A new bound for the capacity of the deletion channel with high deletion probabilities

Marco Dalai, Member, IEEE

Abstract

Let $C(d)$ be the capacity of the deletion channel with deletion probability $d$. It was recently proved by Fertonani and Duman that $\limsup_{d \to 1} C(d)/(1-d) \leq 0.49$. In this paper, it is proved that $\lim_{d \to 1} C(d)/(1-d) = \inf_{d} C(d)/(1-d)$. This result suggests that $C(d)$ may be a convex function of $d$ and it allows to improve the constant 0.49 above to 0.413 using an already available upper bound for $C(0.61)$.

I. INTRODUCTION

A deletion channel $W^d$ is defined to be a channel that drops bits of the input sequence independently with probability $d$. Those bits that are not dropped simply pass through the channel unaltered. While simple to describe, the deletion channel proves to be very difficult to analyze. Dobrushin ([1]) showed that for such a channel it is possible to define a capacity $C(d)$ and that a Shannon like theorem applies to this channel. However, no closed formula expression is known up to now for the capacity $C(d)$, and only upper and lower bounds are currently available (see [2], [3], [4], [5], [6]).

For small values of $d$, it was recently independently proved in [4] and [5] that $C(d) \approx 1 - H(d)$. For values of $d$ close to 1, it is known (see [7], [6]) that $C(d)$ verifies $0.1185(1 - d) \leq C(d) \leq 0.49(1 - d)$.

In this paper, it is proved that

$$\lim_{d \to 1} C(d)/(1-d) = \inf_{d} C(d)/(1-d).$$

This allows to deduce, using currently known bounds, that $C(d) \leq 0.413(1 - d)$ when $d$ approaches 1. Even if equation (I.1) does not formally imply any convexity result$^1$ for $C(d)$ not even in a neighborhood of $d = 1$, the result can be considered at least as an interesting indication of the fact that $C(d)$ may actually be a convex function of $d$.

The main idea used in this paper is that, for a large enough number of input bits $n$, the $d$-deletion channel is fairly well approximated by a channel which drops exactly $dn$ bits selected uniformly at random. In particular, we show that a channel $W_{n,k}$ with $n$-bits input and $k$-bits output, selected uniformly within the $k$-bits subsequences of the input, has a capacity that is close to $C(1 - k/n)$ for large enough $n$. Using this result, we build upon the work in [6] to prove (I.1) and to obtain the improved bound $\lim_{d \to 1} C(d)/(1-d) \leq 0.413$.

The author is with the Department of Information Engineering, University of Brescia, via Branze 38 - 25123, Brescia, Italy. Email: marco.dalai@ing.unibs.it

$^1$It is not difficult to construct functions that satisfy (I.1) but are not convex in any interval $(1 - \epsilon, 1)$.
II. DEFINITION AND REGULARITY OF $C(d)$

Let $W_n^d$ be a channel with an $n$-bit string input whose output is obtained by dropping the bits of the input independently with probability $d$. Let

$$C_n(d) = \frac{1}{n} \max_{p \in \mathcal{P}_n} I(X_n^1; W_n^d(X_n^1)).$$

(II.1)

It was proved by Dobrushin ([1]) that the capacity $C(d)$ of the $d$-deletion channel exists and it holds

$$C(d) = \lim_{n \to \infty} C_n(d).$$

(II.2)

The following lemma gives a quantitative bound on the rate of convergence in (II.2).

Lemma 1: (see also [1], [4], [6]) For every $d \in [0, 1]$ and $n \geq 1$

$$C_n(d) - \frac{\log n}{n} \leq C(d) \leq C_n(d).$$

(II.3)

Proof: As observed in [4], $nC_n(d)$ is subadditive function of $n$. In fact, for an input $X_1^{n+m}$, let $\tilde{Y}(0) = W_n^d(X_n^1)$ and $\tilde{Y}(1) = W_m^d(X_{n+1}^{n+m})$. Note that $Y = W_{n+m}^d(X_1^{n+m})$ can be obtained as a concatenation of the strings $\tilde{Y}(0)$ and $\tilde{Y}(1)$. Thus, $X_1^n \to (\tilde{Y}(0), \tilde{Y}(1)) \to Y$ is a Markov chain. Hence, $(n + m)C_{n+m}(d) = \max_{p \in \mathcal{P}_{X_1^n}} I(X_1^n; Y) \leq \max_{p \in \mathcal{P}_{X_1^n}} I(X_1^n; (\tilde{Y}(0), \tilde{Y}(1))) \leq nC_n(d) + mC_m(d).$ This implies by Fekete’s lemma that the limit $C(d) = \lim_{n \to \infty} C_n(d)$ exists and it verifies $C(d) = \inf_{n \geq 1} C_n(d)$. This proves the right hand side inequality.

Take now an integer $h > 1$ and consider, for an input $X_1^{hn}$, the output $Y = W_{hn}^d(X_1^{hn})$ as the concatenation of the $h$ outputs $\tilde{Y}(i) = W_{n+1}^d(X_{ni+1}^{ni+1})$, $i = 1, \ldots, h-1$. It is clear that $X_1^{hn} \to (\tilde{Y}(0), \tilde{Y}(1), \ldots, \tilde{Y}(h-1)) \to Y$ is a Markov Chain. Let $L_i$ be the length of $\tilde{Y}(i)$. We thus have

$$hnC_{hn}(d) = \max_{p \in \mathcal{P}_{X_1^{hn}}} I(X_1^{hn}, Y)$$

$$= \max_{p \in \mathcal{P}_{X_1^{hn}}} [I(X_1^n; \tilde{Y}(h-1)) - I(X_1^n; \tilde{Y}(h-1)|Y)]$$

$$\geq \max_{p \in \mathcal{P}_{X_1^{hn}}} [I(X_1^n; \tilde{Y}(h-1)) - H(\tilde{Y}(h-1)|Y)]$$

$$= \max_{p \in \mathcal{P}_{X_1^{hn}}} I(X_1^n; \tilde{Y}(h-1)) - H(L_0^{h-1}|Y)]$$

$$\geq \max_{p \in \mathcal{P}_{X_1^{hn}}} I(X_1^n; \tilde{Y}(h-1)) - (h-1) \log n$$

$$= hnC_n(d) - (h - 1) \log n.$$

Hence

$$C(d) = \lim_{h \to \infty} C_{hn}(d)$$

$$\geq \lim_{h \to \infty} \left[ C_n(d) - \frac{h - 1 \log n}{n} \right]$$

$$= C_n(d) - \frac{\log n}{n}.$$ See [6] for a less simple but tighter bound. As a consequence of Lemma 1 we have the following regularity result for $C(d)$.
Lemma 2: The function $C(d)$ is uniformly continuous in $[0, 1]$. Thus, for every $\beta > 0$ there is a $\alpha = \alpha(\beta)$ such that $|d_1 - d_2| < \alpha \Rightarrow |C(d_1) - C(d_2)| < \beta$.

Proof: It is obvious that, for every $n$, $C_n(d)$ is continuous in $d$ since for every $X^n_i$ the mutual information $I(X^n_i; W^n_n(X^n_i))$ is a continuous function of $d$. As shown in Lemma 1, the functions $C_n(d)$ tend to $C(d)$ uniformly in $d$ and thus $C(d)$ is continuous. Since the domain of $C(d)$ is compact, by the Heine-Cantor theorem $C(d)$ is also uniformly continuous.

III. Exact deletion channel

Let now $W_{n,k}$, $k \leq n$, be a channel with $n$-bits input whose output is uniformly chosen within the $\binom{n}{k}$ $k$-bits subsequences of the input. This channel was efficiently used as an auxiliary channel in [5], [6]. Let then

$$C_{n,k} = \frac{1}{n} \max_{f_{X^n_i}} I(X^n_i; W_{n,k}(X^n_i)).$$

(III.1)

Lemma 3: For every random $X^n_i$, if $k_1 \geq k_2$ then

$$I(X^n_i; W_{n,k_1}(X^n_i)) \geq I(X^n_i; W_{n,k_2}(X^n_i)).$$

(III.2)

Proof: Simply note that the $W_{n,k_2}$ channel can be obtained as a cascade of $W_{n,k_1}$ and $W_{k_1,k_2}$. Thus, $X^n_i \rightarrow W_{n,k_1}(X^n_i) \rightarrow W_{n,k_2}(X^n_i)$ is a Markov chain and the lemma follows from the data processing inequality.

The following lemma bounds the capacity of the $W_n^d$ channel in terms of the capacity of opportune exact deletion channels.

Lemma 4: For every $\varepsilon > 0$, $d \in [\varepsilon, 1 - \varepsilon]$, and $n \geq 1$

$$C_{n,[(1-d-\varepsilon)n]} - 2e^{-\frac{\varepsilon^2}{2}n} \leq C_n(d) \leq C_{n,[(1-d+\varepsilon)n]} + 2e^{-\frac{\varepsilon^2}{2}n}.$$  

(III.3)

Proof: We first prove the right hand side inequality. For an input $X^n_i$, let $Y = W^d_n(X^n_i)$ and let $L = |Y|$ be the length of $Y$. First note that $X^n_i \rightarrow Y \rightarrow L$ is a Markov chain. So, by applying the chain rule to $I(X^n_i; Y, L)$, considered that $I(X^n_i; L) = 0$ since $L$ is independent from $X^n_i$, it is easily seen that $I(X^n_i; Y) = I(X^n_i; Y | L)$. Define $T = \{j : \frac{1}{n} - (1-d) \leq \varepsilon\}$, that is $j \in T$ if and only if $\frac{1 - d - \varepsilon}{n} \leq j \leq \frac{1 - d + \varepsilon}{n}$. Let now $X^n_i$ be distributed according to the optimal distribution for the $W_n^d$ channel. Then we have

$$nC_n(d) = I(X^n_i; Y | L)$$

$$= \sum_{j=0}^{n} p_L(j) I(X^n_i; Y | L = j)$$

$$= \sum_{j \in T} p_L(j) I(X^n_i; Y | L = j) + \sum_{j \notin T} p_L(j) I(X^n_i; Y | L = j)$$

$$\leq \sum_{j \in T} p_L(j) I(X^n_i; Y | L = \left\lfloor (1-d+\varepsilon)n \right\rfloor) + \sum_{j \notin T} p_L(j)n$$

$$\leq nC_{n,\left\lfloor (1-d+\varepsilon)n \right\rfloor} \sum_{j \in T} p_L(j) + n \sum_{j \notin T} p_L(j)n$$

$$\leq nC_{n,\left\lfloor (1-d+\varepsilon)n \right\rfloor} + n \cdot 2e^{-\frac{\varepsilon^2}{2(1-d+\varepsilon)n}}$$

$$\leq nC_{n,\left\lfloor (1-d+\varepsilon)n \right\rfloor} + 2ne^{-\frac{\varepsilon^2}{2(1-d+\varepsilon)n}}.$$
where (a) follows from Lemma 3 and the definition of $T$ and (b) follows from the Chernoff bound. Dividing by $n$ we get the desired inequality.

As for the left hand side inequality, let now $X^n_1$ be distributed according to the optimal distribution for the $W_{n,[(1-d-\varepsilon)n]}$ channel. Then we have

$$nC_n d \geq I(X^n_1;Y|L)$$

$$= \sum_{j=0}^{n} p_L(j) I(X^n_1;Y|L = j)$$

$$= \sum_{j \in T} p_L(j) I(X^n_1;Y|L = j) + \sum_{j \notin T} p_L(j) I(X^n_1;Y|L = j)$$

$$(a) \geq \sum_{j \in T} p_L(j) I(X^n_1;Y|L = [(1-d-\varepsilon)n])$$

$$= nC_n,[(1-d-\varepsilon)n] \sum_{j \in T} p_L(j)$$

$$(b) \geq nC_n,[(1-d+\varepsilon)n](1-2e^{-\frac{\varepsilon^2}{2}n})$$

$$(c) \geq nC_n,[(1-d+\varepsilon)n] - 2ne^{-\frac{\varepsilon^2}{2}n},$$

where (a) follows again from Lemma 3, (b) follows from the Chernoff bound, and (c) follows from the obvious fact that $C_n,[(1-d+\varepsilon)n] \leq 1$. Dividing by $n$ the desired result is obtained. \hfill \blacksquare

The following lemma bounds the capacity of the exact deletion channel $W_{n,k}$ in terms of $C(d)$ for opportune values of $d$.

**Lemma 5:** For every $\varepsilon > 0$ and integers $n$ and $k$

$$C(1-k/n + \varepsilon) - 2e^{-\frac{\varepsilon^2}{2}n} \leq C_{n,k} \leq C(1-k/n - \varepsilon) + 2e^{-\frac{\varepsilon^2}{2}n} + \log n/n. \quad (III.4)$$

**Proof:** Take $d = 1 - k/n - \varepsilon$ in Lemma 4 to obtain $C_{n,k} \leq C_n(1-k/n - \varepsilon) + 2e^{-\frac{\varepsilon^2}{2}n} \leq C(1-k/n - \varepsilon) + 2e^{-\frac{\varepsilon^2}{2}n} + \log n/n$, by virtue of Lemma 1. Then take $d = 1 - k/n + \varepsilon$ in Lemma 4 to obtain $C_{n,k} \geq C_n(1-k/n + \varepsilon) - 2e^{-\frac{\varepsilon^2}{2}n} \geq C(1-k/n + \varepsilon) - 2e^{-\frac{\varepsilon^2}{2}n}$. \hfill \blacksquare

**Lemma 6:** For every $\beta > 0$, there is an $\bar{n} = \bar{n}(\beta)$ such that

$$|C_{n,k} - C(1-k/n)| < \beta \quad \forall n \geq \bar{n}, k = 1, \ldots, n. \quad (III.5)$$

**Proof:** First note that, for $\varepsilon > 0$, $C(1-k/n + \varepsilon) \leq C(1-k/n) \leq C(1-k/n - \varepsilon)$. Hence, $C(1-k/n)$ satisfies the two inequalities satisfied by $C_{n,k}$ in equation (III.4). So, $|C_{n,k} - C(1-k/n)|$ is bounded by the difference between the right hand side and the left hand side of equation (III.4), that is

$$|C_{n,k} - C(1-k/n)| \leq C(1-k/n - \varepsilon) - C(1-k/n + \varepsilon) + 4e^{-\frac{\varepsilon^2}{2}n} + \frac{\log n}{n}. \quad (III.6)$$

With the notation of Lemma 2, take $\varepsilon < \alpha(\beta/2)/2$ so that $C(1-k/n - \varepsilon) - C(1-k/n + \varepsilon) < \beta/2$. Once $\varepsilon$ is fixed, choose $\bar{n}$ such that $4e^{-\frac{\varepsilon^2}{2}n} + \frac{\log n}{n} < \beta/2$ to complete the proof. Note that $\bar{n}$ is a function of $\beta$ only and that the result holds for every $k \leq n$. \hfill \blacksquare

We can now state the first result of this paper.

**Theorem 1:** Let $k_n$ be an integer valued sequence such that $k_n/n$ tends to $1-d$ as $n$ goes to infinity. Then

$$\lim_{n \to \infty} C_{n,k_n} = C(d). \quad (III.7)$$
Proof: It follows easily from Lemma 6 by continuity of $C(d)$. □

The following result from [6] is fundamental.

**Lemma 7 (Fertonani and Duman, [6]):** For every $n, k$

$$\limsup_{d \to 1} \frac{C(d)}{1 - d} \leq \frac{n C_n + 1}{k + 1}. \quad \text{(III.8)}$$

**Remark 1:** In [6] the authors state that, for every $n$ and $k$, $\lim_{d \to 1} C(d)/(1 - d) \leq n C_n + 1/k + 1$. However, we are not aware of a previous formal proof that $\lim_{d \to 1} C(d)/(1 - d)$ exists. This fact is proved in the following theorem.

We are finally ready to state the main result of the paper.

**Theorem 2:** It holds that

$$\lim_{d \to 1} \frac{C(d)}{1 - d} = \inf_{d \in (0, 1)} \frac{C(d)}{1 - d}. \quad \text{(III.9)}$$

Proof: For every $d' \in (0, 1)$, let $k_n$ be a sequence such that $k_n/n$ tends to $1 - d'$. Then the right hand side of (III.8) tends to $C(d')/(1 - d')$. Since $d'$ is arbitrary, Lemma 7 implies that $\limsup_{d \to 1} C(d)/(1 - d) \leq \inf_{d' \in (0, 1)} \frac{C(d')}{1 - d'}$. However, it is obvious that $\inf_{d \in (0, 1)} C(d)/(1 - d) \geq \inf_{d' \in (0, 1)} \frac{C(d')}{1 - d'}$. Thus $\lim_{d \to 1} C(d)/(1 - d)$ exists and it equals $\inf_{d' \in (0, 1)} \frac{C(d')}{1 - d'}$. □

A direct consequence of Theorem 2 is the following improved bound on $C(d)$.

**Corollary 1:**

$$\lim_{d \to 1} \frac{C(d)}{1 - d} \leq 0.413. \quad \text{(III.10)}$$

Proof: As far as the author knows, the best known numerical bound obtained for $\inf_{d} C(d)/(1 - d)$ is 0.4128 obtained using the bound $C(0.61) \leq C_{17}(0.61) = 0.161$, evaluated in [6]. □

IV. ACKNOWLEGDMENTS

The author would like to thank Dario Fertonani for providing numerical data used in Corollary 1.

REFERENCES

[1] R. L. Dobrushin, “Shannon’s theorems for channels with synchronization errors,” *Problems of Information Transmission*, vol. 3, no. 4, pp. 11–26, 1967.
[2] S. Diggavi and M. Grossglauser, “On information transmission over a finite buffer channel,” *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1226–1237, 2006.
[3] M. Drinea and M. Mitzenmacher, “Improved lower bounds for the capacity of i.i.d. deletion and duplication channels,” *IEEE Trans. on Inform. Theory*, vol. 53, no. 8, pp. 2693–2714, 2007.
[4] Y. Kanoria and A. Montanari, “On the deletion channel with small deletion probability,” submitted.
[5] A. Kalai, M. Mitzenmacher, and M. Suda, “Tight asymptotic bounds for the deletion channel with small deletion probabilities,” in *Proc. IEEE Intern. Symp. on Inform. Theory*, 2010.
[6] D. Fertonani and T. M. Duman, “Novel bounds on the capacity of the binary deletion channel,” accepted *IEEE Trans. Inform. Theory*, 2010.
[7] M. Drinea and M. Mitzenmacher, “A simple lower bound for the capacity of the deletion channel,” *IEEE Trans. on Inform. Theory*, vol. 52, no. 10, pp. 4657–4660, 2006.