, k)-constrained sequences, denoted $C_{d,k}$, is defined and characterized as follows:

$$C_{d,k} \triangleq \lim_{N \to \infty} \frac{\log_2 |\mathcal{X}_{d,k}^N|}{N} = \log_2(1/\lambda), \tag{1}$$
$$\lambda \in (0, 1) \text{ solves } z^{k+2} - z^{d+1} - z + 1 = 0.$$

For $k = \infty$, the above characterization holds with the term z^{k+2} set as 0.

Consider a binary-input memoryless channel with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y}$ and transition probability denoted $p_{Y|X}(y|x)$. All channels considered in this paper will have binary input. The output alphabet $\mathcal{Y}$ may be either discrete or continuous. For discrete $\mathcal{Y}$, $p_{Y|X}$ will be denoted as an $|\mathcal{X}| \times |\mathcal{Y}|$ matrix with the (x, y)-th entry being $p_{Y|X}(y|x)$. For continuous $\mathcal{Y}$, we will specify the conditional probability density function (PDF) $p_{Y|X=x}$ for $x = 0, 1$. The capacity of a channel $p_{Y|X}$, denoted $C(p_{Y|X})$, is given by $C(p_{Y|X}) = \max_{p_X} I(X; Y)$. Standard convex optimization methods can be used for computing the capacity. Explicit algebraic expressions or small-scale numerical computations are available for computing the capacity of standard channels such as the binary erasure channel (BEC), binary symmetric channel (BSC) and the binary-input additive white Gaussian noise (BIAWGN) channel.

By a (d, k)-constrained channel $p_{Y|X}$, we refer to a channel $p_{Y|X}$ whose input is constrained to be a (d, k)-constrained binary sequence. Specifically, if the channel $p_{Y|X}$ is used N times, its input $X^N = [X_1, X_2, \ldots, X_N]$ is a length-N, (d, k)-constrained binary sequence, i.e. $X^N \in \mathcal{X}_{d,k}^N$. The output $Y^N = [Y_1, Y_2, \ldots, Y_N]$ obeys a memoryless channel transition law $p(y^N|x^N) = \prod_{i=1}^N p_{Y|X}(y_i|x_i)$. The capacity of the N-letter, (d, k)-constrained channel $p_{Y|X}$, denoted $C_{d,k}^N(p_{Y|X})$, is given by

$$C_{d,k}^N(p_{Y|X}) = \max_{p(x^N):x^N \in \mathcal{X}_{d,k}^N} I(X^N; Y^N). \tag{2}$$

The capacity of the (d, k)-constrained channel $p_{Y|X}$, denoted $C_{d,k}(p_{Y|X})$, is defined as

$$C_{d,k}(p_{Y|X}) = \lim_{N \to \infty} \frac{1}{N} C_{d,k}^N(p_{Y|X}). \tag{3}$$

Characterizing the (d, k)-constrained capacity of channels has proven to be considerably more difficult because the memory in the input makes the computation of (2) dependent on N, which grows to infinity. The main result of this paper is the derivation of upper bounds for $C_{d,k}(p_{Y|X})$ that either have simple algebraic characterizations (like that of $C_{d,k}$ in (1)) or small-scale numerical computation procedures. In particular, the computations are independent of N. The bounds, in many cases, are seen to be extremely close to achievable rates showing that they are tight.

In the rest of the paper, we will refer to the channel $p_{Y|X}$ with (d, k)-constrained input simply as the (d, k)-constrained $p_{Y|X}$. Standard channels considered are the following:

1) BEC(ϵ) with

$$p_{Y|X} = \begin{bmatrix} 1 - \epsilon & \epsilon & 0 \\ 0 & \epsilon & 1 - \epsilon \end{bmatrix}.$$

2) BSC(p) with

$$p_{Y|X} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}.$$

3) BIAWGN(σ^2) with

$$p_{Y|X=x} \sim N((-1)^x, \sigma^2),$$

where $N(m, s)$ denotes the Gaussian distribution with mean m and variance s.

We will use the notation (d, k)-BEC(ϵ), (d, k)-BSC(p) and (d, k)-BIAWGN(σ^2), respectively, for the (d, k)-constrained versions, and the notation $C_{d,k}(\epsilon)$, $C_{d,k}(p)$ and $C_{d,k}(\sigma)$ for the (d, k)-constrained capacities of the standard channels.

B. State diagram with memory μ for the (d, k)-constraint

Let μ be a positive integer satisfying

$$\mu \geq \begin{cases} k, & k : \text{finite}, \\ d, & k = \infty. \end{cases} \tag{4}$$

The state diagram for the (d, k) constraint with memory μ is an edge-labeled, directed graph, denoted $G_{d,k}^{\mu}$, and defined as follows. The vertex set of $G_{d,k}^{\mu}$ is the set $\mathcal{X}_{d,k}^{\mu}$ of (d, k)-constrained sequences of length μ. From a vertex $(x_1 x_2 \ldots x_\mu)$, an arc with label $x_{\mu+1} \in \{0, 1\}$ is drawn whenever $(x_1 x_2 \ldots x_\mu x_{\mu+1})$ is a valid length-$(\mu + 1)$, (d, k)-constrained sequence. The arc ends in the vertex $(x_2 \ldots x_{\mu+1})$. Examples of state diagrams with memory are shown in Fig. 2.
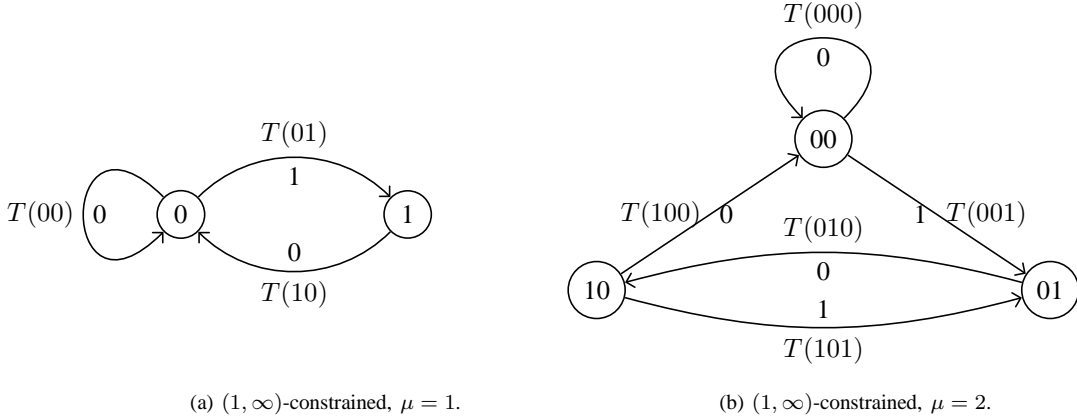

(a) $(1, \infty)$-constrained, $\mu = 1$.   (b) $(1, \infty)$-constrained, $\mu = 2$.

Fig. 2. $G_{1,\infty}^{\mu}$: Memory-μ state diagram for the $(1, \infty)$ constraint.

In $G_{d,k}^{\mu}$, the length-l cycle $\{v_1, v_2, \ldots, v_l, v_1\}$ is considered *equivalent* to the cycle $\{v_i, v_{i+1}, \ldots, v_l, v_1, \ldots, v_{i-1}, v_i\}$, which is simply the same cycle traversed with a different starting point. Let the set of cycles of $G_{d,k}^{\mu}$ be denoted $\mathcal{C}_{d,k}^{\mu}$ with the convention that no two cycles in it are equivalent. The length of a cycle $c \in \mathcal{C}_{d,k}^{\mu}$ is denoted $l(c)$.

Consider a length-l cycle $c = \{v_1, v_2, \ldots, v_l, v_1\} \in \mathcal{C}_{d,k}^{\mu}$ with $e_i = (v_i, v_{i+1})$, $1 \leq i \leq l$. Suppose $v_1 = (x_1 x_2 \ldots x_\mu) \in \mathcal{X}_{d,k}^{\mu}$ and let the label of edge e_i be denoted $x_{\mu+i}$. The sequence $(x_1 \ldots x_\mu x_{\mu+1} \ldots x_{\mu+l})$ is a valid (d, k)-sequence of length $\mu + l$, which we will denote $x(c)$ and associate with the cycle c. For example, in $G_{1,\infty}^2$ shown in Fig. 2, the length-3 cycle $c = \{00, 01, 10, 00\}$ is associated with the length-5 $(1, \infty)$-sequence $x(c) = (00100)$.

C. Markov test distributions

The upper bound on $C_{d,k}(p_{Y|X})$ is expressed using certain Markov test distributions on sequences of channel outputs. A sequence or chain of random variables $(Y_1, Y_2, \ldots)$ (with Y_i taking values in $\mathcal{Y}$) is said to be Markov with memory μ if Y_i is conditionally independent of $Y_{[1:i-\mu-1]}$ given $Y_{[i-\mu:i-1]}$. The distribution of a Markov chain is specified by the transition probability distributions $\Pr(Y_{\mu+1} = y_{\mu+1} | Y^\mu = y^\mu)$ and the initial distribution $\Pr(Y^\mu = y^\mu)$ for all $y^{\mu+1} \in \mathcal{Y}^{\mu+1}$. The distributions are specified as Probability Mass Functions (PMFs) or Probability Density Functions (PDFs) depending on whether $\mathcal{Y}$ is discrete or continuous.

Let $\mathcal{Q}_\mu$ refer to the collection of transition probabilities of Markov chains $(Y_1, Y_2, \ldots)$ with memory μ. A specific $q \in \mathcal{Q}_\mu$ is specified by providing

$$q(y_{\mu+1} | y^\mu) \triangleq \Pr(Y_{\mu+1} = y_{\mu+1} | Y^\mu = y^\mu)$$

for $y^{\mu+1} \in \mathcal{Y}^{\mu+1}$. For a Markov chain $(Y_1, Y_2, \ldots)$ with transition probability q and a channel output sequence $y^N \in \mathcal{Y}^N$, we have

$$\Pr(Y^N = y^N) = \prod_{i=1}^{\mu} q(y_i|y^{i-1}) \prod_{i=\mu+1}^{N} q(y_i|y_{[i-\mu:i-1]}), \tag{5}$$

where, for $1 \leq i \leq \mu$,

$$q(y_i|y^{i-1}) \triangleq \Pr(Y_i = y_i|Y^{i-1} = y^{i-1})$$

is specified by the initial distribution. We will refer to the set of transition probabilities q as Markov distributions of memory μ on the alphabet $\mathcal{Y}$.

D. Metric in the state diagram

Let $D\big(p_1(\cdot)\,\|\,p_2(\cdot)\big) = \sum_{y \in \mathcal{Y}} p_1(y) \log_2 \frac{p_1(y)}{p_2(y)}$ denote the relative entropy of two distributions p_1 and p_2 on the same alphabet $\mathcal{Y}$. Given a Markov distribution q of memory μ on the channel output alphabet $\mathcal{Y}$, we associate a metric to every edge in the graph $G_{d,k}^{\mu}$. The metric for the edge from vertex $(x_1 \ldots x_\mu)$ with label $x_{\mu+1}$, denoted $T_{q,p_{Y|X}}(x_1 \ldots x_{\mu+1})$, is defined as follows:

$$T_{q,p_{Y|X}}(x_1 \ldots x_{\mu+1}) = \sum_{y^\mu \in \mathcal{Y}^\mu} p_{Y^\mu|X^\mu}(y^\mu|x^\mu) D\big(p_{Y|X}(\cdot\,|x_{\mu+1})\,\|\,q(\cdot\,|y^\mu)\big). \tag{6}$$

In Fig. 2, the edge metrics are shown on the edges of the state diagram. The subscripts q and $p_{Y|X}$ are dropped in the notation for the edge metric when it is either clear from the context or not important.

1) Example - BEC(ϵ): To further illustrate, let us consider the channel BEC(ϵ), memory $\mu = 1$, and the Markov distribution

$$q(y_2|y_1) = \begin{bmatrix} \beta(1-\epsilon) & \epsilon & \overline{\beta}(1-\epsilon) \\ \alpha(1-\epsilon) & \epsilon & \overline{\alpha}(1-\epsilon) \\ 1-\epsilon & \epsilon & 0 \end{bmatrix}, \tag{7}$$

where the rows correspond to $y_1 = 0, ?, 1$, the columns correspond to $y_2 = 0, ?, 1$ in that order (? denotes erasure symbol), and α, β are parameters satisfying $0 \leq \alpha, \beta \leq 1$ and $\overline{\beta} = 1 - \beta$, $\overline{\alpha} = 1 - \alpha$. For this example, we see that

$$\begin{aligned}
T(00) &= \sum_{y_1 \in \{0,?,1\}} p_{Y|X}(y_1|0) D\big(p_{Y|X}(\cdot\,|0)\,\|\,q(\cdot\,|y_1)\big) \\
&= (1-\epsilon)\, D\big([1-\epsilon, \epsilon, 0]\,\|\,[\beta(1-\epsilon), \epsilon, \overline{\beta}(1-\epsilon)]\big) \\
&\quad + \ \epsilon \quad D\big([1-\epsilon, \epsilon, 0]\,\|\,[\alpha(1-\epsilon), \epsilon, \overline{\alpha}(1-\epsilon)]\big) \\
&= (1-\epsilon)^2 \log_2(1/\beta) + \epsilon(1-\epsilon) \log_2(1/\alpha).
\end{aligned} \tag{8}$$

Similar expressions derived for $T(01)$ and $T(10)$ are

$$T(01) = (1-\epsilon)^2 \log_2(1/\overline{\beta}) + \epsilon(1-\epsilon) \log_2(1/\overline{\alpha}), \tag{9}$$

$$T(10) = \epsilon(1-\epsilon) \log_2(1/\alpha). \tag{10}$$

2) Metric of walks and cycles: A valid (d,k) sequence $x^N \in \mathcal{X}_{d,k}^N$ corresponds to the walk in $G_{d,k}^{\mu}$ of length $N - \mu$ with sequence of vertices $(x_{[1:\mu]}, x_{[2:\mu+1]}, \ldots, x_{[N-\mu+1:N]})$. The metric of a walk in $G_{d,k}^{\mu}$ is defined to be the sum of the metrics on the edges in the walk. Since a cycle is a walk, the same definition extends for cycles as well. Consider a cycle c of length $l(c)$ in $G_{d,k}^{\mu}$ associated to the sequence $(x_1 \ldots x_{\mu+l(c)})$. The metric of the cycle, denoted $T_{q,p_{Y|X}}(c)$, is readily seen to be

$$T_{q,p_{Y|X}}(c) = \sum_{i=1}^{l(c)} T_{q,p_{Y|X}}(x_i \ldots x_{\mu+i}). \tag{11}$$

As an example, for the cycle in $G_{1,\infty}^{2}$ associated to the sequence (00100), the metric is $T(001) + T(010) + T(100)$.

III. UPPER BOUNDS

The main upper bound and several examples are presented and discussed in this section with the proofs and details of computations to be provided later. A dual capacity upper bound for the capacity of the (d,k)-constrained channel $p_{Y|X}$ is given in the following theorem and its corollary, which use the notation described in Section II.

Theorem 1 (Dual bound). *Let q be a Markov distribution with memory μ over the output alphabet $\mathcal{Y}$ satisfying the constraint that*

$$\frac{T_{q,p_{Y|X}}(c)}{l(c)} = \frac{T_{q,p_{Y|X}}(c')}{l(c')} \tag{12}$$

for any two cycles $c, c' \in \mathcal{C}_{d,k}^{\mu}$. Then, the capacity of the (d, k)-constrained channel $p_{Y|X}$ is upper bounded by the common value in (12), i.e.,

$$C_{d,k}(p_{Y|X}) \leq \frac{T_{q,p_{Y|X}}(c)}{l(c)}$$

for any $c \in \mathcal{C}_{d,k}^{\mu}$.

Since the upper bound holds for every q satisfying the constraint (12), the bound can be improved by minimizing over a set of Markov distributions satisfying the constraint. This observation results in the following corollary, which is useful in deriving bounds for standard channels.

Corollary 2 (Dual bound). *Let Q be a collection of Markov distributions with memory μ over the output alphabet $\mathcal{Y}$. Then,*

$$C_{d,k}(p_{Y|X}) \leq \min_{q \in Q} \ t(q, p_{Y|X}) \tag{13}$$

$$\text{subject to } t(q, p_{Y|X}) = \frac{T_{q,p_{Y|X}}(c)}{l(c)}, \quad \forall c \in \mathcal{C}_{d,k}^{\mu}.$$

The main observation about the dual upper bounds in Theorem 1 and Corollary 2 is that they are independent of N, the length of the channel input, occurring in the definition of $C_{d,k}(p_{Y|X})$. In the rest of this paper, we will refer to the above theorem and corollary as the dual bound theorem and the dual bound corollary, respectively. We will show by several examples that the dual bound theorem and corollary result in simple algebraic characterizations or small scale numerical computations for the upper bound.

A. Binary erasure channel

For the (d, k)-BEC(ϵ) with specific values of d and k, upper bounds on the capacity are given in the following theorems. Recall that the capacity is denoted $C_{d,k}(\epsilon)$.

Theorem 3. *(1) For the $(1, \infty)$-BEC(ϵ), the following upper bound holds:*

$$C_{1,\infty}(\epsilon) \leq (1 - \epsilon)^2 \log_2(1/\beta^*) + \epsilon(1 - \epsilon) \log_2(2 - \beta^*), \tag{14}$$

$$\beta^* \in (0, 1] \text{ solves } \beta^{2(1-\epsilon)} = 1 - \beta.$$

(2) For the $(1, \infty)$-BEC(ϵ), the following upper bound holds:

$$C_{(1,\infty)}(\epsilon) \leq (1 - \epsilon) \left[(1 - \epsilon)^2 \log_2 \frac{1}{\beta} + \epsilon(1 - \epsilon) \log_2 \frac{\alpha^2}{\gamma_1(2\alpha - 1)} + \epsilon^2 \log_2 \frac{1}{\alpha} \right], \tag{15}$$

where $\gamma_1 = 1 - \alpha - \beta + 2\alpha\beta$, and $\alpha, \beta \in [0, 1]$ solve

$$(1 - \epsilon)^2 \log_2 \frac{\beta^2(1 - \alpha)}{\gamma_2} + \epsilon(1 - \epsilon) \log_2 \frac{(2\alpha - 1)^2 \gamma_1}{\alpha^2 \gamma_2}$$

$$= (1 - \epsilon)^2 \log_2 \frac{\beta^2(1 - \alpha)}{(1 - \beta)^2(2\alpha - 1)} + 2\epsilon(1 - \epsilon) \log_2 \frac{\gamma_1}{\alpha(1 - \beta)}$$

$$= \epsilon^2 \log_2 \frac{1 - \alpha}{\alpha} \tag{16}$$

with $\gamma_2 = 2 - 3\alpha - \beta + 2\alpha\beta$.

The first part of the above theorem is obtained by using the dual bound corollary with $\mu = 1$ and the collection of test distributions in (7). Note that the expression is reminiscent of the noiseless capacity of $(1, \infty)$ sequences, which is given by

$$C_{1,\infty} = \log_2(1/\beta), \tag{17}$$

$$\beta = (\sqrt{5} - 1)/2 \text{ solves } \beta^2 = 1 - \beta.$$

The second part uses a collection of memory-2 Markov test distributions and involves more computations. The details are provided in later sections.

Fig. 3 shows plots of the dual bounds from Theorem 3. For comparison, a feedback-based upper bound from [5] and an achievable rate computed using the simulation method of [2] are shown. While all four lines are reasonably close in the figure, the zoomed inset shows that the dual capacity bounds are better with the $\mu = 2$ bound almost meeting the achievable rate.

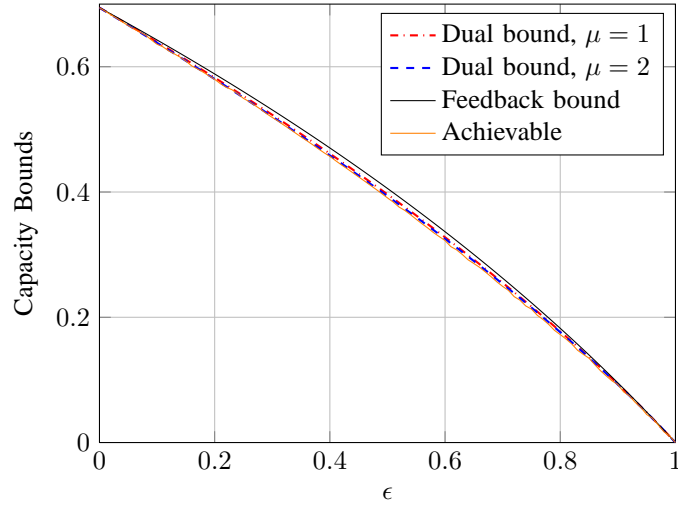Next, we provide dual bounds for the $(1, 2)$-BEC(ϵ) using $\mu = 2$ and $\mu = 3$ Markov test distributions.

Fig. 3. Capacity upper bounds for $(1, \infty)$-constrained BEC(ϵ).

Theorem 4. *(1) The capacity of the* $(1, 2)$*-BEC(ϵ) satisfies the upper bound*

$$C_{1,2}(\epsilon) \leq \frac{1 - \epsilon}{2} \left[(1 - \epsilon)^2 \log_2 \frac{1}{\beta} + \epsilon(1 - \epsilon) \log_2 \frac{(2 - \beta)^2}{\beta} + \epsilon^2 \log_2 \frac{(3 - \beta)^2}{2 - \beta} \right],$$

$$\beta \in (0, 1] \text{ solves } \beta^{3(1-\epsilon)}(2 - \beta)^{2\epsilon - 3\epsilon^2} = (1 - \beta)^{2 + 2\epsilon - 4\epsilon^2}.$$

(2) The capacity of the $(1, 2)$*-BEC(ϵ) satisfies the upper bound*

$$C_{1,2}(\epsilon) \leq \frac{1 - \epsilon}{2} \left[(1 + \epsilon - 2\epsilon^2) \log_2 \frac{1}{\beta} + 2\epsilon^3 \log_2(3 - \beta) + \epsilon^2(3 - 4\epsilon) \log_2(2 - \beta) \right],$$

$$\beta \in (0, 1] \text{ solves}$$

$$\beta^{3(1+\epsilon-2\epsilon^2)}(2 - \beta)^{\epsilon^2(3-4\epsilon)} = (1 - \beta)^{2(1+\epsilon+\epsilon^2-3\epsilon^3)} 2^{4\epsilon^2(1-\epsilon)}.$$

Fig. 4 shows plots of the dual bounds from Theorem 4 along with an achievable rate computed using the simulation method of [2].
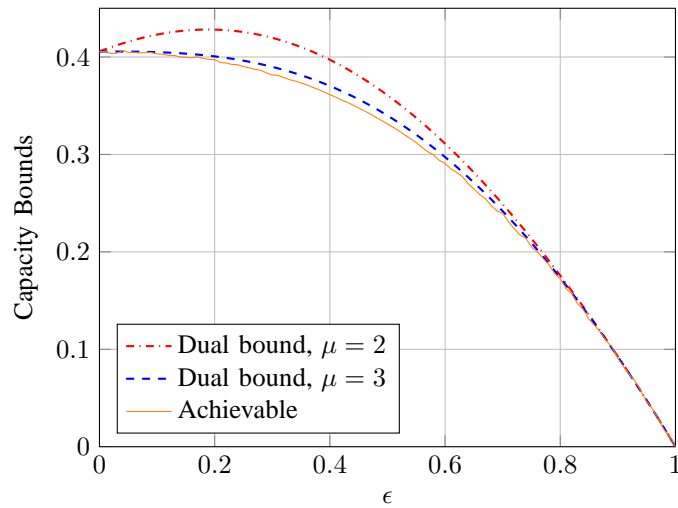

Fig. 4. Capacity upper bounds for $(1, 2)$-constrained BEC(ϵ).

We conclude the results for the BEC with an upper bound for the (d, ∞)-BEC(ϵ). Recall the notation $\overline{x} = 1 - x$ for a fraction x.

Theorem 5. *The capacity of the (d, ∞)-BEC(ϵ) satisfies the upper bound*

$$C_{(d,\infty)}(\epsilon) \leq (1-\epsilon) \sum_{i=0}^{d} B_\epsilon(d,i) \log_2 \frac{1 + i\overline{\alpha_0}}{\alpha_0 + i\overline{\alpha_0}}, \tag{18}$$

where $B_\epsilon(d,i) = \binom{d}{i}\epsilon^i(1-\epsilon)^{d-i}$ and $\alpha_0 \in [0,1]$ solves

$$\sum_{i=0}^{d} B_\epsilon(d,i) \log_2 \frac{(\alpha_0 + i\overline{\alpha_0})^{d-i+1}}{(1 + i\overline{\alpha_0})^{d-i}} = \log_2 \overline{\alpha_0}. \tag{19}$$

B. Binary Symmetric Channel

The next theorem presents an upper bound to the capacity of the $(1, \infty)$-BSC(p). We use the notation $H_2(p) = -p \log_2 p - (1-p) \log_2(1-p)$.

Theorem 6. *The capacity of the $(1, \infty)$-BSC(p) satisfies the upper bound*

$$C_{(1,\infty)}(p) \leq \overline{p}^2 \log_2 \frac{1}{a^*} + p\overline{p} \log_2 \frac{c_1}{\overline{p}^2} + p^2 \log_2 \frac{c_1}{pc_2} - H_2(p), \tag{20}$$

where $c_1 = \overline{a^*} - p^2$, $c_2 = 2\overline{p} - a^*(2-p)$, and $a^* \in [0,1]$ solves

$$a^{2\overline{p}}(\overline{a} - p^2) = p^{2p}\overline{p}^{2(1-2p)}\overline{a}^{2(1-2p)}[2\overline{p} - a(2-p)]^{2p}. \tag{21}$$

The dual bound is plotted along with a feedback-based bound (extension by authors of [5]) and an achievable rate using the simulation-based method of [2] for comparison in Fig. 5. We see that the dual capacity bound improves significantly over the
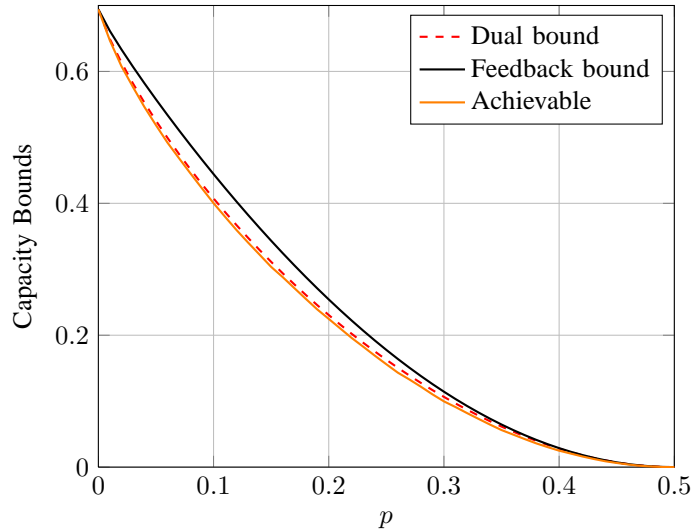

Fig. 5. Capacity bounds for $(1, \infty)$-constrained BSC(p).

feedback-based bound, and is close to the achievable rate as seen in the inset.

C. Binary-input AWGN Channel

Let the Gaussian PDF with mean μ and variance σ^2 be denoted

$$\psi_{\mu,\sigma}(x) \triangleq \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \; x \in (-\infty, \infty). \tag{22}$$

The Gaussian distribution restricted to the interval (a, b) has the PDF $\psi_{\mu,\sigma}(x)/\Psi_{\mu,\sigma}(a, b)$, where $\Psi_{\mu,\sigma}(a, b) \triangleq \int_a^b \psi_{\mu,\sigma}(x)dx$ is the probability that a Gaussian random variable with mean μ and variance σ^2 falls in the interval (a, b).

Consider the binary-input AWGN channel, denoted BIAWGN(σ^2), defined by the relationship $Y = (-1)^X + Z$, where the input $X \in \{0, 1\}$ and $Z \sim N(0, \sigma^2)$. The output alphabet $\mathcal{Y}$ is the set of real numbers. A memory-1 Markov distribution on

the output alphabet is specified by the conditional probability density function (PDF) $q(y_2|y_1)$, where y_1 and y_2 are real-valued variables.

After numerical experimentation, the following form of conditional PDFs was found to result in tight upper bounds:

$$q(y_2|y_1) = \begin{cases} a(y_1,\sigma)\dfrac{\psi_{-1,\sigma}(y_2)}{\Psi_{-1,\sigma}(-\infty, d_2(y_1,\sigma))}, & y_2 < d_2(y_1,\sigma), \\[2ex] b(y_1,\sigma)\dfrac{1}{\Delta(y_1,\sigma)}, & y_2 \in [d_2(y_1,\sigma), d_1(y_1,\sigma)], \\[2ex] c(y_1,\sigma)\dfrac{\psi_{+1,\sigma}(y_2)}{\Psi_{+1,\sigma}(d_1(y_1,\sigma), +\infty)}, & y_2 > d_1(y_1,\sigma), \end{cases} \tag{23}$$

where d_1, d_2, a, b and c are functions of y_1 and σ, and need to be chosen suitably for validity of the PDF and for minimizing the upper bound. We define $\Delta(y_1,\sigma) \triangleq d_1(y_1,\sigma) - d_2(y_2,\sigma)$. The conditional PDF is illustrated in Fig. 6. In the figure, the
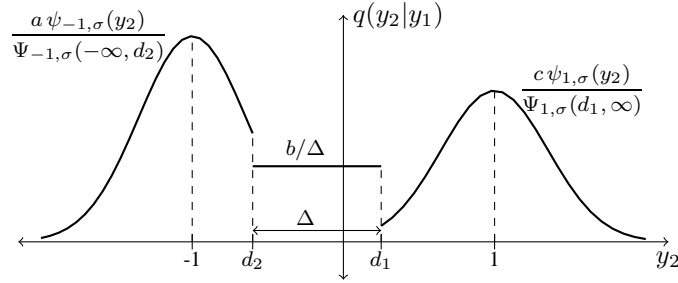

Fig. 6. Conditional PDF for a memory-1 Markov distribution.

dependence on y_1 and σ in the functions is suppressed to reduce clutter. The conditional PDF is restricted Gaussian with means +1 and -1 in the intervals (d_1, ∞) and $(-\infty, d_2)$, respectively, and uniform in the interval $[d_2, d_1]$. The values of a, b, and c are fractions summing to 1 for each value of y_1 and σ, and this constraint makes $q(y_2|y_1)$ a valid PDF for every y_1.

By numerical optimization of the functions d_1, Δ, a and b in the test distribution (note that $d_2 = d_1 - \Delta$ and $c = 1 - a - b$), an upper bound can be obtained for the $(1,\infty)$-constrained BIAWGN(σ^2). The results are shown in Fig. 7. In Fig. 7, a dual
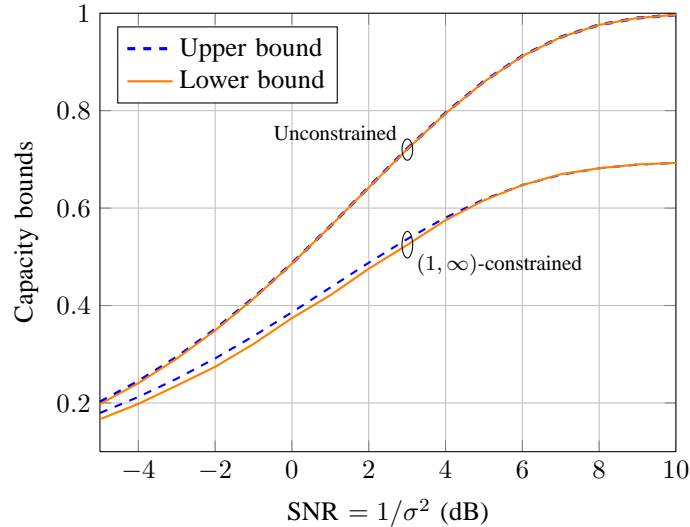

Fig. 7. Capacity bounds for the unconstrained and $(1,\infty)$-constrained BIAWGN(σ^2).

capacity upper bound for the unconstrained BIAWGN channel derived using a test distribution similar to the one shown in Fig. 7 is also shown. Details of the computations are provided in later sections. The upper bounds are seen to be quite tight when compared to lower bounds in both the constrained and the unconstrained case. The lower bound was computed using the simulation method of [2] for the constrained case. We remark that the choice of the test distribution is based on simple heuristics and could possibly be improved to obtain better bounds.

The above examples illustrate the kind of upper bounds that can be derived using the dual method. These are illustrative and not meant to be exhaustive. In general, for discrete channels, an algebraic characterization is possible, while for Gaussian channels, we obtain a small-scale numerical computation for constrained capacity. The computational complexity increases with the memory of the Markov distribution μ, and the optimization problems are nonlinear and may not be convex, in general. However, since their scale is small (number of variables and constraints depend on the memory μ), standard computation packages are effective in solving the optimizations or providing good local minima. The choice of test distribution can be guided by Lagrangian methods, but may also be made using heuristics for minimizing the relative entropy in the dual bound.

IV. PROOFS AND COMPUTATIONS

In this section, we provide proofs for the theorems in Section III and show the computations involved in simplifying the expressions for upper bounds. Some of the routine computations are given in the appendix.

We begin with the basic underlying idea, which is the use of the dual capacity upper bound.

A. Dual capacity bound

Consider the (d, k)-constrained channel $p_{Y|X}$. An upper bound on the N-letter capacity, called the dual capacity upper bound [6][7][8], is given by the following:

$$C_{d,k}^N(p_{Y|X}) \leq \max_{x^N \in \mathcal{X}_{d,k}^N} D\big(p_{Y^N|X^N}(\,\cdot\,|x^N)\,\|\,q_{Y^N}(\,\cdot\,)\big), \tag{24}$$

where q_{Y^N}, called the test distribution, is an arbitrary distribution on the output alphabet $\mathcal{Y}^N$. Note that in the above bound x^N is fixed in the expression $D\big(p_{Y^N|X^N}(\,\cdot\,|x^N)\,\|\,q_{Y^N}(\,\cdot\,)\big)$, and the relative entropy is evaluated between the probability mass functions (PMFs) $p_{Y^N|X^N}(y^N|x^N)$ and $q_{Y^N}(y^N)$ over the alphabet $\mathcal{Y}^N$. The maximum in (24) is over all valid (d, k) sequences x^N.

Since the bound in (24) holds for every q_{Y^N}, we minimize over a family of distributions $\mathcal{Q}$ to improve the upper bound as follows:

$$C_{d,k}^N(p_{Y|X}) \leq \min_{q_{Y^N} \in \mathcal{Q}} \max_{x^N \in \mathcal{X}_{d,k}^N} D\big(p_{Y^N|X^N}(\,\cdot\,|x^N)\,\|\,q_{Y^N}(\,\cdot\,)\big). \tag{25}$$

Given the nature of the input constraint, a good choice for the family of test distributions is $\mathcal{Q}_\mu$, which is the family of Markov distributions on $\mathcal{Y}$ with memory μ. Therefore, an upper bound to the capacity of the (d, k)-constrained $p_{Y|X}$ is

$$C_{d,k}(p_{Y|X}) \leq \min_{q_{Y^N} \in \mathcal{Q}_\mu} \limsup_{N \to \infty} \frac{1}{N} \max_{x^N \in \mathcal{X}_{d,k}^N} D\big(p_{Y^N|X^N}(\,\cdot\,|x^N)\,\|\,q_{Y^N}(\,\cdot\,)\big). \tag{26}$$

The above bound is not directly computable because of the dependence on N. We simplify the bound by application of the chain rule for relative entropy, and restrict the test distributions to obtain the dual bound theorem and corollary. The simplification is described below in detail for completeness.

B. Simplifying the dual capacity bound

For $q \in \mathcal{Q}_\mu$ (the subscript Y^N is suppressed in the notation), $x^N \in \mathcal{X}_{d,k}^N$ and a channel $p_{Y|X}$, the relative entropy term in the dual capacity bound can be simplified as follows:

$$D\big(p_{Y^N|X^N}(\,\cdot\,|x^N)\,\|\,q(\,\cdot\,)\big)$$

$$= \sum_{y^N \in \mathcal{Y}^N} p_{Y^N|X^N}(y^N|x^N) \log_2 \frac{\prod_{n=1}^N p_{Y|X}(y_n|x_n)}{\prod_{n=1}^\mu q(y_n|y^{n-1}) \prod_{n=\mu+1}^N q(y_n|y_{[n-\mu:n-1]})}$$

$$= \sum_{y^N \in \mathcal{Y}^N} \left(\prod_{i=1}^N p_{Y|X}(y_i|x_i)\right) \left[\sum_{n=1}^\mu \log_2 \frac{p_{Y|X}(y_n|x_n)}{q(y_n|y^{n-1})} + \sum_{n=\mu+1}^N \log_2 \frac{p_{Y|X}(y_n|x_n)}{q(y_n|y_{[n-\mu:n-1]})}\right]$$

$$\overset{(a)}{=} \sum_{n=1}^\mu \left[\sum_{y^{n-1} \in \mathcal{Y}^{n-1}} \left(\prod_{i=1}^{n-1} p_{Y|X}(y_i|x_i)\right) \sum_{y_n \in \mathcal{Y}} p_{Y|X}(y_n|x_n) \log_2 \frac{p_{Y|X}(y_n|x_n)}{q(y_n|y^{n-1})}\right] +$$

$$\sum_{n=\mu+1}^N \left[\sum_{y_{[n-\mu:n-1]} \in \mathcal{Y}^\mu} \left(\prod_{i=n-\mu}^{n-1} p_{Y|X}(y_i|x_i)\right) \sum_{y_n \in \mathcal{Y}} p_{Y|X}(y_n|x_n) \log_2 \frac{p_{Y|X}(y_n|x_n)}{q(y_n|y_{[n-\mu:n-1]})}\right],$$

$$\overset{(b)}{=} \sum_{n=1}^\mu \left[\sum_{y^n \in \mathcal{Y}^n} p_{Y^{n-1}|X^{n-1}}(y^{n-1}|x^{n-1}) D\big(p_{Y|X}(\,\cdot\,|x_n)\,\|\,q(\,\cdot\,|y^{n-1})\big)\right] +$$

$$\sum_{n=\mu+1}^{N} \left[\sum_{y^\mu \in \mathcal{Y}^\mu} p_{Y^\mu|X^\mu}(y^\mu|x_{[n-\mu:n-1]}) D\big(p_{Y|X}(\,\cdot\,|x_n) \,\|\, q(\,\cdot\,|y^\mu)\big) \right], \tag{27}$$

where (a) follows by interchanging the order of summation (over y^N and over n) and marginalizing, and (b) follows by identifying the expressions as relative entropies and changing the dummy summation variables from $y_{[n-\mu:n-1]}$ to y^μ.

The first summation term from $n=1$ to $n=\mu$ in (27) is $o(N)$ (assuming that the choice of q is such that the relative entropy is finite). Since the dual bound (see (26)) involves division by N, which tends to infinity, any $o(N)$ term is insignificant in the final bound on $C_{d,k}$. Moreover, the second summation term from $n=\mu+1$ to $n=N$ in (27) is readily identified as the sum of edge metrics (see (6)) of the walk in the graph $G_{d,k}^\mu$ corresponding to the (d,k) sequence x^N. So, we have

$$\frac{1}{N} D\big(p_{Y^N|X^N}(\,\cdot\,|x^N) \,\|\, q(\,\cdot\,)\big) = \frac{1}{N} \sum_{n=\mu+1}^{N} T_{q,p_{Y|X}}(x_{[n-\mu:n]}) + o(1), \tag{28}$$

where $o(1)$ tends to 0 as $N \to \infty$. Therefore, the edge metrics on the walk corresponding to x^N add up to form the asymptotically significant part of the upper bound on capacity.

C. Decomposing walks into cycles

A standard result in graph theory is that any walk in a directed graph can be decomposed into a path and a set of cycles (see [13], Exercises of Chapter 1). For completeness and future use, we provide a brief algorithmic proof of this result for the graph $G_{d,k}^\mu$.

Like mentioned before, a length-N, (d,k)-sequence $x^N \in \mathcal{X}_{d,k}^N$ corresponds to a walk $w(x^N)$ of length $N-\mu$ with sequence of vertices $v_i = x_{[i:i+\mu-1]}$ for $i=1,2,\ldots,N-\mu+1$. In the walk $w(x^N)$, let i_1 be the least integer for which v_{i_1} occurs more than once in $w(x^N)$, and let i_2 be the least integer such that $i_2 > i_1$ and $v_{i_1} = v_{i_2}$. Basically, $c_1(x^N) = (v_{i_1}, v_{i_1+1}, \ldots, v_{i_2})$ is the sequence of vertices of the first cycle traversed in $w(x^N)$, and its length is $l(c_1(x^N)) = i_2 - i_1$.

Now, remove the edges of $c_1(x^N)$ from the walk $w(x^N)$, and consider the walk $w_1 = (v_1, \ldots, v_{i_1}, v_{i_2+1}, \ldots)$. The first cycle in w_1 is denoted $c_2(x^N)$, and the same process is continued iteratively till a path $w^*(x^N)$ without any repeating vertices results after, say, $n_c(x^N)$ steps. The cycles resulting in this process $\{c_i(x^N)\}$ for $i=1,2,\ldots,n_c(x^N)$ and the final path $w^*(x^N)$ form the decomposition of the walk $w(x^N)$ into a set of cycles and a path in $G_{d,k}^\mu$.

We now use the above decomposition of the walk $w(x^N)$ in the dual capacity expression in (28). Since the metric of the path $w^*(x^N)$ is $o(N)$, we have that

$$\frac{1}{N} D\big(p_{Y^N|X^N}(\,\cdot\,|x^N) \,\|\, q(\,\cdot\,)\big) = \frac{1}{N} \sum_{i=1}^{n_c(x^N)} T_{q,p_{Y|X}}(c_i(x^N)) + o(1). \tag{29}$$

Since the length of the path $w^*(x^N)$ is $o(N)$, we have that

$$\frac{1}{N} \sum_{i=1}^{n_c(x^N)} l(c_i(x^N)) = 1 + o(1). \tag{30}$$

The above two equations resulting from the decomposition of walks into cycles along with a cycle metric restriction on the Markov distribution result in the proof of dual bound theorem. The restriction is described next.

D. KKT-constrained Markov test distributions

It is easy to show that the dual capacity upper bound results in equality if the test distribution is set to be equal to the capacity-achieving output distribution. Now, one of the classic results in computation of capacity of a channel $X \to Y$ with channel transition probability $p(y|x)$ is the set of Karush-Kuhn-Tucker (KKT) conditions, which require that

$$D\big(p(y|x) \,\|\, p(y)\big) = C, \text{ if } p(x) > 0,$$
$$D\big(p(y|x) \,\|\, p(y)\big) \le C, \text{ if } p(x) = 0,$$

for the capacity-achieving output distribution $p(y)$. Therefore, in the dual capacity bound calculation for the (d,k)-constrained channel $p_{Y|X}$, a simplifying condition that could result in good bounds is to require that the test distribution $q(y^N)$ satisfies

$$\frac{1}{N} D\big(p_{Y^N|X^N}(\,\cdot\,|x^N) \,\|\, q(\,\cdot\,)\big) = \text{constant} + o(1), \text{ if } x^N \in \mathcal{X}_{d,k}^N, \tag{31}$$

where the constant is independent of x^N, but possibly dependent on q and $p_{Y|X}$.

Consider the restricted set of Markov test distributions

$$\mathcal{Q}_\mu^* \triangleq \left\{ q \in \mathcal{Q}_\mu : \frac{T_{q,p_{Y|X}}(c)}{l(c)} = \frac{T_{q,p_{Y|X}}(c')}{l(c')} \forall c, c' \in C_{d,k}^\mu \right\}. \tag{32}$$

In words, for $q \in \mathcal{Q}_\mu^*$, the length-normalized metric of every cycle in $G_{d,k}^\mu$ is a constant. The common value of the length-normalized cycle metric is denoted $t(q, p_{Y|X})$, i.e.,

$$t(q, p_{Y|X}) \triangleq \frac{T_{q,p_{Y|X}}(c)}{l(c)}, q \in \mathcal{Q}_\mu^*, c \in C_{d,k}^\mu. \tag{33}$$

Using the cycle decomposition of walks and the results in (29), (30), we readily see that, for $q \in \mathcal{Q}_\mu^*$ and $x^N \in \mathcal{X}_{d,k}^N$,

$$\frac{1}{N} D\big(p_{Y^N|X^N}(\cdot|x^N) \,\|\, q(\cdot)\big) = \frac{1}{N} \sum_{i=1}^{n_c(x^N)} T_{q,p_{Y|X}}(c_i(x^N)) + o(1)$$

$$= t(q, p_{Y|X}) \left(\frac{1}{N} \sum_{i=1}^{n_c(x^N)} l(c_i(x^N))\right) + o(1)$$

$$= t(q, p_{Y|X}) + o(1). \tag{34}$$

We will refer to the set of distributions $\mathcal{Q}_\mu^*$ as the KKT-constrained Markov test distributions of memory μ. This is to be contrasted with the choice of test distributions made using the maximizing branch transition probability method in [3].

E. Proof of Theorem 1 and Corollary 2

The proof of the dual bound theorem and corollary is now immediate. Using (34), we see that

$$\limsup_{N\to\infty} \frac{1}{N} \max_{x^N \in \mathcal{X}_{d,k}^N} D\big(p_{Y^N|X^N}(\cdot|x^N) \,\|\, q_{Y^N}(\cdot)\big)$$

$$= t(q, p_{Y|X}). \tag{35}$$

Hence, the dual capacity bound for a KKT-constrained Markov test distribution results in

$$C_{d,k}(p_{Y|X}) \le t(q, p_{Y|X}), \quad q \in \mathcal{Q}_\mu^*, \tag{36}$$

which proves Theorem 1. Minimization over $q \in \mathcal{Q}_\mu^*$ proves Corollary 2.

F. Binary Erasure Channel

1) $(1, \infty)$-BEC(ϵ) (Proof of Theorem 3 - Part (1)): For this bound, the dual bound corollary is used with a memory-1 Markov test distribution given in (7), which is reproduced here for convenience.

$$q(y_2|y_1) = \begin{bmatrix} \beta(1-\epsilon) & \epsilon & \overline{\beta}(1-\epsilon) \\ \alpha(1-\epsilon) & \epsilon & \overline{\alpha}(1-\epsilon) \\ 1-\epsilon & \epsilon & 0 \end{bmatrix}, \tag{37}$$

where the rows correspond to $y_1 = 0, ?, 1$, the columns correspond to $y_2 = 0, ?, 1$ in that order (? denotes erasure symbol), and α, β are parameters satisfying $0 \le \alpha, \beta \le 1$. The specific choice in (37) is motivated by the minimization involved in the upper bound. If $y_{n-1} = 1$, we have $x_{n-1} = 1$ because the channel is a BEC. Now, if $x_{n-1} = 1$, by the $(1, \infty)$ constraint, we have that $x_n = 0$. So, $q(y_2|1)$ can be set to be $p_{Y|X}(y_2|0) = [1 - \epsilon \ \epsilon \ 0]$ to ensure that $D\big(p_{Y|X}(\cdot|x_n) \,\|\, q(\cdot|y_{n-1})\big) = 0$ whenever $y_{n-1} = 1$. The choice of $q(?|y_1) = \epsilon$ can be shown to be best possible for the BEC by using complementary slackness in the optimization problem of Corollary 2, and we skip the details.

With the choice of $q(y_2|y_1)$ as in (37), the metrics $T(00)$, $T(01)$ and $T(10)$ were computed earlier and are given in (8), (9) and (10). We will drop the subscripts from the notation $T_{q,p_{Y|X}}(\cdot)$ to reduce clutter. Note that the metrics depend on the parameters α, β and the channel erasure probability ϵ.

As seen in Fig. 2(a), there are two cycles in $G_{1,\infty}^1$ - a length-1 cycle associated to the sequence (00), and a length-2 cycle associated to the sequence (010). The length-normalized cycle metrics are $T(00)$ and $(T(01) + T(10))/2$, which have to be equal for a KKT-constrained Markov test distribution. So, we have the constraint $T(00) = (T(01) + T(10)/2$. By the dual bound corollary, we get the bound

$$C_{1,\infty}(\epsilon) \le \min_{\substack{q: \\ T(01)+T(10)=2T(00)}} T(00). \tag{38}$$

The above constrained optimization can be solved by standard Lagrangian techniques (as shown in the Appendix) resulting in the upper bound in Theorem 3 - Part (1).

In the optimization problem of (38) and those in the ensuing examples, while our methods might be resulting in global minima, we have not made an attempt to prove the same. The optimization problems could be non-convex in some cases, but Lagrangian methods result in good bounds even though there is a chance that they might be local minima for the problems.

TABLE I
CHOICE OF $q(y_3|y_1, y_2)$, $(rs) \in \{00, 0?, ?0, ??, 10, 1?\}$.

$(y_1 y_2)$	$q(y_3|y_1, y_2)$
01, ?1	$[1 - \epsilon \quad \epsilon \quad 0 \quad]$
rs	$[\alpha_{rs}(1 - \epsilon) \; \epsilon \; \overline{\alpha}_{rs}(1 - \epsilon)]$

2) $(1, \infty)$-BEC(ϵ) (Proof of Theorem 3 - Part (2)): For this bound, the dual bound corollary is used with a memory-2 Markov test distribution shown in Table I. The choices for $q(y_3|y_1, y_2)$ have been made to suit the constraint and the channel. If $y_2 = 1$, then we have $x_2 = 1$ since the channel is a BEC. This implies that $x_3 = 0$. So, we set $q(y_3|y_1, 1) = p_{Y|X}(y|0)$. For other values of $y_1 = r$, $y_2 = s$, $q(y_3|y_1, y_2)$ is parameterized by $\alpha_{rs} \in [0, 1]$.

As seen from Fig. 2(b), there are three directed cycles in the state diagram $G_{1,\infty}^2$ - (1) length-1 cycle associated to (000), (2) length-2 cycle associated to (0101), and (3) length-3 cycle associated to (00100). Equating the length-normalized cycle metrics, we get the constraints

$$T(010) + T(101) = 2T(000),$$
$$T(001) + T(010) + T(100) = 3T(000).$$

Note that the metrics $T(\cdot)$ are functions of the parameters α_{rs} and the erasure probability ϵ. By the dual bound corollary, we get the bound

$$C_{(1,\infty)}(\epsilon) \leq \min_{\substack{q: \\ T(010)+T(101)=2T(000) \\ T(001)+T(010)+T(100)=3T(000)}} T(000). \tag{39}$$

The above constrained optimization can be solved by standard Lagrangian techniques (as shown in the Appendix) resulting in the upper bound in Theorem 3 - Part (2).

3) $(1, 2)$-BEC(ϵ) (Proof of Theorem 4): The state diagrams for the $(1, 2)$ constraint with $\mu = 2, 3$ are shown in Fig. 8. The

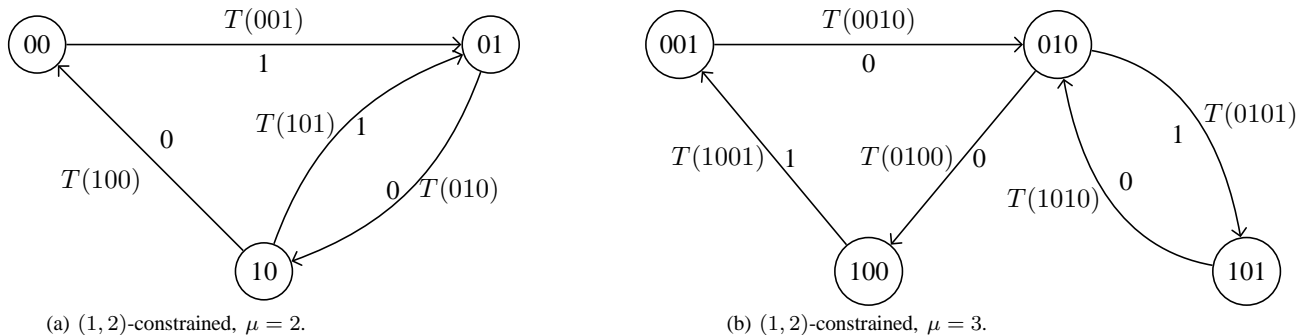(a) $(1, 2)$-constrained, $\mu = 2$.

(b) $(1, 2)$-constrained, $\mu = 3$.

Fig. 8. $G_{1,2}^\mu$: Memory-μ state diagram for the $(1, 2)$ constraint.

Markov test distribution used for $\mu = 2$ is given in Table II. The choices are made to suit the constraint. For $(y_1 y_2) \in \{01, ?1\}$,

TABLE II
TEST DISTRIBUTION FAMILY FOR $\mu = 2$. $(rs) \in \{10, 0?, ?0, 1?, ??\}$.

$(y_1 y_2)$	$q(y_3|y_1, y_2)$
00	$[0 \quad \epsilon \; 1 - \epsilon]$
01, ?1	$[1 - \epsilon \; \epsilon \quad 0 \quad]$
rs	$[\alpha_{rs}(1 - \epsilon) \; \epsilon \; \overline{\alpha}_{rs}(1 - \epsilon)]$

we have $x_3 = 0$. So, the test distribution $q(y_3|0, 1)$ and $q(y_3|?, 1)$ are both made equal to $p_{Y|X}(\cdot|0) = [1 - \epsilon \; \epsilon \; 0]$. Likewise, when $(y_1 y_2) = 00$, we have $x_3 = 1$ by the $(1, 2)$ constraint. So, we set $q(y_3|0, 0)$ as $p_{Y|X}(\cdot|1) = [0 \; \epsilon \; 1 - \epsilon]$. For other possibilities, we introduce parameters $\alpha_{rs} \in [0, 1]$.

In $G_{1,2}^2$, as seen in Fig. 8(a), there are two cycles - (1) the length-2 cycle associated to (0101), and (2) the length-3 cycle associated to (00100). Equating the length-normalized cycle metrics, we get the constraint

$$\frac{T(010) + T(101)}{2} = \frac{T(001) + T(010) + T(100)}{3}. \tag{40}$$

Note that the metrics $T(\cdot)$ are functions of the parameters α_{rs} and the erasure probability ϵ. By the dual bound corollary, we get the bound

$$C_{(1,2)}(\epsilon) \leq \min_{\substack{q: \\ T(010)+T(101)=2t \\ T(001)+T(010)+T(100)=3t}} t. \tag{41}$$

The above constrained optimization can be solved by standard Lagrangian techniques (as shown in the Appendix) resulting in the upper bound in Theorem 4 - Part (1).

The Markov test distribution used for $\mu = 3$ is given in Table III. The choices are made to suit the constraint. For $(y_1 y_2 y_3) \in$

TABLE III
TEST DISTRIBUTION FAMILY FOR $\mu = 3$. $(rst) \in \{10?, 0??, ?0?, ??0, 1??, ???\}$.

$(y_1 y_2 y_3)$	$q(y_4 \mid y_1, y_2, y_3)$
1?0, ?00, 100	$[0 \quad \epsilon \ 1-\epsilon]$
??1, 0?1, 00?, 001, 1?1, ?01, 101	$[1-\epsilon \ \epsilon \quad 0 \quad]$
?10, 01?, 0?0, ?1?, 010	$[\alpha_{010}(1-\epsilon) \ \epsilon \ \overline{\alpha}_{010}(1-\epsilon)]$
rst	$[\alpha_{rst}(1-\epsilon) \ \epsilon \ \overline{\alpha}_{rst}(1-\epsilon)]$

$\{1?0, ?00, 100\}$, we have $x_4 = 1$. So, the conditional test distribution for these cases is set as $p_{Y|X}(\cdot|1) = [0 \ \epsilon \ 1-\epsilon]$. Likewise, when $y_3 = 1$ or $(y_1 y_2 y_3) = 00?$, we have $x_4 = 0$ by the $(1,2)$ constraint. So, we set the conditional test distribution for these cases as $p_{Y|X}(\cdot|0) = [1 - \epsilon \ \epsilon \ 0]$. For $(y_1 y_2 y_3) \in \{?10, 01?, 0?0, ?1?, 010\}$, we see that $(x_1 x_2 x_3) = 010$. So, we set the conditional test distribution for these cases as $[\alpha_{010}(1-\epsilon) \ \epsilon \ \overline{\alpha}_{010}(1-\epsilon)]$ with a common parameter α_{010}. For every other possibility, a parameter $\alpha_{rst} \in [0,1]$ is used.

In $G_{1,2}^3$, as seen in Fig. 8(b), there are two cycles - (1) the length-2 cycle associated to (01010), and (2) the length-3 cycle associated to (001001). Equating the length-normalized cycle metrics, we get the constraint

$$\frac{T(0101) + T(1010)}{2} = \frac{T(0010) + T(0100) + T(1001)}{3}. \tag{42}$$

Note that the metrics $T(\cdot)$ are functions of the parameters α_{rst} and the erasure probability ϵ. By the dual bound corollary, we get the bound

$$C_{(1,2)}(\epsilon) \leq \min_{\substack{q: \\ T(0101)+T(1010)=2t \\ T(0010)+T(0100)+T(1001)=3t}} t. \tag{43}$$

The above constrained optimization can be solved by standard Lagrangian techniques (as shown in the Appendix) resulting in the upper bound in Theorem 4 - Part (2).

4) (d, ∞)-BEC(ϵ) (Proof of Theorem 5): For the (d, ∞) constraint with a test distribution of memory $\mu = d$, the graph $G_{d,\infty}^d$ has $d+1$ states and is isomorphic to the graph shown in Fig. 1(b) with the labels of states replaced with (d, ∞)-constrained sequences of length d. The state labels in $G_{d,\infty}^d$ are the $d+1$ sequences in $\mathcal{X}_{d,\infty}^d = \{u_0, u_1, \ldots, u_d\}$, where u_0 is the all-zero sequence of length d and u_i is the length-d binary sequence with a single 1 in the i-th position for $1 \leq i \leq d$.

For the bound in Theorem 5, the dual bound corollary is used with the memory-d Markov test distribution $q(y_{d+1}|y^d)$ defined as follows:

$$q(y_{d+1}|y^d) = \begin{cases} [1-\epsilon \ \epsilon \ 0] & \text{if } y^d \text{ has a 1,} \\ [\alpha_0(1-\epsilon) \ \epsilon \ \overline{\alpha_0}\,\overline{\epsilon}] & \text{if } y^d = 0^d, \\ [\alpha_k(1-\epsilon) \ \epsilon \ \overline{\alpha_k}\,\overline{\epsilon}] & \text{if } w_?(y^d) = k, \end{cases} \tag{44}$$

where $w_?(y^d)$ denotes the number of erasures in y^d, $\overline{x} \triangleq 1 - x$, and $\overline{\alpha_k} = \overline{\alpha_0}/(1 + k\overline{\alpha_0})$. Note that a valid y^d can have at most a single 1. Using Lagrangian techniques, the choice of a common parameter α_k for all y^d with $w_?(y^d) = k$ can be shown to be optimal for this case, and we skip the details.

There are two cycles in $G_{d,\infty}^d$ - (1) the length-1 cycle associated with $(u_0 0)$, and (2) the length-$(d+1)$ cycle associated with $(u_0 \, 1 \, u_0)$. Equating the length-normalized cycle metrics, we get the constraint

$$\frac{T(u_0 1) + \sum_{i=1}^d T(u_i 0)}{d+1} = T(u_0 0). \tag{45}$$

Note that the metrics are functions of the parameter α_0 and the erasure probability ϵ.

By the dual bound corollary, we get the bound

$$C_{(d,\infty)}(p_{Y|X}) \leq \min_{\substack{q: \\ T(u_0 1)+\sum_{i=1}^d T(u_i 0)=(d+1)T(u_0 0)}} T(u_0 0). \tag{46}$$

The above constrained optimization can be solved by standard Lagrangian techniques (as shown in the Appendix) resulting in the upper bound in Theorem 5.

G. Binary Symmetric Channel - Proof of Theorem 6

We will use the memory-1 Markov test distribution $q(y_2|y_1)$ defined as

$$q(y_2|y_1) = \begin{bmatrix} a & 1-a \\ b & 1-b \end{bmatrix}, \tag{47}$$

where the rows correspond to $y_1 = 0, 1$, columns correspond to $y_2 = 0, 1$ in that order, and a, b are parameters satisfying $0 \leq a, b \leq 1$.

Using the dual bound corollary, we get the bound

$$C_{1,\infty}(p) \leq \min_{a,b:T(01)+T(10)=2T(00)} T(00). \tag{48}$$

For the choice of test distribution over the binary symmetric channel BSC(p), the metrics evaluate to the following:

$$T(00) = \overline{p}^2 \log_2 \frac{\overline{p}}{a} + p\overline{p} \log_2 \frac{p\overline{p}}{b\overline{a}} + p^2 \log_2 \frac{p}{b},$$

$$T(01) = \overline{p}^2 \log_2 \frac{\overline{p}}{\overline{a}} + p\overline{p} \log_2 \frac{p\overline{p}}{a\overline{b}} + p^2 \log_2 \frac{p}{b},$$

$$T(10) = \overline{p}^2 \log_2 \frac{\overline{p}}{b} + p\overline{p} \log_2 \frac{p\overline{p}}{a\overline{b}} + p^2 \log_2 \frac{p}{a},$$

where $\overline{x} \triangleq 1 - x$. Further, the constraint simplifies as follows:

$$T(10) + T(01) - 2T(00) = (1 - 2p) \left[(1 - 2p) \log_2 \frac{a(1-b)}{b(1-a)} - \log_2 \frac{1-b}{a} \right] = 0. \tag{49}$$

Therefore, the optimization in (48) simplifies to a problem in two variables a, b taking values in $[0, 1]$. This problem is solved using standard Lagrangian methods to obtain the bound in Theorem 6.

H. $(1,\infty)$-constrained binary-input AWGN Channel

For $x \in \{0,1\}$, let $s(x) \triangleq (-1)^x$ denote the standard BPSK modulation. Recall the notation for the Gaussian PDF $\psi_{\mu,\sigma} = e^{-(x-\mu)^2/(2\sigma^2)}/(\sqrt{2\pi}\sigma)$ and its integral $\Psi_{\mu,\sigma}(a, b) = \int_a^b \psi_{\mu,\sigma}(y)dy$ introduced earlier.

As can be expected, the computations for the binary-input Gaussian channel, defined by the conditional PDF $p_{Y|X}(y|x) = \psi_{s(x),\sigma}(y)$, are highly numerical in nature. We use the dual bound corollary with the class of memory-1 Markov test distribution defined by the conditional PDF $q(y_2|y_1)$ given in (23) and depicted in Fig. 6.

The relative entropy $D\big(p_{Y|X}(\cdot|x) \,\|\, q(\cdot|y_1)\big)$, denoted as $D_x(y_1)$, simplifies to the following:

$$D_x(y_1)/\log_2 e = \Psi_{s(x),\sigma}(-\infty, d_2(y_1,\sigma)) \log_e \frac{e^{\frac{1+s(x)}{\sigma^2}} \Psi_{-1,\sigma}(-\infty, d_2(y_1,\sigma))}{a(y_1,\sigma)} +$$

$$\Psi_{s(x),\sigma}(d_2(y_1,\sigma), d_1(y_1,\sigma)) \log_e \frac{\Delta(y_1,\sigma)}{\sqrt{2\pi}\sigma e^{\frac{s(x)}{2}} b(y_1,\sigma)} +$$

$$\Psi_{s(x),\sigma}(d_1(y_1,\sigma), \infty) \log_e \frac{e^{\frac{1-s(x)}{\sigma^2}} \Psi_{1,\sigma}(d_1(y_1,\sigma), \infty)}{c(y_1,\sigma)} +$$

$$\left(\frac{d_1(y_1,\sigma) - 1}{2} - i(x) \right) \psi_{s(x),\sigma}(d_1(y_1,\sigma)) - \left(\frac{d_2(y_1,\sigma) + 1}{2} + \overline{i(x)} \right) \psi_{s(x),\sigma}(d_2(y_1,\sigma)),$$

where $i(x)$ is x seen as an integer. In the graph $G_{1,\infty}^1$, the metrics $T(x_1 x_2)$, for $(x_1 x_2) \in \{00, 01, 10\}$, are given by the following:

$$T(x_1 x_2) = \int_{-\infty}^{\infty} D_{x_2}(y_1) p_{Y|X}(y_1|x_1) dy_1$$

$$= \int_{-\infty}^{\infty} D_{x_2}(y_1) \psi_{s(x_1),\sigma}(y_1) dy_1. \tag{50}$$

The metrics above depend on the choice of the functions d_1, Δ, a and b, which are all functions of y_1 and σ. Note that $d_2 = d_1 - \Delta$ and $c = 1 - a - b$ are dependent on the other functions. Also, the functions a, b and c take nonnegative fractional values adding to 1, and the function Δ takes nonnegative values.

The choice of the conditional PDF, its shape, its piecewise nature, and its dependence on y_1 and σ are motivated by the $(1,\infty)$ constraint and the minimization of the relative entropy in the dual upper bound. There are three piecewise shapes for $q(y_2|y_1)$ - (1) shape of $N(-1, \sigma^2)$ for $y_2 < d_2(y_1, \sigma)$, (2) constant for $d_2(y_1, \sigma) \leq y_2 \leq d_1(y_1, \sigma)$, and (3) shape of $N(1, \sigma^2)$ for $y_2 > d_1(y_1, \sigma)$. These shapes are weighted by the fractions $a(y_1, \sigma)$, $b(y_1, \sigma)$ and $c(y_1, \sigma)$.

If y_1 is large and negative, then $x_1 = 1$ with high probability and this implies $x_2 = 0$ with high probability because of the $(1, \infty)$ constraint. Therefore, for large negative values of y_1, $q(y_2|y_1)$ can be of the shape of $N(1, \sigma)$ for significant values of y_2 by choosing the value of the function $d(y_1, \sigma)$ and the fractions $a(y_1, \sigma)$, $b(y_1, \sigma)$ and $c(y_1, \sigma)$ appropriately. As the value of y_1 increases and becomes large and positive, the probability of $x_1 = 0$ increases to 1. So, x_2 can be 0 or 1 with some nonzero probability as per the $(1, \infty)$ constraint. So, as y_1 increases, the function $d(y_1, \sigma)$ needs to increase to positive values. There are similar heuristics used to motivate the other aspects of $q(y_2|y_1)$.

Consider the class of functions

$$\mathcal{F}_{\alpha_{[1:4]}}(y_1, \sigma) = \frac{\alpha_1 e^{y_1/\sigma^2} + \alpha_2 e^{-y_1/\sigma^2}}{\alpha_3 e^{y_1/\sigma^2} + \alpha_4 e^{-y_1/\sigma^2}}, \tag{51}$$

where α_i, $1 \leq i \leq 4$, are real-valued parameters. After some experimentation, we have found that the functions d_1, Δ, a and b can be chosen from the above class of functions with some suitable restrictions on the values of parameters $\alpha_{[1:4]}$. The $\tanh$-like choice is motivated by the form of the posterior probabilities over the BIAWGN channel, which have the $\tanh$ form. In our computations, we use the following set $\mathcal{F}$ for the choice of the functions d_1, Δ, a and b:

$$\mathcal{F} = \{d_1, \Delta, a, b :$$
$$d_1 \in \mathcal{F}_{\delta_{[1:4]}}(y_1, \sigma), \ \delta_2 \in [-1, 1], \delta_1, \delta_2, \delta_3 \in [0, 1],$$
$$\Delta \in \mathcal{F}_{\Delta_{[1:4]}}(y_1, \sigma), \ \Delta_i \in [0, 1],$$
$$a \in \mathcal{F}_{\alpha_{[1:4]}}(y_1, \sigma), \ \alpha_i \in [0, 1],$$
$$b \in \mathcal{F}_{\beta_{[1:4]}}(y_1, \sigma), \ \beta_i \in [0, 1],$$
$$\max\left(\frac{\alpha_1}{\alpha_3}, \frac{\alpha_2}{\alpha_4}\right) + \max\left(\frac{\beta_1}{\beta_3}, \frac{\beta_2}{\beta_4}\right) \leq 1\}.$$

The above choices allow d_1 to be negative and ensures that Δ is positive and that a, b and $c = 1 - a - b$ are nonnegative fractions adding to 1.

Using the dual bound corollary, we obtain the bound

$$C_{(1,\infty)}(\sigma) \leq \min_{\substack{d_1, \Delta, a, b \in \mathcal{F}: \\ T(01) + T(10) = 2T(00)}} T(00). \tag{52}$$

The above problem is non-linear, and local minima can be found using numerical optimization procedures. This bound is plotted in Fig. 7 as the $(1, \infty)$-constrained upper bound. The lower bound is by using the method of [2].

For the unconstrained BIAWGN(σ^2) channel, the dual upper bound is evaluated using the following test distribution

$$q(y) = \begin{cases} \dfrac{1 - a(\sigma)}{2\Psi_{-1,\sigma}(-\infty, -\Delta(\sigma))} \psi_{-1,\sigma}(y), & y < -\Delta(\sigma), \\[2ex] \dfrac{a(\sigma)}{2\Delta(\sigma)}, & |y| \leq \Delta(\sigma), \\[2ex] \dfrac{1 - a(\sigma)}{2\Psi_{1,\sigma}(\Delta(\sigma), \infty)} \psi_{1,\sigma}(y), & y > \Delta(\sigma). \end{cases} \tag{53}$$

parameterized by a positive real-valued function $\Delta(\sigma)$ and $a(\sigma)$ is a function taking values in $[0, 1]$ and will be chosen to get the best bound. The relative entropy $D\big(p_{Y|X}(\cdot|x) \| q(\cdot)\big)$ for both $x = 0, 1$ simplifies to the following expression:

$$D\big(p_{Y|X}(\cdot|x) \| q(\cdot)\big) / \log_2 e = \big(1 - \Psi_{1,\sigma}(-\Delta, \Delta)\big) \log_e \frac{2\Psi_{1,\sigma}(\Delta, \infty)}{1 - a} + \Psi_{1,\sigma}\big(-\Delta, \Delta\big) \log_e \frac{2\Delta}{a\sqrt{2\pi}\sigma} + \tag{54}$$
$$\frac{2}{\sigma^2}\Psi_{1,\sigma}(-\infty, -\Delta) + 2\psi_{1,\sigma}(-\Delta) +$$
$$\frac{1}{2}\left[(\Delta - 1)\psi_{1,\sigma}(\Delta) + (\Delta + 1)\psi_{1,\sigma}(-\Delta) - \Psi_{1,\sigma}(-\Delta, \Delta)\right],$$

where the dependence of Δ and a on σ is suppressed in the notation to reduce clutter. From the above, it is seen that the choice

$$a(\sigma) = \Psi_{1,\sigma}(-\Delta(\sigma), \Delta(\sigma)) \tag{55}$$

minimizes the relative entropy. The upper bound on the capacity of the unconstrained BIAWGN channel, shown in Fig. 7, is obtained by setting $a(\sigma)$ as in (55) and numerically finding the best $\Delta(\sigma)$ that minimizes (54).

V. Concluding remarks

The dual capacity bound is useful in scenarios where characterizing the exact capacity is difficult. In particular, restricting the test distributions to those that satisfy the Karush-Kuhn-Tucker (KKT) conditions on the capacity-achieving output distribution appears to be an important idea for obtaining tight bounds with simple characterizations. In this paper, KKT-constrained test distributions were explored for runlength constrained binary channels.

For runlength constrained channels, the KKT constraint is converted into a condition on the metrics of cycles in the state diagram. For larger memory of the test distribution, the state diagram has many cycles and computation complexity of the method increases, but, interestingly, the bounds for low memory appear to be tight in many cases of theoretical and practical interest. Characterizing the gap between the upper bound and achievable rates analytically is an interesting problem to pursue in the future. Extending the method to other channels with memory, such as the Inter Symbol Interference (ISI) channel, is another interesting avenue for future work.

References

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

[2] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, Aug 2006.

[3] P. Vontobel and D. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Information Theory Workshop, 2001. Proceedings. 2001 IEEE*, 2001, pp. 147–149.

[4] Y. Li and G. Han, "Input-constrained erasure channels: Mutual information and capacity," in *Information Theory (ISIT), 2014 IEEE International Symposium on*, June 2014, pp. 3072–3076.

[5] O. Sabag, H. H. Permuter, and N. Kashyap, "The feedback capacity of the binary erasure channel with a no-consecutive-ones input constraint," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 8–22, Jan 2016.

[6] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Sci. Math. Hungar.*, vol. 2, pp. 291–292, 1967.

[7] J. Kemperman, "On the Shannon capacity of an arbitrary channel," *Indagationes Mathematicae (Proceedings)*, vol. 77, no. 2, pp. 101 – 115, 1974.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[9] A. Lapidoth, S. Moser, and M. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct 2009.

[10] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, Oct 2003.

[11] G. Durisi, "On the capacity of the block-memoryless phase-noise channel," *IEEE Communications Letters*, vol. 16, no. 8, pp. 1157–1160, August 2012.

[12] A. Thangaraj, G. Kramer, and G. Bcherer, "Capacity upper bounds for discrete-time amplitude-constrained awgn channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2321–2325.

[13] J. Bang-Jensen and G. Gutin, *Digraphs: Theory, Algorithms and Applications*. Springer London, 2013.

Appendix: Langrangian computations

A. $(1, \infty)$-BEC(ϵ): Theorem 3 - Part (1)

For the optimization problem in (38), the Lagrangian is defined as follows:

$$
\begin{aligned}
L &= T(00) + \lambda(T(01) + T(10) - 2T(00)), \\
&= (1 - 2\lambda)T(00) + \lambda T(01) + \lambda T(10).
\end{aligned} \tag{56}
$$

The partial derivative of L with respect to α simplifies as follows:

$$
\begin{aligned}
(\log_e 2)\frac{\partial L}{\partial \alpha} &= (1 - 2\lambda)\frac{-\epsilon(1 - \epsilon)}{\alpha} + \lambda\frac{\epsilon(1 - \epsilon)}{\overline{\alpha}} + \lambda\frac{-\epsilon(1 - \epsilon)}{\alpha} \\
&= \epsilon(1 - \epsilon)\left[\frac{\lambda}{1 - \alpha} - \frac{1 - \lambda}{\alpha}\right].
\end{aligned} \tag{57}
$$

Equating to zero, we get

$$
\alpha = 1 - \lambda. \tag{58}
$$

Similarly, equating the partial derivative of L with respect to β to zero, we get

$$
\lambda = \frac{1 - \beta}{2 - \beta}. \tag{59}
$$

Using (58) and (59) in the objective function $T(00)$ and the constraint $T(00) + T(01) - 2T(00) = 0$, we get the statement of Theorem 3 - Part (1).

B. $(1, \infty)$-BEC(ϵ): Theorem 3 - Part (2)

The metrics in $G_{1,\infty}^2$ are as follows:

$$T(000) = (1-\epsilon)^3 \log_2 \frac{1}{\alpha_{00}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\alpha_{0?}} + \epsilon(1-\epsilon)^2 \log \frac{1}{\alpha_{?0}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\alpha_{??}},$$

$$T(100) = (1-\epsilon)^3 \log_2 \frac{1}{\alpha_{10}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\alpha_{1?}} + \epsilon(1-\epsilon)^2 \log \frac{1}{\alpha_{?0}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\alpha_{??}},$$

$$T(010) = \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\alpha_{0?}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\alpha_{??}},$$

$$T(001) = (1-\epsilon)^3 \log_2 \frac{1}{\overline{\alpha}_{00}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\overline{\alpha}_{0?}} + \epsilon(1-\epsilon)^2 \log \frac{1}{\overline{\alpha}_{?0}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\overline{\alpha}_{??}},$$

$$T(100) = (1-\epsilon)^3 \log_2 \frac{1}{\overline{\alpha}_{10}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\overline{\alpha}_{1?}} + \epsilon(1-\epsilon)^2 \log \frac{1}{\overline{\alpha}_{?0}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\overline{\alpha}_{??}}.$$

The Lagrangian for the optimization problem in (39) is

$$L = (1 - 2\lambda_1 - 3\lambda_3)T(000) + (\lambda_1 + \lambda_2)T(010) + \lambda_1 T(101) + \lambda_2 T(100) + \lambda_2 T(001). \tag{60}$$

Equating the partial derivatives with respect to α_{rs}, $(rs) \in \{00, 0?, ?0, ??, 10, 1?\}$, we obtain the following relationships:

$$\alpha_{00} = 1 - \frac{\lambda_2}{1 - 2\lambda_1 - 2\lambda_2}, \quad \alpha_{0?} = 1 - \frac{\lambda_2}{1 - \lambda_1 - \lambda_2},$$

$$\alpha_{?0} = 1 - \frac{\lambda_1 + \lambda_2}{1 - \lambda_1 - \lambda_2}, \quad \alpha_{??} = 1 - \lambda_1 - \lambda_2,$$

$$\alpha_{10} = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \quad \alpha_{1?} = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Setting $\beta = \alpha_{00}$ and $\alpha = \alpha_{??}$, we express all variables in terms of α and β. Expressing the objective function and the constraints in terms of α and β and simplifying results in the statement of Theorem 3 - Part (2).

C. $(1,2)$-BEC(ϵ): Theorem 4

1) Part (1): The metrics in $G_{1,2}^1$ are as follows:

$$T(001) = \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\overline{\alpha}_{0?}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\overline{\alpha}_{?0}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\overline{\alpha}_{??}}$$

$$T(010) = \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\alpha_{0?}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\alpha_{??}}$$

$$T(100) = (1-\epsilon)^3 \log_2 \frac{1}{\alpha_{10}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\alpha_{?0}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\alpha_{1?}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\alpha_{??}}$$

$$T(101) = (1-\epsilon)^3 \log_2 \frac{1}{\overline{\alpha}_{10}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\overline{\alpha}_{?0}} + \epsilon(1-\epsilon)^2 \log_2 \frac{1}{\overline{\alpha}_{1?}} + \epsilon^2(1-\epsilon) \log_2 \frac{1}{\overline{\alpha}_{??}}$$

The Lagrangian for the optimization problem in (41) can be written as

$$L = \frac{1}{2}(T(010) + T(101)) + \lambda(T(010) + 3T(101) - 2T(100) - 2T(001))$$

$$= (\frac{1}{2} + \lambda)T(010) + (\frac{1}{2} + 3\lambda)T(101) - 2\lambda T(100) - 2\lambda T(001). \tag{61}$$

Equating partial derivatives of L with respect to the parameters α_{rs}, $(rs) \in \{10, 0?, ?0, 1?, ??\}$, we get

$$\alpha_{10} = \alpha_{1?} = \frac{4\lambda}{1 - 2\lambda}, \quad \alpha_{0?} = \frac{1 - 2\lambda}{1 + 2\lambda},$$

$$\alpha_{?0} = \frac{4\lambda}{1 + 2\lambda}, \quad \alpha_{??} = \frac{1}{2} + \lambda.$$

Setting $\beta = \overline{\alpha}_{10} = \frac{1-6\lambda}{1-2\lambda}$, we express all variables in terms of β. Expressing the objective function and the constraint in terms of β and simplifying results in the statement of Theorem 4 - Part (1).

2) Part (2): The metrics in $G_{1,2}^2$ are as follows:

$$T(0101) = (1 - \epsilon)\left[\epsilon^3 \log_2 \frac{1}{\overline{\alpha}_{???}} + \epsilon^2(1 - \epsilon)(\log_2 \frac{1}{\overline{\alpha}_{0??}} + \log_2 \frac{1}{\overline{\alpha}_{??0}}) + \right.$$
$$\left. ((1 - \epsilon)^3 + 3\epsilon(1 - \epsilon)^2 + \epsilon^2(1 - \epsilon)) \log_2 \frac{1}{\overline{\alpha}_{010}}\right],$$

$$T(0100) = (1 - \epsilon)\left[\epsilon^3 \log_2 \frac{1}{\alpha_{???}} + \epsilon^2(1 - \epsilon)(\log_2 \frac{1}{\alpha_{0??}} + \log_2 \frac{1}{\alpha_{??0}}) + \right.$$
$$\left. ((1 - \epsilon)^3 + 3\epsilon(1 - \epsilon)^2 + \epsilon^2(1 - \epsilon)) \log_2 \frac{1}{\alpha_{010}}\right],$$

$$T(1010) = (1 - \epsilon)\left[\epsilon^3 \log_2 \frac{1}{\alpha_{???}} + \epsilon(1 - \epsilon)^2 \log_2 \frac{1}{\alpha_{10?}} + \epsilon^2(1 - \epsilon)(\log_2 \frac{1}{\alpha_{?0?}} + \log_2 \frac{1}{\alpha_{1??}})\right],$$

$$T(0010) = (1 - \epsilon)\left[\epsilon^3 \log_2 \frac{1}{\alpha_{???}} + \epsilon^2(1 - \epsilon)(\log_2 \frac{1}{\alpha_{0??}} + \log_2 \frac{1}{\alpha_{?0?}})\right],$$

$$T(1001) = (1 - \epsilon)\left[\epsilon^3 \log_2 \frac{1}{\overline{\alpha}_{???}} + \epsilon(1 - \epsilon)^2 \log_2 \frac{1}{\overline{\alpha}_{10?}} + \epsilon^2(1 - \epsilon)(\log_2 \frac{1}{\overline{\alpha}_{??0}} + \log_2 \frac{1}{\alpha_{?0?}} + \log_2 \frac{1}{\alpha_{1??}})\right].$$

The Lagrangian for the optimization problem in (43) can be written as follows:

$$L = \frac{1}{2}(T(0101) + T(1010)) + \lambda(2T(0010) + 2T(0100) + 2T(1001) - 3T(0101) - 3T(1010))$$
$$= (\frac{1}{2} - 3\lambda)(T(0101) + T(1010)) + 2\lambda(T(0010) + T(0100) + T(1001)). \tag{62}$$

Equating partial derivatives of L with respect to the parameters α_{rst}, $rst \in \{010, 10?, 0??, ?0?, ??0, 1??, ???\}$, we get

$$\alpha_{010} = \frac{4\lambda}{1 - 2\lambda}, \ \alpha_{10?} = \alpha_{1??} = \frac{1 - 6\lambda}{1 - 2\lambda},$$
$$\alpha_{0??} = \frac{8\lambda}{1 + 2\lambda}, \ \alpha_{??0} = \frac{4\lambda}{1 + 2\lambda},$$
$$\alpha_{?0?} = \frac{1 - 2\lambda}{1 + 2\lambda}, \ \alpha_{???} = \frac{1}{2} + \lambda.$$

Setting $\beta = \overline{\alpha}_{010} = \frac{1 - 6\lambda}{1 - 2\lambda}$, we express all variables in terms of β. Expressing the objective function and the constraint in terms of β and simplifying results in the statement of Theorem 4 - Part (2).

D. (d, ∞)-BEC(ϵ): Theorem 5

The metrics in the graph $G_{d,\infty}^d$ can be written as follows:

$$T(u_00) = \sum_{k=0}^{d} \binom{d}{k} \epsilon^k (1 - \epsilon)^{d-k+1} \log_2 \frac{1}{\alpha_k},$$
$$T(u_i0) = \sum_{k=1}^{d} \binom{d-1}{k-1} \epsilon^k (1 - \epsilon)^{d-k+1} \log_2 \frac{1}{\alpha_k},$$
$$T(u_01) = \sum_{k=0}^{d} \binom{d}{k} \epsilon^k (1 - \epsilon)^{d-k+1} \log_2 \frac{1}{\overline{\alpha}_k},$$

The Lagrangian for the optimization problem in (46) is

$$L = T(u_00) + \lambda(T(u_01) + \sum_{i=1}^{d} T(u_i0) - (d + 1)T(u_00)). \tag{63}$$

Equating partial derivatives of L with respect to the parameters α_k, $0 \leq k \leq d$, we get

$$\alpha_k = \frac{(d - k + 1)\lambda - 1}{(d - k)\lambda - 1}. \tag{64}$$

Expressing all variables in terms of α_0 and simplifying the objective function and the constraint results in the statement of Theorem 5.