On the Capacity of Generalized Ising Channels

Artyom Sharov  Ron M. Roth
Computer Science Department, Technion, Haifa 32000, Israel.
{sharov, ronny}@cs.technion.ac.il

Abstract—Nearly tight lower and upper bounds on the capacity of generalized Ising channels are presented. For the case where feedback is allowed, a closed-form expression for the capacity is found for channel error probability \( p \in [0, p_0) \), where \( p_0 \approx 0.398324 \). Two lower bounds on that capacity for larger values of \( p \) are presented.

Index Terms—granular media, dynamic programming, channel capacity, magnetic recording, feedback

I. INTRODUCTION

Let \( \mathbb{Z}^+ \) denote the set of positive integers. For a real parameter \( p \in [0, 1] \), we define a one-dimensional generalized Ising channel \( \mathcal{I}_p \) through the following relation between an input random process \( \mathcal{X} = (X_t)_{t \in \mathbb{Z}^+} \) over \( \Sigma = \{0, 1\} \) and an output random process \( \mathcal{Y} = (Y_t)_{t \in \mathbb{Z}^+} \) over \( \Sigma \), for any \( t \in \mathbb{Z}^+ \):

\[
Y_t = X_t - Z_t, \tag{1}
\]

where \( Z = (Z_t)_{t \in \mathbb{Z}^+} \) is an i.i.d. Bernoulli process characterized by \( p \) (in other words, \( \text{Prob}(Z_t = 1) = p \) independently of the other entries of \( Z \)), and \( X_0 \overset{\text{iid}}{=} 1 \). Defining a state process \( S = (S_t)_{t \in \mathbb{Z}^+} \) by \( S_t = X_{t-1} \), gives an alternative way of representing the channel \( \mathcal{I}_p \), specifically,

\[
Y_t = \begin{cases} 
X_t & \text{with probability } 1-p \\
S_t & \text{with probability } p
\end{cases}
\]

for any \( t \in \mathbb{Z}^+ \). Figure 1 demonstrates how this channel operates at any time \( t \) given the state \( S_t \). To the best of our knowledge, this family of channels was first defined in [10, Sec. 4], where the first lower and upper bounds on its capacity were given.

One particular representative of this family of channels, \( \mathcal{I}_p(0.5) \), is known in the literature as the one-dimensional Ising channel [4], which bears its name due to its affinity to the Ising model in statistical mechanics [9][11]. Berger and Bonomi studied the information-theoretic properties of \( \mathcal{I}_p(0.5) \), specifically, they established the value 0.5 of the zero-error capacity [4, Th. 6] and gave the first numeric lower and upper bounds on the value of the channel (Shannon) capacity [4, Th. 4, 5], which was found to be strictly greater than 0.5. The zero-error capacity of \( \mathcal{I}_p(0.5) \), attained by a simple code construction, also appears in more recent works in the context of grain errors under a combinatorial error model [12, Sec. 2], as well as in a probabilistic setting [10, Prop. 5]. Elishco and Permuter [6] introduced feedback to \( \mathcal{I}_p(0.5) \) and found its capacity to be approximately 0.575522 along with a simple zero-error coding scheme achieving the capacity [6, Th. 1, 2]. The idea of adding feedback can be generalized to any channel

\[ S_t = 0 \quad S_t = 1 \]

\[
\begin{array}{c}
0 \quad 1 \quad 0 \quad 1-p \\
X_t \quad Y_t \quad X_t \quad p \\
1 \quad 1-p \quad 1 \quad 1
\end{array}
\]

Fig. 1. The channel \( \mathcal{I}_p \) as a function of the state \( S_t \).

\[ \mathcal{I}_{\text{FB}}(p) \]

with \( p \in [0, 1] \), thereby resulting in a new channel denoted hereafter by \( \mathcal{I}_{\text{FB}}(p) \). Figure 2 shows schematically how decoding of the information message \( M \) to \( \widetilde{M} \) is performed in \( \mathcal{I}_{\text{FB}}(p) \). The idea of [6], as well as of several other earlier papers [2][13], is based on the observation that the capacity of certain channels with feedback can be recast as an infinite-horizon dynamic program [5, Ch. 8], whose optimal average reward equals the capacity of the channel. This technique was proposed by Tatikonda in his Ph.D. thesis [15], followed by his subsequent works [16][18][19]. The optimal average reward of the dynamic program in these works was found by solving the matching Bellman equation [1, Sec. 3][3].

The study of \( \mathcal{I}_p \) is motivated by the behavior of bit-patterned media during shingled writing [8][10, Sec. 3], where the magnetic field emanating from the write head may cause overlapping patterns of substitution errors, resembling those of \( \mathcal{I}_p \). In our earlier work on granular media [14], we modeled these substitution errors as overlapping grain errors. The storage capacity of the medium can be increased by supplementing the shingled magnetic recording with a hardware that detects during a write whether the previously-written bit along the track has been overrun by the current write (yet no backtracking is allowed to correct that bit); it can be shown that introducing such a capability is equivalent

\[ \mathcal{I}_{\text{FB}}(p) \]

Fig. 2. The channel \( \mathcal{I}_{\text{FB}}(p) \).

This work was supported in part by Grant No. 1092/12 from the Israel Science Foundation.
II. Bounds on \( \text{cap}(p) \)

Let \( \{s, s\}' \) denote the set \( \{s, s+1, s+2, \ldots, s'\} \) for any \( s, s' \in \mathbb{Z}^+ \) such that \( s \leq s' \). For an infinite vector \( (A_t)_{t \in \mathbb{Z}^+} \) and \( s, s' \in \mathbb{Z}^+ \) such that \( s \leq s' \), let \( A_s^{s'} \) denote the finite sub-vector \( (A_t)_{t \in \mathbb{Z}^+} \) related by \( (1) \), let \( I(X_1^n; Y_1^n) \) be the mutual information between \( X_1^n \) and \( Y_1^n \), and denote by \( \text{cap}(p) \) the capacity of \( \text{Is}(p) \), namely,

\[
\text{cap}(p) = \sup_{\alpha, \beta} \liminf_{n \to \infty} \frac{1}{n} I(X_1^n; Y_1^n)
\]

[7, Sec. 4.6, 5.9]. In the following two theorems we establish improved lower and upper bounds on \( \text{cap}(p) \).

**Theorem 2.1:** Let \( p \in [0, 1] \) and let \( \alpha, \zeta \in \mathbb{Z}^+ \). Then

\[
\text{cap}(p) \geq \min (\mu, \zeta, p) \overset{\Delta}{=} \sup_{\alpha, \beta} \min \left\{ H(Y_{\zeta+1} | Y_\zeta X_1^n) - H(p) \cdot \text{Prob}(X_2 \neq X_1) \right\},
\]

where \( M_\mu \) stands for the set of all the stationary Markov chains of order \( \mu \) over \( \Sigma \) and \( H(\cdot|\cdot) \) stands for the conditional binary entropy function.

**Theorem 2.2:** Let \( p \in [0, 1] \) and let \( \alpha_0, \alpha_{1;1}, \alpha_1, \alpha_{1;2}, \alpha_2, \alpha_3 \in [0, 1] \) be real numbers such that

\[
\alpha_{1;1} \leq \alpha_1, \alpha_{1;2} \leq \alpha_2, \quad \sum_{i \in \{1,3\}} \alpha_i \leq \alpha_0, \quad 4\alpha_0 - \sum_{i \in \{1,3\}} (4-i)\alpha_i \leq 1.
\]

For each \( i \in \{0, 2\} \), let \( \beta_i \in [0, 1] \) be real numbers such that

\[
\beta_0 \overset{\Delta}{=} \alpha_0 - 2p(1-p)\alpha_1,
\]

\[
\beta_{1;1} \overset{\Delta}{=} p^2 \alpha_1 + p(1-p)\alpha_2 + (\alpha_1 - p\alpha_{1;1})(1-p)^2,
\]

\[
\beta_{1;2} \overset{\Delta}{=} p^2 \alpha_2 + p(1-p)(\alpha_3 + \alpha_1 - p\alpha_{1;1})(1-p)^2 + (\alpha_2 - p\alpha_{1;2})(1-p)^2.
\]

Also, let

\[
\begin{align*}
& t(\beta_0, \beta_1, \beta_2) \overset{\Delta}{=} (1-2\beta_0 + \beta_1)H\left(\frac{\beta_2 - \beta_1 - \beta_2}{1-2\beta_0 + \beta_1}\right) \\
& \quad + \beta_0 \left(\frac{\beta_2}{\beta_0}\right) + (\beta_0 - \beta_1)H\left(\frac{\beta_3 - \beta_0}{\beta_0 - \beta_1}\right).
\end{align*}
\]

Then

\[
\text{cap}(p) \leq \rho(p) \overset{\Delta}{=} \max_{(\alpha_0, \alpha_{1;1}, \alpha_1, \alpha_{1;2}, \alpha_2, \alpha_3) \in U} t(\beta_0, \beta_1, \beta_2) - \alpha_0 \cdot H(p),
\]

where \( U \subseteq [0, 1]^6 \) is the region given by the inequalities (2).

The proofs of Theorems 2.1 and 2.2, which are omitted due to space limitations, follow a standard technique of bounding the mutual information from below and above. The lower bound of Theorem 2.1 arises after assuming a certain Markovian property on the distribution of \( \mathcal{X} \), whereas the upper bound of Theorem 2.2 emerges as a result of employing the method of types both on \( \mathcal{X} \) and \( \mathcal{Y} \).

The quantities \( \alpha_0, \alpha_{1;1}, \alpha_1, \alpha_{1;2}, \alpha_2, \alpha_3 \) in Theorem 2.2 stand for numbers of runs of various lengths in the typical word \( x \) of a capacity-achieving probability distribution of the input process \( X^n \), normalized by the word length \( n \). As a result of the transmission of any such \( x \), any typical output word will have asymptotically the same number \( \beta_0 \) of runs and the same number \( \beta_1 \) of runs of length \( i \), for each \( i \in \{1, 2\} \), per symbol. The quantity \( t(\beta_0, \beta_1, \beta_2) \) stands for the growth rate (with respect to \( n \)) of a set of words of length \( n \) with \( \beta_0 \) runs and \( \beta_1 \) runs of length \( i \), for each \( i \in \{1, 2\} \), per symbol.

Figure 3 shows \( \rho(\mu=6, \zeta=7, p) \) and \( \rho(p) \) alongside the best lower bound \( C_{M_1}(p) \) derived from [10, Eq. (12)] and the best upper bound \( C_{ge}(p) \overset{\triangle}{=} 1-p(1-p) \) from [10, Eq. (3)] on \( \text{cap}(p) \). Due to [10, Prop. 1], which shows that \( \text{cap}(p) = \text{cap}(1-p) \), we can limit ourselves to the range \( [0, 0.5] \) of \( p \) (as is done in Figure 3), since all the lower and upper bounds in Figure 3 on the range \([0.5, 1]\) are obtained by their reflection with respect to \( p = 0.5 \). One can observe the improvement of \( \rho(6, 7, p) \) and \( \rho(p) \) over \( C_{M_1}(p) \) and \( C_{ge}(p) \), and the near-tightness of the new lower and upper bounds.

**Remark 2.3:** Using a generalization of the classical Blahut-Arimoto algorithm [17], one can obtain another set of lower bounds on \( \text{cap}(p) \); however, with comparable computational power and running time, the bounds that we obtained with this method were looser than \( \rho(\mu, \zeta, p) \).

III. THE VALUE OF \( \text{cap}_{FB}(p) \) FOR THE LOW RANGE OF \( p \)

The main result of this section is the following theorem whose proof will take the rest of this section.
Theorem 3.1: Let $p_0 \approx 0.398324$ be the unique solution of the equality \[ \frac{1}{1-p} = \frac{2^p}{2^{p+1}} \] on $[0, 1]$. Then, for $p \in [0, p_0]$, \[ \text{cap}_{\text{FB}}(p) = H\left(\frac{1}{2^{p+1}}\right) - H\left(\frac{p}{2^{p+1}}\right). \]

In Section III-A, we present a general formulation of an infinite-horizon average-reward dynamic program (DP), which is instantiated in Section III-B with quantities related to $\text{Is}_{\text{FB}}(p)$ to represent the problem of finding $\text{cap}_{\text{FB}}(p)$ as a DP. In Section III-C, we present the specialization of the Bellman equation to our case by substituting the quantities from Section III-B into the general form of that equation. All three sections closely follow the presentation in [6, Sec. 4], albeit with several changes in notation.

A. DP formulation

The definitions of this section can be found in the survey [1] on average-cost dynamic programming.

An average-reward DP problem describes a dynamic system evolving in (discrete) time $t \in \mathbb{Z}^+$ according to the septuple $\mathcal{T} = (B, A, D, \iota, \sigma, f, r)$, where

- $B$ is a Borel space containing all the system states $b_t$;
- $A$ is a compact subset of a Borel space containing all the system actions $a_t$;
- $D$ is a measurable space of all system disturbances $d_t$;
- $\iota$ is a probability distribution of the initial state $b_0$;
- $\sigma : D \times B \times A \to [0, 1]$ is a conditional probability distribution of the disturbance $d_t$ given $b_t$ and $a_t$;
- $f : B \times A \times D \to B$ is the function by which the system evolves from one state to another in the following fashion: for any $t \in \mathbb{Z}^+$, $b_{t+1} = f(b_t, a_t, d_t)$;
- $r : B \times A \to \mathbb{R}$ is a bounded real reward function.

The system transitions from one state $b_t$ to another in discrete time $t \in \mathbb{Z}^+$ while experiencing disturbances $d_t$ and undertaking actions $a_t$ as a response to the disturbances. The goal of the system is to maximize the average reward gained by it over time, namely, to maximize

$$\phi = \phi(\mathcal{T}) = \sup_{\pi} \lim_{n \to \infty} \frac{1}{n} E\left(\sum_{t \in \{1, n\}} r(b_t, a_t)\right),$$

where $\pi = (\pi_t)_{t \in \mathbb{Z}^+}$ is the system policy, given by an infinite vector of functions $\pi_t : B \times D_{t-1} \to \mathcal{A}$ mapping $(b_1, d_1, d_2, d_3, \ldots, d_{t-1})$ into $a_t$; in other words, $\pi_t$ determines which action is undertaken by the system given the initial state and the entire history of disturbances prior to time $t$, for each $t \in \mathbb{Z}^+$.

B. Recasting the problem of finding $\text{cap}_{FB}(p)$ as a DP

Given channel $\text{Is}_{\text{FB}}(p)$ with channel input process $X = (X_t)_{t \in \mathbb{Z}^+}$, output process $Y = (Y_t)_{t \in \mathbb{Z}^+}$, and state process $S = (S_t)_{t \in \mathbb{Z}^+}$ taking values $(x_t), (y_t), \text{ and } (s_t)$, respectively, we associate the following quantities with the components of the septuple $\mathcal{T}$ for all $t \in \mathbb{Z}^+$.

Define the DP state at time $t$ as the following vector of length 2

$$b_t = (b_t(s))_{s \in \Sigma} \overset{\Delta}{=} (\text{Prob}(S_t = s \mid Y_1^{t-1} = y_1^{t-1}))_{s \in \Sigma},$$

where $y_1^{t-1} = (y_t)_{t=1}^{t-1} \in \Sigma^{t-1}$. The DP action at time $t$ is given by the $2 \times 2$ stochastic matrix

$$a_t = (a_t(s, x))_{s, x \in \Sigma} \overset{\Delta}{=} (\text{Prob}(X_t = x \mid S_t = s, Y_1^{t-1} = y_1^{t-1}))_{s, x \in \Sigma};$$

the DP disturbance at time $t$ is simply $d_t \overset{\Delta}{=} y_t$. Set

$$\zeta(y, s, x) \overset{\Delta}{=} \text{Prob}(Y_t = y | S_t = s, X_t = x),$$

and notice that this distribution is completely defined by the channel model, as it was presented in Figure 1. The initial distribution $\iota$ of $b_1$ is defined as $\iota(s) = 1$ at $s = 1$ and $\iota(s) = 0$ otherwise; the disturbance distribution is given by

$$\sigma(y \mid b_t, a_t) = \sum_{s, s' \in \Sigma} b_t(s) a_t(s', s') \zeta(y, s, s').$$

The evolution function $f$ can be expressed recursively for $\tilde{s} \in \Sigma$ and $t \in \mathbb{Z}^+$ as

$$b_{t+1}(\tilde{s}) = \sum_{s, s' \in \Sigma} b_t(s) a_t(s', s') \zeta(y_t, s, s'),$$

finally, the reward function is the conditional mutual information

$$r(b_t, a_t) \overset{\Delta}{=} I(S_t, X_t; Y_t | Y_1^{t-1}),$$

which depends only on the distribution $\text{Prob}(S_t = s, X_t = x, Y_t = y \mid Y_1^{t-1} = y_1^{t-1})$, which, in turn, is a product of $b_t(s), a_t(s, x)$ and $\zeta(y, s, x)$, as they appeared in (4)–(6).

Substituting (9) into (3) and noticing that looking for the supremum on the range of policies $\pi$ in our case is equivalent to optimizing on the distribution $\text{Prob}(X_t | S_t, Y_1^{t-1})$ yields

$$\phi = \sup_{\pi \in \text{Prob}(X_t \mid S_t, Y_1^{t-1})} \lim_{n \to \infty} \frac{1}{n} E\left(\sum_{t \in \{1, n\}} I(S_t, X_t; Y_t | Y_1^{t-1})\right).$$

It was proved in [13, Th. 3] that the right-hand side of (10) equals the capacity of the underlying channel if its state process $S$ evolves according to the equation $s_{t+1} = g(s_t, x_t, y_t)$ for some function $g(\cdot)$ and if there is a nonzero probability of reaching any channel state $s$ from any other channel state $s'$ in a finite number of steps. Clearly, both prerequisites hold in our case, therefore (10) transforms to merely $\phi = \text{cap}_{\text{FB}}(p)$.

C. Statement and solution of the Bellman equation

The Bellman equation in its general form ([1, Th. 2.1]), which comes next, gives an alternative characterization of the optimal average reward $\phi$. We present a slightly modified discrete form of that integral equation below (see [1, Th. 5.1]).

Theorem $3.2 \overset{(11)}{\text{: The optimal average reward of a dynamic system characterized by } \mathcal{T} \text{ equals } \phi \in \mathbb{R} \text{ if there exists a function } h : B \to \mathbb{R} \text{ such that for all } b \in B$

$$\phi + h(b) = \sup_{a \in \mathcal{A}} \left( r(b, a) + \sum_{d \in \mathcal{D}} \sigma(d | b, a) h(f(b, a, d)) \right). \quad (11)$$

To state a specialized form of the identity (11), we first represent the expressions (7)–(9) as functions of $b_t(0), \gamma_t \overset{\Delta}{=} \frac{\text{cap}_{\text{FB}}(p)}{2^p}$.

2To see how the recurrence (8) is obtained, the reader is referred to [13, Eq. (35)].
describes both an encoder $E$ from the set of binary vectors of length $A$. Generalization of the encoder from [6, Th. 2] (Section IV-A), whereas the second one is an iterative technique (14), respectively:

\[
\begin{align*}
\sigma(0 | b_t, a_t) &= \sigma \\
\sigma(1 | b_t, a_t) &= 1 - \sigma,
\end{align*}
\]

transmitting the following bits of the message $M$ in order. Next, we consider the case where the last two received bits $Y_t$ and $Y_{t-1}$ are the same. This can happen either when the last transmitted bits are the same ($X_t = X_{t-1}$), or when there was an error in the last transmitted bit and no error in the previous one ($X_{t-1} = Y_{t-1}$, and $X_t \neq Y_t$); in both events we have $X_t = Y_{t-1}$, enter the state $S_0$, and, on the next time slot $t+1$, retransmit the last bit $X_t$. From the state $S_0$ we always transition back to the state $S_1$. This way, $(1-p)(\xi \pm \epsilon)n$ bits of $M$ are transmitted once and the remaining $(1-(1-p)(\xi \pm \epsilon))n$ bits — twice. Since the exponential growth rate of words with $(\xi \pm \epsilon)n$ runs, for $\epsilon \to 0$, tends to $H(\xi)$, the average rate of the encoder in Figure 4 is

\[
\eta_{EP}(p) = \frac{H(\xi)}{1 - \xi(1-p) + H(\xi)(1-\xi(1-p))} = \frac{H(\xi)}{2 - \xi(1-p)}. \tag{16}
\]

The decoder $\mathcal{D}$ is equivalent to the decoder given in [6, Th. 2]; for completeness, we briefly describe its operation next. It starts in the state $S_1$ upon receiving the first bit $Y_1 = 1$ from the channel. In the state $S_1$ at time $t$, we always decode the next message bit as $Y_{t+1}$, whereas in the state $S_0$ we wait one time slot. As with the encoder $\mathcal{E}$, we remain in $S_1$, as long as the last two received bits are different ($Y_t \neq Y_{t-1}$) and transition to $S_0$ otherwise; from $S_0$ we deterministically transition to $S_1$. This scheme guarantees error-free decoding, as the encoder at time $t+1$ always retransmits the $t$-th bit, once it detects that $Y_t = Y_{t-1}$, hence after waiting one time slot in the state $S_0$, we are guaranteed to receive $Y_{t+1} = X_t$, which represents the next message bit sent by the encoder.

The aforementioned decoding scheme was proved in [6] to be capacity-achieving at $p = 0.5$ with code rate $\approx 0.575522$. Surprisingly, it is also capacity-achieving at $p = p_0$, where $\text{cap}_{\text{FB}}(p)$, as it appears in Corollary 3.3, and the code rate $\eta_{EP}(p)$ from (16) can be shown to be equal (to approximately 0.595045).

B. Iterative method

Another technique for obtaining simple lower bounds on $\text{cap}_{\text{FB}}(p)$ is to start with a "guess" $h_0$ bounding the function $h^*$ from below, and perform the following iterations for $i \in \mathbb{Z}^+$:

\[
h_i(b) = \sup_{(\gamma, \delta, b) \in [0,1] \times [0,1]} u(\gamma, \delta, b; h_{i-1}), \quad b \in [0,1]
\]

Since for any solution $h$ of the Bellman equation (15), the function $h+c$ is also a solution for any constant $c \in \mathbb{R}$, for the

IV. LOWER BOUNDS ON $\text{cap}_{\text{FB}}(p)$ FOR $p \in (p_0, 1]$

In this section, we present two lower bounds on $\text{cap}_{\text{FB}}(p)$, for values of $p$ in the range $(p_0, 1]$. The first lower bound is a generalization of the coding scheme of [6, Th. 2] (presented in Section IV-A), whereas the second one is an iterative technique based on the Bellman equation (presented in Section IV-B).

A. Generalization of the encoder from [6, Th. 2]

The finite-state machine appearing in Figure 4 schematically describes both an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$ for $I_{\text{FB}}(p)$. The input to the encoder is a message $M = (M_i)_{i \in \{1, n\}}$ drawn from the set of binary vectors of length $n$ starting with a 1, whose number of runs is within $(\xi \pm \epsilon)n$ for an arbitrarily small $\epsilon > 0$, where $\xi = \xi(p) \in [0, 1]$ maximizes the expression $\frac{H(\xi)}{2 - \xi(1-p)}$. In other words, the probability of alternation from a 0 to a 1 or from a 1 to a 0, as we move along the bits of $M$, is within $(\xi \pm \epsilon)$. After transmitting the first bit $X_1 = M_1 = 1$ (and receiving the first output bit $Y_1 = 1$ through the feedback), we enter state $S_1$ of $\mathcal{E}$. We remain in this state as long as the bit $Y_t$ received through the feedback at time $t$ is different from the previously received bit $Y_{t-1}$ and keep

3Messages with a prescribed number of runs can be generated from the i.i.d. and uniformly distributed input bits via enumerative coding.
purposes of this method we can choose $h^*(0) = 0$. Using the fact that $u(\gamma, \delta; b) \mid h$ is monotonically increasing in $h$, one can prove by induction that $h_i(b) \leq h^*(b) + \gamma \cdot \phi^*$ for any $i \in \mathbb{Z}^+ \cup \{0\}$ and $b \in [0, 1]$, thus $h_0(b) \leq \phi^*$. A natural candidate for the initial guess is $h_0 \equiv 0$. One technical difficulty arising while calculating lower bounds of this kind is the continuous nature of the functions $h_i$. While it is relatively easy to find $h_1$ analytically, the computation of the functions $h_i$ for larger values of $i$ becomes unwieldy. This problem, in turn, is remedied by discretizing the values of the parameter $p \in [0, 1]$ at the expense of some slack in the bound. Figure 5 shows a procedure that gets a vector $\chi = (\chi_j)_{j \in (0, \lambda)}$ of length $\lambda+1$, representing a discretized guess of the function $h^*$, bounding it from below, and returns a vector $\chi' = (\chi'_j)_{j \in (0, \lambda)}$ of the same length, which is an improvement over $\chi$. With a concave initial guess $h_0$, one can prove the concavity of all the functions $h_i$, wherefrom one obtains that $\chi'$ is a lower bound on $h_i(b)$; we omit this proof due to space limitations. We can apply this procedure till the difference between two consecutive lower bounds $\chi^{(m)} - \chi^{(m-1)}$ is sufficiently small. If we start, as suggested before, with the initial guess of $\chi^{(0)} = (0)_{j \in (0, \lambda)}$, the quantity $q_T(p, \lambda) = \chi^{(m)}$ obtained while reaching the stopping condition after $m$ iterations, will represent a lower bound on $\phi^*$. Readily, the larger $\lambda$ is, the tighter the obtained lower bound will be.

Figure 6 shows the curve $p \mapsto \text{cap}_{\text{FB}}(p)$ on the interval $[0, p_0]$ alongside the lower bounds $p \mapsto \varrho_{\text{FP}}(p)$ and $p \mapsto \varrho_{\text{IT}}(p, \lambda=1000)$, which are the best lower bounds on $\text{cap}_{\text{FB}}(p)$ on the intervals $[p_0, 0.525]$ and $[0.525, 1]$, respectively. For $p = p_0$ and $p = 0.5$, explicit capacity-achieving coding schemes are known; the corresponding values of capacity are marked by bold dots in Figure 6. Interestingly, at $p = p_0$ the obtained piecewise curve is $C^1$ continuous: the left derivative of the expression for $\text{cap}_{\text{FB}}(p)$ in Corollary 3.3 and the right derivative of $\varrho_{\text{FP}}(p)$ are equal (to approximately $-0.20111$). Thanks to the near-tightness of the obtained lower bounds, we can also conclude that the curve $p \mapsto \text{cap}_{\text{FB}}(p)$ is not symmetric with respect to $p = 0.5$, unlike the equivalent without feedback [10, Prop. 1]. The dotted curve shows the extension of the curve $p \mapsto \varrho_{\text{FP}}(p)$ to $p < p_0$; one can observe that it is strictly below capacity for $p \in (0, p_0)$.

REFERENCES