Comments on "Information Nonanticipative Rate Distortion Function And Its Applications"

Milan S. Derpich

Abstract

In [1, Theorem III.6] it is claimed that, for a one-sided random source $\mathrm{x}_1^\infty = \mathrm{x}_1, \mathrm{x}_2, \ldots$, the search for the non-anticipative (i.e., causal) rate distortion function can be restricted to reconstructions y_1^∞ which are jointly stationary with x_1^∞. In this technical report we show that the proof of [1, Theorem III.6] is invalid because it relies on [1, Theorem III.5], the proof of which, as we also show, is flawed.

I. Introduction

The manuscript [1] utilizes [2, Theorem 4] to prove the claim that, for one-sided sources x_1^∞, the non-anticipative (i.e., causal) rate-distortion function can be realized by a reconstruction process y_1^∞ which is jointly stationary with x_1^∞. To do so, it relies on [1, Theorem III.5].

In this note we argue that the proof of [1, Theorem III.5], and hence that of [1, Theorem III.6], are flawed. For that purpose, we will first recall the assumptions and definitions utilized in [2]. After that, we will present the definitions introduced in [1] and show, under the conditions stated there, the requirements needed by [2, Theorem 4] (the basis of [1, Theorem III.6]) of are not met.

II. A Brief Review of [2]

Throughout [2], the search in the infimizations associated with various types of "nonanticipatory" (i.e., causal) rate-distortion functions is stated over sets of *joint* probability distributions between source and reconstruction (as opposed to the usual definitions, in which the search is over *conditional* distributions, see [3, Chapter 10], [4]). Since the distribution of the source is given, it is required that for every $k_2 > k_1 \in \mathbb{Z}$, all the joint distributions $P_{\mathrm{x}_{k_1}^{k_2}, \mathrm{y}_{k_1}^{k_2}}$ to be considered yield $\mathrm{x}_{k_1}^{k_2}$ having the

Milan S. Derpich is with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Av. España 1680, Valparaíso, Chile.

same (given) distribution of the source for the corresponding block, say $P_{\mathring{x}_{k_1}^{k_2}}$. This requirement can be formalized as requiring that $P_{x_{k_1}^{k_2},y_{k_1}^{k_2}} \in \mathcal{P}^{k_1,k_2}$, for a set of admissible joint distributions $\mathcal{P}^{k_1,k_2}$ defined as

$$\mathcal{P}^{k_1,k_2} \triangleq \left\{ P : P(E \times \mathcal{Y}_{k_1}^{k_2}) = P_{\mathring{x}_{k_1}^{k_2}}(E), \quad \forall E \in \mathcal{B}(\mathcal{X}_{k_1}^{k_2}) \right\}, \quad k_1 \leq k_2 \in \mathbb{Z}, \tag{1}$$

where $\mathcal{X}_{k_1}^{k_2}$ and $\mathcal{Y}_{k_1}^{k_2}$ are, respectively, the alphabets to which $x_{k_1}^{k_2}$ and $y_{k_1}^{k_2}$ belong, and $\mathcal{B}(\mathcal{X}_{k_1}^{k_2})$ is a σ-algebra over $\mathcal{X}_{k_1}^{k_2}$. In [2], this admissibility requirement is embedded in the definition of the sets of distributions which meet the distortion constraint, described next.

The fidelity criterion for every pair of integers[1] $k_1 \leq k_2$ is expressed in [2] as requiring $P_{x_{k_1}^{k_2},y_{k_1}^{k_2}}$ to belong to a non-empty set of distributions (hereafter referred to as *distortion-feasible set*) $\mathcal{W}_D^{k_1,k_2}$, a condition written as $(x_{k_1}^{k_2}, y_{k_1}^{k_2}) \in (\mathcal{W}_D^{k_1,k_2})$. In this definition, the number $D \geq 0$ represents an admissible distortion level. Notice that such general formulation of a fidelity criteria does not need a distortion function and does not necessarily involve an expectation.

As mentioned above, the admissibility requirement $P_{x_{k_1}^{k_2}} \in \mathcal{P}^{k_1,k_2}$ is expressed in the distortion-feasible sets in [2, eqn. (2.1)]. The latter equation can be written as

$$\mathcal{W}_D^{k_1,k_2} \subset \mathcal{P}^{k_1,k_2}. \tag{2}$$

In [2, eqs. (2.4) and (2.5)], the distortion-feasible sets are assumed to satisfy the "concatenation" condition

$$(x_{k_1}^{k_2}, y_{k_1}^{k_2}) \in (\mathcal{W}_D^{k_1,k_2}) \wedge (x_{k_2+1}^{k_3}, y_{k_2+1}^{k_3}) \in (\mathcal{W}_D^{k_2+1,k_3}) \Longrightarrow (x_{k_1}^{k_3}, y_{k_1}^{k_3}) \in (\mathcal{W}_D^{k_1,k_3}). \tag{3}$$

With this, [2, eqn. (2.9)] defined the "nonanticipatory epsilon entropy" of the set of distributions[2] $\mathcal{W}_D^{k_1,k_2}$ as

$$H^0(\mathcal{W}_D^{k_1,k_2}) \triangleq \inf I(x_{k_1}^{k_2}; y_{k_1}^{k_2}), \tag{4}$$

where the infimum is taken over all pairs of random sequences $(x_{k_1}^{k_2}, y_{k_1}^{k_2}) \in (\mathcal{W}_D^{k_1,k_2})$ such that the causality Markov chains

$$x_{k+1}^{k_2} \longleftrightarrow x_{k_1}^{k} \longleftrightarrow y_{k_1}^{k}, \quad k_1 \leq k \leq k_2 \tag{5}$$

are satisfied. Then [2, eq. (2.13)] defines the "nonanticipatory message generation rate" as

$$\overline{H_D^0} \triangleq \lim_{k_2-k_1\to\infty} \frac{1}{k_2-k_1} H^0(\mathcal{W}_D^{k_1,k_2}) \tag{6}$$

[1] The analysis in [2] considered both dicrete- and continuous-time processes, but here we only refer to the discrete-time scenario.

[2] The actual term employed in [2] is "nonanticipatory epsilon entropy of the message $(\mathcal{W}_D^{k_1,k_2})$" where the term "message" refers to the random ensembles in $(\mathcal{W}_D^{k_1,k_2})$.

(when the limit exists).

An alternative "nonanticipatory message generation rate" is also considered in [2] by defining the set of distortion-admissible process distributions $\mathcal{W}_D$ as follows:

Definition 1. *The set* $(\mathcal{W}_D)$ *consists of all two-sided random process pairs* $(\mathrm{x}_{-\infty}^{\infty}, \mathrm{y}_{-\infty}^{\infty}) \in (\mathcal{W}_D)$ *for which there exist integers* $\cdots < k_{-1} < k_0 < k_1 < \cdots$ *such that* $\lim_{i\to\pm\infty} k_i = \pm\infty$ *and*

$$(\mathrm{x}_{k_i}^{k_{i+1}-1}, \mathrm{y}_{k_i}^{k_{i+1}-1}) \in (\mathcal{W}_D^{k_i, k_{i+1}-1}), \quad \forall i \in \mathbb{Z}. \tag{7}$$

▲

With this, [2, eq. (2.14)] defines

$$\overrightarrow{H_D^0} \triangleq \inf \lim_{k_2 - k_1 \to \infty} \frac{1}{k_2 - k_1} I(\mathrm{x}_{k_1}^{k_2}; \mathrm{y}_{k_1}^{k_2}) \tag{8}$$

(when the limit exists), where the infimum is taken over all pairs of processes $(\mathrm{x}_{-\infty}^{\infty}, \mathrm{y}_{-\infty}^{\infty}) \in (\mathcal{W}_D)$ satisfying the causality Markov chains

$$\mathrm{x}_{k+1}^{\infty} \longleftrightarrow \mathrm{x}_{-\infty}^{k} \longleftrightarrow \mathrm{y}_{-\infty}^{k}, \quad \forall k \in \mathbb{Z}. \tag{9}$$

III. The Problems With [1]

The proof of [1, Theorem III.6] relies on the claim stated in [1, Theorem III.5], namely, that an equality similar to

$$\overrightarrow{H_D^0} = \overline{H_D^0} \tag{10}$$

holds.

We demonstrate that the proof of [1, Theorem III.5] is not valid (and hence that of [1, Theorem III.6] is flawed). We do this by showing next that [1, Theorem III.5] has two problems, namely: a) one of the causal IRDFs considered in it does not coincide with $\overrightarrow{H_D^0}$, and b) the proof of [1, Theorem III.5] is invalid.

A. The First Problem

The already mentioned first problem of [1, Theorem III.5] as a basis for [1, Theorem III.6] follows from the fact that [1] defines its alternative causal IRDF function $\overrightarrow{R}^{na}(D)$ as ([1, II.9])

$$\overrightarrow{R}^{na}(D) \triangleq \inf_{P_{\mathrm{y}_1^\infty | \mathrm{x}_1^\infty} \in \overrightarrow{\mathcal{Q}}_{1,\infty}(D)} \lim_{n\to\infty} \frac{1}{n} I(\mathrm{x}_1^n \to \mathrm{y}_1^n), \tag{11}$$

where (as defined in the text just below equation (II.6) in [1]) $\overrightarrow{\mathcal{Q}}_{1,\infty}(D)$ is the set of conditional distributions of y_1^∞ given x_1^∞ such that $(\mathrm{x}_1^\infty, \mathrm{y}_1^\infty)$ satisfies the causality Markov chains

$$\mathrm{x}_{k+1}^{\infty} \longleftrightarrow \mathrm{x}_1^k \longleftrightarrow \mathrm{y}_1^k, \quad k = 1, 2, \ldots. \tag{12}$$

and the asymptotic distortion constraint

$$\lim_{n\to\infty} \frac{1}{n}\,\mathrm{E}[d_{1,n}(\mathrm{x}_1^n, \mathrm{y}_1^n)] \leq D. \tag{13}$$

Next, [1] states in its equation (III.2) that [2] defined

$$\overrightarrow{R}^{\varepsilon}(D) \triangleq \inf_{P_{\mathrm{y}_1^\infty \mid \mathrm{x}_1^\infty} \in \overrightarrow{\mathcal{Q}}_{1,\infty}(D)} \lim_{n\to\infty} \frac{1}{n} I(\mathrm{x}_1^n; \mathrm{y}_1^n). \tag{14}$$

Thanks to (12), it readily follows that $\overrightarrow{R}^{\varepsilon}(D) = \overrightarrow{R}^{na}(D)$ (although this equality is not explicitly stated in [1]).

Since the only causal IRDF defined in [2] as an $\inf \lim$ is $\overrightarrow{H_D^0}$, one must conclude that [1] regards $\overrightarrow{R}^{\varepsilon}(D)$ as equivalent to $\overrightarrow{H_D^0}$. However, in view of Definition 1 and (8), such equivalence is not valid (since the distortion feasible sets of Definition 1 are not compatible with the distortion constraint (13)). Therefore, when in [1, Theorem III.5] it is stated that $R^{na}(D) = \overrightarrow{R}^{na}(D)$ (and hence $\overline{H_D^0} = \overrightarrow{R}^{\varepsilon}(D)$), it does not mean that $\overline{H_D^0}$ equals $\overrightarrow{H_D^0}$. As a consequence, one of the necessary conditions for [2, Theorem 4] is not shown to hold.

B. *The Second Problem*

The second issue with [1, Theorem III.5] is the validity of its proof. To begin with, the only argument used in it is that the source is stationary and [2, Theorem 2]. However, the latter theorem only says that $\overrightarrow{H_D^0} \leq \overline{H_D^0}$, and thus the proof of [1, Theorem III.5] presented there is flawed.

Although not referred to in that proof, the reverse inequality claimed in [1, Lemma III.4] would be all that is required to show that $\overrightarrow{R}^{\varepsilon}(D) = \overline{H_D^0}$. However, the proof of [1, Lemma III.4], reproduced below, is clearly invalid. It starts by noting that, by definition,

$$R_{1,n}^{na}(D) \leq I(\mathrm{x}_1^n, \mathrm{y}_1^n), \quad \forall (\mathrm{x}_1^n, \mathrm{y}_1^n) \in (\overrightarrow{\mathcal{Q}}_{1,n}(D)). \tag{15}$$

Then it proceeds by saying that "taking the limit on both sides we obtain"

$$\lim_{n\to\infty} \frac{1}{n} R_{1,n}^{na}(D) \leq \lim_{n\to\infty} \frac{1}{n} I(\mathrm{x}_1^n, \mathrm{y}_1^n), \quad \forall (\mathrm{x}_1^\infty, \mathrm{y}_1^\infty) \in (\overrightarrow{\mathcal{Q}}_{1,\infty}(D)) \tag{16}$$

and then that the claim follows by taking the infimum over $\overrightarrow{\mathcal{Q}}_{1,\infty}(D)$. The problem with this reasoning is that (16) does not follow from (15). A rigorous reasoning reveals that when taking the limit as $n \to \infty$, (15) translates to

$$\lim_{n\to\infty} R_{1,n}^{na}(D) \leq \lim_{n\to\infty} \frac{1}{n} I({}^{(n)}\mathrm{x}_1^n; {}^{(n)}\mathrm{y}_1^n), \quad \forall \{{}^{(n)}\mathrm{x}_1^n, {}^{(n)}\mathrm{y}_1^n\}_{n\in\mathbb{N}} \text{ such that } P_{{}^{(n)}\mathrm{y}_1^n \mid {}^{(n)}\mathrm{x}_1^n} \in \overrightarrow{\mathcal{Q}}_{1,n}(D) \tag{17}$$

Thus, one cannot choose to infimize the RHS of this inequality over $\overrightarrow{\mathcal{Q}}_{1,\infty}(D)$ and expect the inequality to hold, since one can easily find a pair of processes $(\mathrm{x}_1^\infty, \mathrm{y}_1^\infty)$ whose conditional distribution $P_{\mathrm{y}_1^\infty \mid \mathrm{x}_1^\infty}$ belongs to $\overrightarrow{\mathcal{Q}}_{1,\infty}(D)$ and yet $P_{\mathrm{y}_1^n \mid \mathrm{x}_1^n} \notin \overrightarrow{\mathcal{Q}}_{1,n}(D)$ (because the normalized expectations on the LHS of (13) are allowed to reach the limit D from above).

In order to arrive to (16), one should first show that

$$R^{na}_{1,n}(D) \leq \frac{1}{n} I(\mathrm{x}_1^n, \mathrm{y}_1^n), \quad \forall P_{\mathrm{y}_1^\infty \,|\, \mathrm{x}_1^\infty} \in \overrightarrow{\mathcal{Q}}_{1,\infty}(D). \tag{18}$$

Unfortunately, the latter is not true since, as already mentioned, $\overrightarrow{\mathcal{Q}}_{1,\infty}(D)$ allows pairs of random processes $(\mathrm{x}_1^\infty, \mathrm{y}_1^n)$ such that $\mathrm{E}[d_{1,n}(\mathrm{x}_1^n, \mathrm{y}_1^n)] > D$, for all $n \in \mathbb{N}$ (thus reaching the limit distortion (13) from above), and thus such that $P_{\mathrm{y}_1^n \,|\, \mathrm{x}_1^n} \notin \overrightarrow{\mathcal{Q}}_{1,n}(D)$, for all $n \in \mathbb{N}$. Therefore, (18) does not hold. Indeed, the latter reasoning reveals that

$$R^{na}_{1,n}(D) \geq \inf_{(\mathrm{x}_1^\infty, \mathrm{y}_1^\infty) \in (\overrightarrow{\mathcal{Q}}_{1,\infty}(D))} \frac{1}{n} I(\mathrm{x}_1^n, \mathrm{y}_1^n), \quad n \in \mathbb{N}, \tag{19}$$

leading to an inequality in the same direction as the one provided by [2, Theorem 2], i.e., that $R^{na}(D) \geq \overrightarrow{R}^{\varepsilon}(D)$.

References

[1] P. Stavrou, C. K. Kourtellaris, and C. D. Charalambous, "Information nonanticipative rate distortion function and its applications," *CoRR*, vol. abs/1405.1593v2, 2015. [Online]. Available: http://arxiv.org/abs/1405.1593v2

[2] A. Gorbunov and M. Pinsker, "Non anticipatory and prognostic epsilon entropies and message generation rates," *Probl. Inf. Transm.*, vol. 9, no. 3, pp. 12–21, July–Sept. 1973, translation from Problemi Peredachi Informatsii, vol. 23, no. 2, pp. 3–8, April-June 1973.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, N.J: Wiley-Interscience, 2006.

[4] T. Berger, *Rate distortion theory: a mathematical basis for data compression*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.