# One shot approach to lossy source coding under average distortion constraints 

Nir Elkayam Meir Feder<br>Department of Electrical Engineering - Systems<br>Tel-Aviv University, Israel<br>Email: nirelkayam@post.tau.ac.il, meir@eng.tau.ac.il


#### Abstract

This paper presents a one shot analysis of the lossy compression problem under average distortion constraints. We calculate the exact expected distortion of a random code. The result is given as an integral formula using a newly defined functional $\tilde{D}\left(z, Q_{Y}\right)$ where $Q_{Y}$ is the random coding distribution and $z \in[0,1]$. When we plug in the code distribution as $Q_{Y}$, this functional produces the average distortion of the code, thus provide a converse result utilizing the same functional. Two alternative formulas are provided for $\tilde{D}\left(z, Q_{Y}\right)$, the first involves a supremum over some auxiliary distribution $Q_{X}$ which has resemblance to the channel coding meta-converse and the other involves an infimum over channels which resemble the well known Shannon distortion-rate function.


## I. Introduction

The single shot approach aims to find informational quantities that govern the optimal performance of an operational problems of interest, e.g., channel coding and lossy compression. In both cases, the problem settings pose a random object that we want to control. In the channel coding problem this is the random channel which abstracts the medium we want to use to enable a reliable communication. In the lossy compression this is the source we want to "compress" to a minimum number of bits subject to a distortion constraint. The single shot approach tries to solve the problem by providing achievable and converse bounds without any assumption on the random object.

A "good" solution should have the following properties:

1) Tightness: The relation between the achievable and converse bounds should be clarifies and quantified. Preferably, the gap between the bounds should be "small".
2) Computation: The bounds should be computable. Since we generally deal with a high dimensional problem space for which the exact description might not even be feasible, we relax the computability to convexity, i.e., the bounds should be presented as a minimization of some convex function on some convex domain. For such a problems, symmetries might solve the problem entirely or substantially reduce the effective size, see e.g.[1] Theorem 20].
3) Generalization: The bounds can be relaxed to other known bounds.

In this paper we deal with the lossy source coding problem. In [2] we presented a general approach to the one-shot coding problem. In this paper we borrow and extends ideas from [2] and provide a novel analysis of the lossy compression problem. We derive an achievable bound using random coding and a corresponding converse bound. Both bounds are given in term of a newly defined functional $\tilde{D}\left(z, Q_{Y}\right)$ where $z \in[0,1]$ and $Q_{Y}$ denote a distribution over the reproduction space $\mathcal{Y}$. The functional $\tilde{D}\left(z, Q_{Y}\right)$ is shown to be convex and has similarity to the channel coding meta-converse [3] Theorem 27].

## A. Notation

Throughout this paper, scalar random variables are denoted by capital letters (e.g. $X$ ), sample values are denoted by lower case letters (e.g. $x$ ) and their alphabets are denoted by their respective calligraphic letters, (e.g. $\mathcal{X}$ ).

The set of all distributions (probability mass functions) supported on alphabet $\mathcal{Y}$ is denoted as $\mathscr{P}(\mathcal{Y})$. The set of all conditional distributions (i.e., channels) with the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ is denoted by $\mathscr{P}(\mathcal{Y} \mid \mathcal{X})$. If $X$ has distribution $P_{X}$, we write this as $X \sim P_{X}$. The uniform probability distribution over $[0,1]$ is denoted throughout by $\mathcal{U}$. The probability (expectation) of an event (random variable) $\mathcal{A}$ under the distribution $P_{X}$ is denoted by $\mathbb{P}_{P_{X}}\{\mathcal{A}\}\left(\mathbb{E}_{P_{X}}(\mathcal{A})\right)$ respectively, e.g. $\mathbb{P}_{P_{X}}\{X \geq \alpha\}$ and $\mathbb{E}_{P_{X}}(f(X))$. In some cases, we abbreviate the notation and write $P_{X}\{\mathcal{A}\}$ instead of $\mathbb{P}_{P_{X}}\{\mathcal{A}\}$, e.g. $P_{X}(X \geq \alpha)=\mathbb{P}_{P_{X}}\{X \geq \alpha\}$. In some cases, we write $\mathbb{E}_{\mu}(f(X))$ where $\mu$ is $\sigma$-finite measure and not a probability measure.

## II. Problem setting

Let $X$ denote the random variable on $\mathcal{X}$, representing the source we want to compress. The elements of $\mathcal{X}$ are the input symbols. Denote by $P_{X}$ the distribution of $X$. Let $\mathcal{Y}$ denote the set of reproduction symbols. Let $d: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^{+}$denote the distortion function. Any subset $\mathcal{C} \subset \mathcal{Y}$ is a code and the average distortion associated with this code is:

$$
\begin{equation*}
D\left(P_{X}, \mathcal{C}\right) \triangleq \mathbb{E}_{P_{X}}\left(\min _{y \in \mathcal{C}}\{d(X, y)\}\right) \tag{1}
\end{equation*}
$$

Let:

$$
\begin{equation*}
D\left(P_{X}, R\right) \triangleq \min _{\mathcal{C} \subset \mathcal{Y}:|\mathcal{C}|=e^{R}} D\left(P_{X}, \mathcal{C}\right) \tag{2}
\end{equation*}
$$

denote the optimal distortion-rate function. The goal is to find upper and lower bounds on $D\left(P_{X}, R\right)$.
Throughout this paper we assume that both $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. Thus the distribution $P_{X}$ is discrete and the distortion is bounded by some $d_{\max }$. The results in this paper can be extended quite straight forwardly to rather general alphabets $\mathcal{X}, \mathcal{Y}$ and appropriate $\sigma$-algebras, as long as the probability distribution $P_{X}$ is well defined. The boundedness of the distortion can be relaxed to the following: There exist a "small" finite set $\mathcal{R} \subset \mathcal{Y}$ such that $\mathbb{E}_{P_{X}}\left(\min _{y \in \mathcal{R}} d(X, y)\right)=d_{\max }<\infty$.

## III. ACHIEVABILITY BOUND

For the achievable argument we use the random coding approach. Let $Q_{Y} \in \mathscr{P}(\mathcal{Y})$ denote a given distribution on $\mathcal{Y}$. A random code of rate $R$ with $M=e^{R}+1$ codewords is $C=\left\{Y_{0}, \ldots, Y_{e^{R}}\right\}$ where each $Y_{i}$ is drawn from $Q_{Y}$ independently of the other code words. The average distortion of the random code is:

$$
\begin{equation*}
D\left(P_{X}, Q_{Y}, R\right) \triangleq \mathbb{E}_{C}\left(D\left(P_{X}, C\right)\right) \tag{3}
\end{equation*}
$$

## A. pairwise correct probability

In [2] the pairwise error probability was defined as the random variable representing the probability of error given the sent and received symbols. Here, we define the pairwise correct probability, which represent the probability of drawing a reproduction symbol that is better than a given reproduction symbol.
Definition 1. For $x \in \mathcal{X}, y \in \mathcal{Y}$ and $u \in[0,1]$ Let:

$$
\begin{align*}
p_{c, x, y, u} & \triangleq Q_{Y}\{d(x, Y)<d(x, y)\} \\
& +u \cdot Q_{Y}\{d(x, Y)=d(x, y)\} \tag{4}
\end{align*}
$$

The pairwise correct decoding probability is the random variable: $p_{c, x, y, U}$ where $U \sim \mathcal{U}$ is uniform over $[0,1]$.
The following proposition summaries the properties we need about the pairwise correct decoding $p_{c, x, y, u}$ :

## Proposition 1.

1) For any $w \in[0,1]$ and $x$ there exist $y$ and $\tau$ such that:

$$
w=p_{c, x, y, \tau}
$$

2) $d\left(x, y_{1}\right)<d\left(x, y_{2}\right) \Rightarrow p_{c, x, y_{1}, U_{1}} \leq p_{c, x, y_{2}, U_{2}}$. If $Q_{Y}\left(y_{1}\right)>0$ or $Q_{Y}\left(y_{2}\right)>0$ then $p_{c, x, y_{1}, U_{1}}<p_{c, x, y_{2}, U_{2}}$ with probability 1.
3) $p_{c, x, Y, U} \sim W$ where $Y \sim Q_{Y}$ and $U, W$ are uniform over $[0,1]$.

Proof. To prove (1) note that there must exist a $y$ such that:

$$
Q_{Y}\{d(x, Y)<d(x, y)\} \leq w \leq Q_{Y}\{d(x, Y) \leq d(x, y)\}
$$

If $Q_{Y}\{d(x, Y)=d(x, y)\}=0$ then we are done with any $\tau$. If $Q_{Y}\{d(x, Y)=d(x, y)\} \neq 0$ then

$$
\tau=\frac{w-Q_{Y}\{d(x, Y)<d(x, y)\}}{Q_{Y}\{d(x, Y)=d(x, y)\}} \leq 1
$$

satisfies the requirement. To prove (2):

$$
\begin{aligned}
p_{c, x, y_{1}, U_{1}} & =Q_{Y}\left\{d(x, Y)<d\left(x, y_{1}\right)\right\} \\
& +U_{1} \cdot Q_{Y}\left\{d(x, Y)=d\left(x, y_{1}\right)\right\} \\
& \stackrel{(a)}{\leq} Q_{Y}\left\{d(x, Y) \leq d\left(x, y_{1}\right)\right\} \\
& \leq Q_{Y}\left\{d(x, Y)<d\left(x, y_{2}\right)\right\} \\
& \stackrel{(b)}{\leq} Q_{Y}\left\{d(x, Y)<d\left(x, y_{2}\right)\right\} \\
& +U_{2} \cdot Q_{Y}\left\{d(x, Y)=d\left(x, y_{2}\right)\right\} \\
& =p_{c, x, y_{2}, U_{2}}
\end{aligned}
$$

If $Q_{Y}\left(y_{1}\right)>0$ then $Q_{Y}\left\{d(x, Y)=d\left(x, y_{1}\right)\right\}>0$ thus we have strict inequality in (a). If $Q_{Y}\left(y_{2}\right)>0$ we have strict inequality in (b). To prove (3) let $y_{w}$ and $\tau_{w}$ be such that:

$$
\begin{aligned}
w & =Q_{Y}\left\{d(x, Y)<d\left(x, y_{w}\right)\right\} \\
& +\tau_{w} \cdot Q_{Y}\left\{d(x, Y)=d\left(x, y_{w}\right)\right\}
\end{aligned}
$$

Then:

$$
\begin{aligned}
& \mathbb{P}_{Q_{Y}}\left\{p_{c, x, Y, U}<w\right\} \\
& =\sum_{y: d(x, y)<d\left(x, y_{w}\right)} Q_{Y}(y) \mathbb{P}_{Q_{Y}}\left\{p_{c, x, y, U}<w\right\} \\
& +\sum_{y: d(x, y)=d\left(x, y_{w}\right)} Q_{Y}(y) \mathbb{P}_{Q_{Y}}\left\{p_{c, x, y, U}<w\right\} \\
& +\sum_{y: d(x, y)>d\left(x, y_{w}\right)} Q_{Y}(y) \mathbb{P}_{Q_{Y}}\left\{p_{c, x, y, U}<w\right\}
\end{aligned}
$$

If $d(x, y) \lessgtr d\left(x, y_{w}\right)$ and $Q_{Y}(y)>0$ then:

$$
p_{c, x, y, U} \lessgtr w=p_{c, x, y_{w}, \tau}
$$

according to (2). Thus the first sum gives $Q_{Y}\left\{d(x, Y)<d\left(x, y_{w}\right)\right\}$ and the last sum vanishes. If $d(x, y)=d\left(x, y_{w}\right)$ and $Q_{Y}(y)>0$ then:

$$
\begin{aligned}
\mathbb{P}_{Q_{Y}}\left\{p_{c, x, y, U}<w\right\} & =\mathbb{P}_{Q_{Y}}\left\{p_{c, x, y_{w}, U}<p_{c, x, y_{w}, \tau_{w}}\right\} \\
& =\mathbb{P}\left\{U<\tau_{w}\right\}=\tau_{w}
\end{aligned}
$$

hence the middle sum gives $\tau_{w} \cdot Q_{Y}\left\{d(x, Y)=d\left(x, y_{w}\right)\right\}$. Combined:

$$
\begin{aligned}
\mathbb{P}_{Q_{Y}}\left\{p_{c, x, Y, U}<w\right\} & =Q_{Y}\left\{d(x, Y)<d\left(x, y_{w}\right)\right\} \\
& +\tau_{w} \cdot Q_{Y}\left\{d(x, Y)=d\left(x, y_{w}\right)\right\} \\
& =w
\end{aligned}
$$

## B. Random coding performance

Proposition 1 suggests that for any fixed $x$ we have a correspondence between the elements of $\mathcal{Y}$ and the sub interval

$$
\left[Q_{Y}\{d(x, Y)<d(x, y)\}, Q_{Y}\{d(x, Y) \leq d(x, y)\}\right] \subset[0,1]
$$

given by $y \Longleftrightarrow p_{c, x, y, U}$. We define the distortion function $d$ on $\mathcal{X} \times[0,1]$ according to:

$$
\begin{equation*}
\tilde{d}(x, u) \triangleq d(x, y), \quad u=p_{c, x, y, \tau} \tag{5}
\end{equation*}
$$

This correspondence is well defined almost everywhere with respect to the pair $\left(Q_{Y}, U\right)$ since if $d\left(x, y_{1}\right) \neq d\left(x, y_{2}\right)$, and $Q_{Y}\left(y_{1}\right)>0$ or $Q_{Y}\left(y_{2}\right)>0$ the support of $p_{c, x, y_{1}, U_{1}}$ and $p_{c, x, y_{2}, U_{2}}$ do not overlap with probability 1 . Moreover, this mapping is order preserving, i.e.,

$$
\begin{equation*}
d\left(x, y_{1}\right)<d\left(x, y_{2}\right) \Rightarrow p_{c, x, y_{1}, U_{1}} \leq p_{c, x, y_{2}, U_{2}} \tag{6}
\end{equation*}
$$

The following result provides an exact formula for the average distortion of random code.
Theorem 2 (Exact performance of random coding). The average distortion of random code with $M=e^{R}+1$ codewords $\left\{Y_{i}\right\}$ drawn from $Q_{Y}$ is given by:

$$
\begin{equation*}
\mathbb{E}_{P_{X},\left\{Y_{i}\right\}}\left(\min d\left(X, Y_{i}\right)\right)=\int_{0}^{1} \tilde{D}\left(w, Q_{Y}\right) G_{M}^{\prime}(w) d w \tag{7}
\end{equation*}
$$

where: $\tilde{D}\left(w, Q_{Y}\right)=w^{-1} \cdot \mathbb{E}_{P_{X} \times Q_{Y}}\left(d(X, Y) \cdot \mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}\right)$ and $G_{M}(w)=-(1-w)^{M-1}((M-1) w+1)$.
Corollary 3. For any $\lambda<R$ :

$$
\begin{aligned}
& \mathbb{E}\left(\min d\left(X, Y_{i}\right)\right) \leq \tilde{D}\left(e^{-(R-\lambda)}, Q_{Y}\right) \\
& +\left(\tilde{D}\left(1, Q_{Y}\right)-\tilde{D}\left(e^{-(R-\lambda)}, Q_{Y}\right)\right) \cdot e^{-e^{\lambda}}\left(e^{\lambda}+1\right) \\
& \leq \tilde{D}\left(e^{-(R-\lambda)}, Q_{Y}\right)+d_{\max } \cdot e^{-e^{\lambda}}\left(e^{\lambda}+1\right)
\end{aligned}
$$

For $z \in\left[\tilde{D}\left(0, Q_{Y}\right), \tilde{D}\left(1, Q_{Y}\right)\right]$ let:

$$
\tilde{D}^{-1}\left(z, Q_{Y}\right) \triangleq \inf \left\{w \in[0,1]: \tilde{D}\left(w, Q_{Y}\right) \geq z\right\}
$$

and:

$$
\begin{equation*}
\tilde{R}\left(z, Q_{Y}\right) \triangleq-\log \tilde{D}^{-1}\left(z, Q_{Y}\right) \tag{8}
\end{equation*}
$$

$\tilde{R}\left(z, Q_{Y}\right)$ is the "rate distortion" function associated with the prior distribution $Q_{Y}$.
Corollary 4. For any $Q_{Y}$, let $d_{r e q} \in\left(\tilde{D}\left(0, Q_{Y}\right), \tilde{D}\left(1, Q_{Y}\right)\right)$ denote a desired distortion level. There exist a code with distortion level $d_{\text {req }}$ and rate $R$ such that:

$$
\begin{aligned}
R & \leq \min _{z<d_{r e q}} \tilde{R}\left(z, Q_{Y}\right)+f^{-1}\left(\frac{d_{r e q}-z}{\tilde{D}\left(1, Q_{Y}\right)-z}\right) \\
& \leq \min _{z<d_{r e q}} \tilde{R}\left(z, Q_{Y}\right)+g\left(\frac{\tilde{D}\left(1, Q_{Y}\right)-z}{d_{r e q}-z}\right)
\end{aligned}
$$

where $f(t)=e^{-e^{t}}\left(e^{t}+1\right)$ and $g(x)=\log \log (x)+\log \left(\frac{2}{3}\right)+\log \left(1+\sqrt{1+9(2 \log (x))^{-1}}\right)$
Proof of theorem 2. To calculate the average distortion of a random code with $M$ codewords we proceed as follow:

$$
\begin{aligned}
\mathbb{E}\left(\min d\left(X, Y_{i}\right)\right. & \stackrel{(a)}{=} \mathbb{E}_{P_{X}}\left(\mathbb{E}\left(\min d\left(X, Y_{i}\right) \mid X\right)\right) \\
& \stackrel{(b)}{=} \mathbb{E}_{P_{X}}\left(\mathbb{E}\left(\min _{i}\left\{\tilde{d}\left(X, p_{c, X, Y_{i}, U_{i}}\right) \mid X\right\}\right)\right) \\
& \stackrel{(c)}{=} \mathbb{E}_{P_{X}}\left(\mathbb{E}\left(\tilde{d}\left(X, \min _{i}\left\{p_{c, X, Y_{i}, U_{i}}\right\}\right)\right) \mid X\right) \\
& \stackrel{(d)}{=} \mathbb{E}_{P_{X}}\left(\mathbb{E}_{W_{M}}\left(\tilde{d}\left(X, W_{M}\right) \mid X\right)\right) \\
& \stackrel{(e)}{=} \mathbb{E}_{W_{M}}\left(\mathbb{E}_{P_{X}}\left(\tilde{d}\left(X, W_{M}\right) \mid W_{M}\right)\right) \\
& \stackrel{(f)}{=} \int_{0}^{1} \mathbb{E}_{P_{X}}(\tilde{d}(X, w)) f_{M}(w) d w \\
& \stackrel{(g)}{=} f_{M}(1) \tilde{D}_{1}(1)-f_{M}(0) \tilde{D}_{1}(0) \\
& -\int_{0}^{1} \tilde{D}_{1}(w) f_{M}^{\prime}(w) d w \\
& \stackrel{(h)}{=} M \cdot(M-1) \int_{0}^{1} \tilde{D}_{2}(w) w(1-w)^{M-2} d w \\
& \stackrel{(i)}{=} \int_{0}^{1} \tilde{D}_{2}(w) G_{M}^{\prime}(w) d w
\end{aligned}
$$

where $W_{M}$ is the minimum of $M$ independent uniform random variables, $\tilde{D}_{1}(w)=\int_{0}^{w} \mathbb{E}_{X}(\tilde{d}(X, z)) d z$ and $\tilde{D}_{2}(w)=$ $w^{-1} \cdot \tilde{D}_{1}(w)$.

- (a) is the law of total expectation with respect to $X$.
- (b) follow since $d(x, y)=\tilde{d}\left(x, p_{c, x, y, U}\right)$ according to (5).
- (c) follow since $p_{c, x, y, U}$ preserve the order induce by $\tilde{d}$ according to (6).
- (d) follow since $p_{c, x, Y_{i}, U_{i}}$ are all independent and uniform over $[0,1]$.
- (e) is again the law of total expectation with respect to $X$ and $W_{M}$.
- (f) is the expectation according to the p.d.f $f_{M} 1 \mathbf{1}$. (see 18 in the appendix).
- (g) is integration by parts.
- (h) follow since $\tilde{D}_{1}(0)=f(1)=0$ and:

$$
f_{M}^{\prime}(w)=-M \cdot(M-1)(1-w)^{M-2}
$$

- (i) follow since $G_{M}^{\prime}(w)=M(M-1) w(1-w)^{M-2}$.
${ }^{1}$ The p.d.f of $W_{M}$ is $f_{M}(w)=M \cdot(1-u)^{M-1}$

Finlay:

$$
\begin{aligned}
\tilde{D}_{1}(w) & =\int_{0}^{w} \mathbb{E}_{P_{X}}(\tilde{d}(X, z)) d z \\
& \stackrel{(a)}{=} \mathbb{E}_{P_{X}, W}\left(\tilde{d}(X, W) \cdot \mathbb{1}_{\{W \leq w\}}\right) \\
& \stackrel{(b)}{=} \mathbb{E}_{P_{X} \times Q_{Y}}\left(\tilde{d}\left(X, p_{c, X, Y, U}\right) \cdot \mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}\right) \\
& \stackrel{(c)}{=} \mathbb{E}_{P_{X} \times Q_{Y}}\left(d(X, Y) \cdot \mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}\right) \\
& =w \cdot \tilde{D}\left(w, Q_{Y}\right)
\end{aligned}
$$

where (a) follow since $W$ is uniform over $[0,1]$. (b) follow since for each $x, p_{c, x, Y, U}$ is uniform over $[0,1]$ and (c) is (5). Proof of corollary 3. $\tilde{D}\left(w, Q_{Y}\right)$ is an increasing function, Thus:

$$
\begin{aligned}
\mathbb{E}\left(\min d\left(X, Y_{i}\right)\right) & =\int_{0}^{1} \tilde{D}\left(w, Q_{Y}\right) G_{M}^{\prime}(w) d w \\
& \leq \tilde{D}\left(u, Q_{Y}\right) \int_{0}^{u} G_{M}^{\prime}(w) d w \\
& +\tilde{D}\left(1, Q_{Y}\right) \int_{u}^{1} G_{M}^{\prime}(w) d w \\
& =\tilde{D}\left(u, Q_{Y}\right) \\
& -\left(\tilde{D}\left(1, Q_{Y}\right)-\tilde{D}\left(u, Q_{Y}\right)\right) G_{M}(u)
\end{aligned}
$$

since $G_{M}(0)=-1, G_{M}(1)=0$ and:

$$
\begin{aligned}
-G_{M}(u) & =(1-u)^{M-1}((M-1) u+1) \\
& \leq e^{-u \cdot e^{R}}\left(u \cdot e^{R}+1\right)=e^{-e^{\lambda}}\left(e^{\lambda}+1\right)
\end{aligned}
$$

where $u \cdot e^{R}=e^{\lambda}$. The second bound follow since $\tilde{D}\left(1, Q_{Y}\right) \leq d_{\max }$.
Proof of corollary 4. Using $d_{\text {req }}=\mathbb{E}\left(\min d\left(X, Y_{i}\right)\right)$ and $z=\tilde{D}\left(e^{-(R-\lambda)}, Q_{Y}\right)$ in corollary 3 we have:

$$
d_{r e q} \leq z+\left(\tilde{D}\left(1, Q_{Y}\right)-z\right) \cdot f(\lambda)
$$

Hence:

$$
\lambda \leq f^{-1}\left(\frac{d_{r e q}-z}{\tilde{D}\left(1, Q_{Y}\right)-z}\right)
$$

since $f$ is decreasing. The result follow from:

$$
\tilde{R}\left(z, Q_{Y}\right)=R-\lambda
$$

The second bound follow from the technical lemma 1 given in the appendix.

The standard rate distortion analysis usually employ a "test" channel $W_{Y \mid X}$ that is used for change of measure during the achievability proof and also serves as the encoding function when the converse is proved. The following Proposition suggest such a channel to be used in our achievability theorem.
Proposition 5. Let $W_{Y \mid X=x}^{w} \triangleq w^{-1} \cdot Q_{Y} \cdot \mathbb{1}_{\left\{p_{c, x, Y, U} \leq w\right\}}$, i.e.

$$
W_{Y \mid X=x}^{w}(y)=w^{-1} \cdot Q_{Y}(y) \cdot \mathbb{P}_{U}\left\{p_{c, x, y, U} \leq w\right\}
$$

Then $W_{Y \mid X=x}^{w}$ is a probability distribution over $\mathcal{Y}$ and:

$$
\tilde{D}\left(w, Q_{Y}\right)=\mathbb{E}_{P_{X} \times W_{Y \mid X}^{w}}(d(X, Y))
$$

Proof. $W_{Y \mid X=x}^{w}$ is a distribution since:

$$
\begin{aligned}
w & =\mathbb{P}_{Q_{Y}, U}\left\{p_{c, x, Y, U} \leq w\right\} \\
& =\mathbb{E}_{Q_{Y}, U}\left(\mathbb{1}_{\left\{p_{c, x, Y, U} \leq w\right\}}\right)
\end{aligned}
$$

and:

$$
\begin{aligned}
\tilde{D}\left(w, Q_{Y}\right) & =\mathbb{E}_{P_{X} \times Q_{Y}}\left(d(X, Y) \cdot w^{-1} \cdot \mathbb{1}_{\left\{p_{c, x, Y, U} \leq w\right\}}\right) \\
& =\mathbb{E}_{P_{X} \times W_{Y \mid X}^{w}}(d(X, Y))
\end{aligned}
$$

## IV. Converse bound

The channel $W_{Y \mid X}^{w}$ that was suggested in proposition 5 with the proper distribution $Q_{Y}$ and parameter $w=M^{-1}$ can serve as the encoding function as the next proposition show.
Proposition 6. Let $\mathcal{C} \subset \mathcal{Y}$ be a code with $M$ codewords. Let $\tilde{W}_{Y \mid X}$ denote the optimal encoding of the $\mathcal{X}$ to $\mathcal{C}$, i.e.

$$
\tilde{W}_{Y \mid X}=\underset{c \in \mathcal{C}}{\arg \min }(d(X, c))
$$

where ties are broken arbitrary. Let $Q_{Y}^{\mathcal{C}}$ distribute uniformly over the codewords, i.e. $Q_{Y}^{\mathcal{C}}(y)=M^{-1}$ if $y \in \mathcal{C}$ and $Q_{Y}^{\mathcal{C}}(y)=0$ otherwise. Then:

$$
\tilde{W}_{Y \mid X=x}=W_{Y \mid X=x}^{M^{-1}}
$$

Proof. Recall that

$$
p_{c, x, y, U}=Q_{Y}^{\mathcal{C}}(d(x, Y)<d(x, y))+U \cdot Q_{Y}^{\mathcal{C}}(d(x, Y)=d(x, y))
$$

Case I: $Q_{Y}^{\mathcal{C}}(d(x, Y)<d(x, y))>0$ : In this case there exist $y^{\prime} \neq y$ such that $d\left(x, y^{\prime}\right)<d(x, y)$. Thus: $\tilde{W}_{Y \mid X=x}(y \mid x)=0$ since $x$ cannot encode to $y$ ( $y^{\prime}$ produce lower distortion). From $Q_{Y}^{\mathcal{C}}(d(x, Y)<d(x, y)) \geq M^{-1}$ it follow that $p_{c, x, y, U}>M^{-1}$ with probability 1 and $W_{Y \mid X=x}^{M^{-1}}(y \mid x)=0$ as well.

Case II: $Q_{Y}^{\mathcal{C}}(d(x, Y)<d(x, y))=0$ Let:

$$
Q_{Y}^{\mathcal{C}}(d(x, Y)=d(x, y))=k \cdot M^{-1}
$$

where $k$ is integer greater than 0 . There are $k$ different symbols $y_{1}=y, \ldots, y_{k} \in \mathcal{Y}$ such that $d\left(x, y_{i}\right)=d(x, y)$ and $\tilde{W}_{Y \mid X=x}(\cdot \mid x)$ encode $x$ to one of these symbols randomly. Thus:

$$
\tilde{W}_{Y \mid X=x}\left(y_{1} \mid x\right)=\cdots=\tilde{W}_{Y \mid X=x}\left(y_{k} \mid x\right)=k^{-1}
$$

in particular, $\tilde{W}_{Y \mid X=x}(y \mid x)=k^{-1}$. Since:

$$
p_{c, x, y, U}=U \cdot Q_{Y}^{\mathcal{C}}(d(x, Y)=d(x, y))=U \cdot k \cdot M^{-1}
$$

we have:

$$
\begin{aligned}
\mathbb{P}_{U}\left\{p_{c, x, y, U} \leq M^{-1}\right\} & =\mathbb{P}_{U}\left\{U \cdot k \cdot M^{-1} \leq M^{-1}\right\} \\
& =k^{-1}
\end{aligned}
$$

thus:

$$
W_{Y \mid X=x}^{M^{-1}}(y)=M \cdot Q_{Y}^{\mathcal{C}}(y) \cdot \mathbb{P}_{U}\left\{p_{c, x, y, U} \leq M^{-1}\right\}=k^{-1}
$$

as required.
Theorem 7. For any code $\mathcal{C}$ :

$$
D\left(P_{X}, \mathcal{C}\right)=\tilde{D}\left(M^{-1}, Q_{Y}^{\mathcal{C}}\right)
$$

in particular:

$$
D\left(P_{X}, R\right) \geq \inf _{Q_{Y} \in \mathscr{P}(\mathcal{Y})} \tilde{D}\left(e^{-R}, Q_{Y}\right)
$$

Proof.

$$
\begin{aligned}
D\left(P_{X}, \mathcal{C}\right) & =\mathbb{E}_{P_{X} \times \tilde{W}_{Y \mid X}}(d(X, Y)) \\
& =\mathbb{E}_{P_{X} \times W_{Y \mid X}^{M-1}}(d(X, Y)) \\
& =\tilde{D}\left(M^{-1}, Q_{Y}^{\mathcal{C}}\right)
\end{aligned}
$$

Let:

$$
\hat{D}(z) \triangleq \inf _{Q_{Y}} \tilde{D}\left(e^{-z}, Q_{Y}\right)
$$

Writing the achievable and converse results in terms of $\hat{D}(z)$, we have:

$$
\begin{align*}
\hat{D}(R) & \leq D\left(P_{X}, R\right)  \tag{9}\\
& \leq \inf _{\lambda<R}\left\{\hat{D}(R-\lambda)+d_{\max } \cdot e^{-e^{\lambda}}\left(e^{\lambda}+1\right)\right\} \tag{10}
\end{align*}
$$

This equation exemplify the tightness of our bounds.

## V. VARIATIONAL FORMS of $\tilde{D}\left(z, Q_{Y}\right)$

In this section we present two alternative presentations of $\tilde{D}\left(z, Q_{Y}\right)$. We first recall some definition. The definition of $\beta_{\alpha}(P, Q)$ which represent the optimal performance of a binary hypothesis testing between two $\sigma$-finite measures $P$ and $Q$ over a set $W$ :

$$
\begin{equation*}
\beta_{\alpha}(P, Q)=\min _{\substack{P_{Z \mid W}: \\ \sum_{w \in W} P(w) P_{Z \mid W}(1 \mid w) \geq \alpha}} \sum_{w \in W} Q(w) P_{Z \mid W}(1 \mid w) \tag{11}
\end{equation*}
$$

$P_{Z \mid W}: W \rightarrow\{0,1\}$ is any randomized test between $P$ and $Q$. The minimum is guaranteed to be achieved by the NeymanPearson lemma. The functional $\beta_{\alpha}(P, Q)$ has been proved useful for converse results in channel coding and lossy compression, e.g. [3, Theorem 26], [4, Theorem 8].

The $\infty$-order divergences between $P$ and $Q$ is:

$$
\begin{equation*}
D_{\infty}(P \| Q) \triangleq \log \inf \{\lambda: P(x) \leq \lambda Q(x), \forall x\} \tag{12}
\end{equation*}
$$

Theorem 8 (Variational forms of $\tilde{D}\left(z, Q_{Y}\right)$ ). Let $w=e^{-R}$. Then:

$$
\begin{align*}
& \tilde{D}\left(z, Q_{Y}\right) \\
& =\sup _{Q_{X}} \beta_{w}\left(Q_{X} \times Q_{Y}, P_{X} \times Q_{Y} \times d(X, Y)\right)  \tag{13}\\
& =\inf _{W_{Y \mid X}: D_{\infty}\left(P_{X} \times W_{Y \mid X} \| P_{X} \times Q_{Y}\right) \leq R} \mathbb{E}_{P_{X} \times W_{Y \mid X}}(d(X, Y)) \tag{14}
\end{align*}
$$

Corollary 9. The convexity of $\tilde{D}\left(z, Q_{Y}\right)$ with respect to $Q_{Y}$ readily follows from (13). Since $\beta_{w}\left(Q_{X} \times Q_{Y}, P_{X} \times Q_{Y} \times d(X, Y)\right)$ is convex with respect to $Q_{Y}$ according to [1] Theorem 6] and supremum of convex function is convex as well.

Proof. To prove (13), let $\mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}$ denote a (not necessarily optimal) test between $Q_{X} \times Q_{Y}$ and $P_{X} \times Q_{Y} \times d(X, Y)$. Since $\mathbb{P}_{Q_{Y}}\left\{p_{c, x, Y, U} \leq w\right\}=w$ for any $x$, it follow that for any $Q_{X}$ we have: $\mathbb{P}_{Q_{X} \times Q_{Y}}\left\{p_{c, X, Y, U} \leq w\right\}=w$ :

$$
\begin{aligned}
& \beta_{w}\left(Q_{X} \times Q_{Y}, P_{X} \times Q_{Y} \times d(X, Y)\right) \\
& \leq \mathbb{E}_{P_{X} \times Q_{Y} \times d(X, Y)}\left(\mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}\right) \\
& =\mathbb{E}_{P_{X} \times Q_{Y}}\left(d(X, Y) \cdot \mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}\right)
\end{aligned}
$$

Thus:

$$
\begin{aligned}
& \sup _{Q_{X}} \beta_{u}\left(Q_{X} \times Q_{Y}, P_{X} \times Q_{Y} \times d(X, Y)\right) \\
& \leq \mathbb{E}_{P_{X} \times Q_{Y}}\left(d(X, Y) \cdot \mathbb{1}_{\left\{p_{c, X, Y, U} \leq u\right\}}\right)
\end{aligned}
$$

To show the reverse inequality, we show that there exist $Q_{X}^{c}$ such that:

$$
\begin{aligned}
& \beta_{w}\left(Q_{X}^{c} \times Q_{Y}, P_{X} \times Q_{Y} \times d(X, Y)\right) \\
& =\mathbb{E}_{P_{X} \times Q_{Y}}\left(d(X, Y) \cdot \mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}\right)
\end{aligned}
$$

which will complete the proof. We will construct such a $Q_{X}^{c}$ and show that the optimal likelihood test between $Q_{X}^{c} \times Q_{Y}$ and $P_{X} \times Q_{Y} \times d(X, Y)$ matches the test $\mathbb{1}_{\left\{p_{c, X, Y, U} \leq w\right\}}$. The likelihood ratio is:

$$
L(x, y)=\frac{P_{X}(x) \times Q_{Y}(y) \cdot d(x, y)}{Q_{X}^{c}(x) \times Q_{Y}(y)}=\frac{P_{X}(x)}{Q_{X}^{c}(x)} \times d(x, y)
$$

and the likelihood ratio test is:

$$
P(z \mid x, y)=\mathbb{1}_{\{L(x, y)<\lambda\}}+\sum_{x^{\prime}} \tau_{x^{\prime}} \mathbb{1}_{\left\{L(x, y)=\lambda, x=x^{\prime}\right\}}
$$

where $\lambda$ and $\tau_{x^{\prime}}$ are tuned so that:

$$
\mathbb{P}_{Q_{X}^{c} \times Q_{Y}}\{P(z \mid X, Y)=1\}=w
$$

For any $x$ there exist $y_{x}$ and $\tau_{x}$ such that:

$$
w=Q_{Y}\left(d(x, Y)<d\left(x, y_{x}\right)\right)+\tau_{x} \cdot Q_{Y}\left(d(x, Y)=d\left(x, y_{x}\right)\right)
$$

Define $\lambda$ and $Q_{X}^{c}$ :

$$
\begin{aligned}
\lambda & =\sum_{x} P_{X}(x) \cdot d\left(x, y_{x}\right) \\
Q_{X}^{c}(x) & =\lambda^{-1} \cdot P_{X}(x) \cdot d\left(x, y_{x}\right)
\end{aligned}
$$

Note that $Q_{X}^{c}$ is probability distribution and:

$$
L\left(x, y_{x}\right)=\frac{P_{X}(x)}{Q_{X}^{c}(x)} \cdot d\left(x, y_{x}\right)=\lambda
$$

To show that the tests matches we have to prove:

$$
\mathbb{P}\left\{p_{c, x, y, U} \leq w\right\}=\mathbb{P}\{P(z \mid x, y)=1\}
$$

for any $x$ and $y$. There are 3 cases to consider:

1) $d(x, y)<d\left(x, y_{x}\right)$ : In this case:

$$
\begin{aligned}
p_{c, x, y, U} & \leq Q_{Y}(d(x, Y) \leq d(x, y)) \\
& \leq Q_{Y}\left(d(x, Y)<d\left(x, y_{x}\right)\right) \\
& \leq w
\end{aligned}
$$

thus: $\mathbb{P}\left\{p_{c, x, y, U} \leq w\right\}=1$. On the other hand, since:

$$
\begin{aligned}
L(x, y) & =\frac{P_{X}(x)}{Q_{X}^{c}(x)} \cdot d(x, y) \\
& <\frac{P_{X}(x)}{Q_{X}^{c}(x)} \cdot d\left(x, y_{x}\right) \\
& =\lambda
\end{aligned}
$$

we also have: $\mathbb{P}\{P(z \mid x, y)=1\}=1$.
2) $d(x, y)>d\left(x, y_{x}\right)$ : In this case:

$$
\begin{aligned}
p_{c, x, y, U} & \geq Q_{Y}(d(x, Y)<d(x, y)) \\
& \geq Q_{Y}\left(d(x, Y) \leq d\left(x, y_{x}\right)\right) \\
& \geq w
\end{aligned}
$$

thus: $\mathbb{P}\left\{p_{c, x, y, U} \leq w\right\}=0$. On the other hand, since

$$
\begin{aligned}
L(x, y) & =\frac{P_{X}(x)}{Q_{X}^{c}(x)} \cdot d(x, y) \\
& >\frac{P_{X}(x)}{Q_{X}^{c}(x)} \cdot d\left(x, y_{x}\right) \\
& =\lambda
\end{aligned}
$$

we also have: $\mathbb{P}\{P(z \mid x, y)=1\}=0$.
3) $d(x, y)=d\left(x, y_{x}\right)$ : In this case

$$
\begin{aligned}
\mathbb{P}\left\{p_{c, x, y, U} \leq w\right\} & =\mathbb{P}\left\{p_{c, x, y_{x}, U} \leq w\right\} \\
& =\tau_{x}
\end{aligned}
$$

On the other hand, since

$$
\begin{aligned}
L(x, y) & =\frac{P_{X}(x)}{Q_{X}^{c}(x)} \cdot d(x, y) \\
& =\frac{P_{X}(x)}{Q_{X}^{c}(x)} \cdot d\left(x, y_{x}\right) \\
& =\lambda
\end{aligned}
$$

we also have: $\mathbb{P}\{P(z \mid x, y)=1\}=\tau_{x}$.
Thus, the test $\mathbb{1}_{\left\{p_{c, x, y, U} \leq w\right\}}$ matches the likelihood ratio test and is, in fact, optimal.
To prove (14), recall the channel:

$$
W_{Y \mid X=x}^{e^{-R}}=e^{R} \cdot Q_{Y} \cdot \mathbb{1}_{\left\{p_{c, x, Y, U} \leq e^{-R}\right\}}
$$

For any $y$ such that $Q_{Y}(y)>0$ :

$$
\log \frac{e^{R} \cdot Q_{Y}(y) \cdot \mathbb{P}\left\{p_{c, x, Y, U} \leq e^{-R}\right\}}{Q_{Y}(y)} \leq R
$$

Thus: $D_{\infty}\left(P_{X} \times W_{Y \mid X}^{e^{-R}} \| P_{X} \times Q_{Y}\right) \leq R$ and:

$$
\begin{aligned}
& \tilde{D}\left(e^{-R}, Q_{Y}\right) \\
& \geq \min _{W_{Y \mid X}: D_{\infty}\left(P_{X} \times W_{Y \mid X} \| P_{X} \times Q_{Y}\right) \leq R} \mathbb{E}_{P_{X} \times W_{Y \mid X}}(d(X, Y))
\end{aligned}
$$

On the other hand, if $W_{Y \mid X}$ satisfy:

$$
D_{\infty}\left(P_{X} \times W_{Y \mid X} \| P_{X} \times Q_{Y}\right) \leq R
$$

we have $W_{Y \mid X}(y \mid x) P_{X}(x) \leq e^{R} \cdot Q_{Y}(y) P_{X}(x)$ for each $x$ and $y$. Thus, to get the minimal $\mathbb{E}_{P_{X} \times W_{Y \mid X}}(d(X, Y))$ we will have to assign the maximal probability to the minimal distortion, i.e. $Q_{Y}(y) \cdot e^{R}$. This is exactly what the assignment $Q_{Y} \cdot \mathbb{1}_{\left\{p_{c, X, Y, U} \leq e^{-R}\right\}} \cdot e^{R}$ does which assign $Q_{Y}(y) \cdot e^{R}$ for the smallest distortion values until it exhaust the probability to one.

Let $W_{Y \mid X}$ denote any channel and let $Q_{Y}$ denote the marginal distribution of $W_{Y \mid X} \times P_{X}$. Let $i(x ; y)=\frac{W_{Y \mid X}(y \mid x)}{Q_{Y}(y)}$ denote the information density. Note that in theorem 8 we did not require that $Q_{Y}$ is the marginal distribution. We might have relaxed the requirement $D_{\infty}\left(P_{X} \times W_{Y \mid X} \| P_{X} \times Q_{Y}\right) \leq R$ which amounts to $i(x ; y) \leq R$ for each $x$ and $y$ when $Q_{Y}$ is the marginal distribution to the requirement that $\mathbb{P}_{P_{X} \times W_{Y \mid X}}\{i(X, Y) \leq R-\delta\}$ is close to 1 .
Theorem 10. Let $W_{Y \mid X}$ such that:

$$
\mathbb{P}_{P_{X} \times W_{Y \mid X}}\{i(X, Y) \leq R-\delta\}=e^{-\lambda}
$$

Then:

$$
\tilde{D}\left(e^{-(R-\delta)-\lambda}, Q_{Y}\right) \leq \mathbb{E}_{P_{X} \times W_{Y \mid X}}(d(X, Y)) \cdot e^{\lambda}
$$

Proof.

$$
\begin{aligned}
\mathbb{E}_{P_{X} \times W_{Y \mid X}}(d(X, Y)) & \geq \mathbb{E}_{P_{X} \times W_{Y \mid X}}\left(d(X, Y) \mathbb{1}_{\{i(x, y) \leq R-\delta\}}\right) \\
& \geq \mathbb{E}_{P_{X} \times W_{Y \mid X}^{\prime}}(d(X, Y)) e^{-\lambda} \\
& \stackrel{(a)}{\geq} \tilde{D}\left(e^{-(R-\delta)-\lambda}, Q_{Y}\right) \cdot e^{-\lambda}
\end{aligned}
$$

where $W_{Y \mid X}^{\prime}=e^{\lambda} \cdot W_{Y \mid X} \mathbb{1}_{\{i(x, y) \leq R-\delta\}}$ is a probability distribution. Note that:

$$
\frac{W_{Y \mid X}^{\prime}(y \mid x)}{Q_{Y}(y)}=\frac{e^{\lambda} \cdot W_{Y \mid X}(y \mid x) \mathbb{1}_{\{i(x, y) \leq R-\delta\}}}{Q_{Y}(y)} \leq e^{R-\delta} \cdot e^{\lambda}
$$

Thus $D_{\infty}\left(P_{X} \times W_{Y \mid X}^{\prime} \| P_{X} \times Q_{Y}\right) \leq R-\delta+\lambda$ and (a) follow from (14).

## VI. EXCESS DISTORTION

The excess distortion is a spacial case of the average distortion that we have analyzed. Let $d_{t h}$ denote the target distortion level, replacing $d(x, y)$ with $\mathbb{1}_{\left\{d(x, y)>d_{t h}\right\}}$. Let:

$$
\begin{equation*}
\tilde{D}\left(R, d_{t h}, Q_{Y}\right)=\mathbb{E}_{P_{X} \times Q_{Y}}\left(\mathbb{1}_{\left\{d(X, Y)>d_{t h}\right\}} \mathbb{1}_{\left\{p_{c, X, Y, U}<e^{-R}\right\}}\right) \tag{15}
\end{equation*}
$$

Note that $p_{c, x, y, u}$ is also defined with respect to the "new" distortion:

$$
\begin{aligned}
p_{c, x, y, u} & =Q_{Y}\left(\mathbb{1}_{\left\{d(x, Y)>d_{t h}\right\}}<\mathbb{1}_{\left\{d(x, y)>d_{t h}\right\}}\right) \\
& +u \cdot Q_{Y}\left(\mathbb{1}_{\left\{d(x, Y)>d_{t h}\right\}}<\mathbb{1}_{\left\{d(x, y)>d_{t h}\right\}}\right)
\end{aligned}
$$

equation (14) translates in this case to:

$$
\begin{aligned}
& \tilde{D}\left(R, d_{t h}, Q_{Y}\right) \\
& =\inf _{D_{\infty}\left(P_{X} \times W_{Y \mid X} \| P_{X} \times Q_{Y}\right) \leq R} \mathbb{P}_{P_{X} \times W_{Y \mid X}}\left\{d(X, Y)>d_{t h}\right\}
\end{aligned}
$$

and (8) is:

$$
\begin{aligned}
& \tilde{R}\left(\delta, z, Q_{Y}\right) \\
& =\inf _{\mathbb{P}_{P_{X} \times W_{Y \mid X}}\{d(X, Y)>z\} \leq \delta}^{=} D_{\infty}\left(P_{X} \times W_{Y \mid X} \| P_{X} \times Q_{Y}\right)
\end{aligned}
$$

## VII. Relation to known bounds

The information spectrum approach [5] Theorem 5.5.1] provides a general formula for the rate distortion function. The general formula takes any channel $W_{Y \mid X}$ and with slight abuse of notation, say that the distortion $\mathbb{E}_{P_{X} \times W_{Y \mid X}}(d(X, Y))$ is achievable for a code with rate $R$ such that $\mathbb{P}_{P_{X} \times W_{Y \mid X}}\{i(x ; y) \leq R\}$ approach 1 where $i(x ; y)$ is the information density. The achievability is proved by the random coding argument where the distribution used to draw the codewords is the marginal distribution of $P_{X} \times W_{Y \mid X}$ on $\mathcal{Y}$. In this paper (see also [6]) we started with any distribution $Q_{Y}$ and analyzed the random code performance with no channel in mind. For the achievable part, theorem 10 provides the link between the functional $\tilde{D}\left(z, Q_{Y}\right)$ and the elements in the information spectrum formula. The converse follow since a code $\mathcal{C}$ with rate $R$ satisfies

$$
D_{\infty}\left(P_{X} \times \tilde{W}_{Y \mid X} \| P_{X} \times Q_{Y}^{\mathcal{C}}\right)=R
$$

where $\tilde{W}_{Y \mid X}$ and $Q_{Y}^{\mathcal{C}}$ were defined in the text.
Much attention has been given to the problem of lossy compression with the excess distortion constraint. The tightest results (to the best of our knowledge) appeared in [6]. The converse bound [6. Theorem 2] was shown to be tighter than [7] Theorem 8]. In [8] the author demonstrated how the bound [7] Theorem 8] can be relaxed to all other bounds presented there.

The approach in [6] for the achievability was to analyze the random coding for a given prior distribution $Q_{Y}$ and bound the performance from above using a change of measure from $P_{X} \times Q_{Y}$ to $P_{X} \times W_{Y \mid X}$ where the marginal distribution of $P_{X} \times W_{Y \mid X}$ with respect to $\mathcal{Y}$ does not necessarily matches $Q_{Y}$. Later they optimize for a given channel $W_{Y \mid X}$ the best prior distribution. Thus, their bounds are given as an optimization over the a set of channels. In [6, Theorem 2] the author defined:

$$
M\left(P_{X Y}\right) \triangleq \sum_{y \in \mathcal{Y}} \sup _{x \in \mathcal{X}: P_{X Y}(x, y)>0} P_{Y \mid X}(y \mid x)
$$

and [6, Lemma 4] reads:

$$
\inf _{Q_{Y} \in \mathscr{P}(\mathcal{Y})} D_{\infty}\left(P_{X Y} \| P_{X} \times Q_{Y}\right)=\log M\left(P_{X Y}\right)
$$

Hence:

$$
\begin{aligned}
& \inf _{Q_{Y} \in \mathscr{P}(\mathcal{Y})} \tilde{R}\left(\delta, d_{t h}, Q_{Y}\right) \\
& =\inf _{\mathbb{P}_{P_{X} \times W_{Y \mid X}}\left\{d(X, X: Y)>d_{t h}\right\} \leq \delta} \quad \log M\left(P_{X} \times W_{Y \mid X}\right)
\end{aligned}
$$

and our converse bound matches theirs. Their achievability bound (Theorem 2 ) is slightly tighter than the bound in corollary 4 While their bound reads:

$$
R \leq \min _{\delta<\epsilon} \tilde{R}\left(\delta, d_{t h}\right)+\log \log \left(\frac{1-\delta}{\epsilon-\delta}\right)
$$

our bound is:

$$
R \leq \min _{\delta<\epsilon} \tilde{R}\left(\delta, d_{t h}\right)+g\left(\frac{1-\delta}{\epsilon-\delta}\right)
$$

where:

$$
g(x)=\log \log (x)+\log \left(\frac{2}{3}\right)+\log \left(1+\sqrt{1+9(2 \log (x))^{-1}}\right) .
$$

The difference smaller than 1 nat for all practical cases. Note that, our bound follow from bounding

$$
\tilde{D}\left(z, Q_{Y}\right) \leq \tilde{D}\left(u, Q_{Y}\right) \mathbb{1}_{\{z \leq u\}}+\tilde{D}\left(1, Q_{Y}\right) \mathbb{1}_{\{z>u\}}
$$

and evaluating the exact formula with this bound. Using the exact formula directly or bound $\tilde{D}\left(z, Q_{Y}\right)$ using more points would gives tighter bound. Since their achievability also follow from the random coding, obviously the exact formula is tighter.

## ViII. Prior optimization

The prior optimization problem is the main drawback of our approach. While the optimization problem is convex, we do not have a close form solution to the important case of memoryless source (and channel in the channel coding case) and this should be further investigated. In the channel coding case, the prior optimization problem for memoryless channel is "solved" by [9], Theorem 10] which shows that for memoryless channel, we can resort to memoryless priors. For our functional $\tilde{D}\left(z, Q_{Y}\right)$ (and the one used in the channel coding problem) we do not have such a theorem. Moreover, simulation results show that this is not even true, and sometimes, prior with memory are better than the memoryless priors.

## IX. Concluding remarks

In this paper we presented a novel analysis for the lossy compression problem. We have analyzed the general case of average distortion constraints. We presented tight achievable and converse bounds. Both bounds are given in terms of the functional $\tilde{D}\left(z, Q_{Y}\right)$ which has resemblance to the meta-converse in channel coding. The problem of finding distribution minimizing $\tilde{D}\left(z, Q_{Y}\right)$ is still open in general although the $\tilde{D}\left(z, Q_{Y}\right)$ is convex with respect to $Q_{Y}$.

## Appendix

Lemma 1. For any $x \in(0,1)$ let $\lambda$ be the solution to $x=e^{-e^{\lambda}}\left(e^{\lambda}+1\right)$, then:

$$
\begin{align*}
& \lambda-\log (-\log (x)) \leq \log \left(\frac{2}{3}\right)+\log \left(1+\sqrt{1-\frac{9}{2 \log (x)}}\right)  \tag{16}\\
& \lambda-\log (-\log (x)) \geq-\log (2)+\log \left(1+\sqrt{1-\frac{8}{\log (x)}}\right) \tag{17}
\end{align*}
$$

Proof. Let $t=e^{\lambda}$ and $z=-\log (x)$. Then:

$$
z=t-\log (1+t)
$$

Using: $\log (1+t) \leq \frac{t(6+t)}{2(3+2 t)}([10$, Eq. 22]) we have:

$$
\begin{aligned}
z & \geq t-\frac{t(6+t)}{2(3+2 t)} \\
& =\frac{t^{2}}{2+\frac{4}{3} t}
\end{aligned}
$$

For $t, z \geq 0$ the only solution to: $z=\frac{t^{2}}{2+\frac{4}{3} t}$ is: $t=\frac{2}{3} z\left(1+\sqrt{1+\frac{9}{2 z}}\right)$. Thus:

$$
\begin{aligned}
& z \geq \frac{3 t^{2}}{2(3+2 t)} ; t, z \geq 0 \Longleftrightarrow \\
& t \leq \frac{2}{3} z\left(1+\sqrt{1+\frac{9}{2 z}}\right)
\end{aligned}
$$

Hence:

$$
\begin{aligned}
\lambda & =\log (t) \\
& \leq \log \left(\frac{2}{3} z\left(1+\sqrt{1+\frac{9}{2 z}}\right)\right) \\
& =\log (-\log (x))+\log \left(\frac{2}{3}\right)+\log \left(1+\sqrt{1-\frac{9}{2 \log (x)}}\right)
\end{aligned}
$$

The lower bound follow along the same line using $\log (1+t) \geq \frac{2 t}{2+t}$.
Proposition 11. Let $U_{i}$ denote $M$ independent uniform random variables and let $W_{M}=\min _{i}\left\{U_{i}\right\}$. The c.d.f of $W_{M}$ is:

$$
\begin{equation*}
f_{M}(w)=M \cdot(1-u)^{M-1} \tag{18}
\end{equation*}
$$

Proof.

$$
\begin{aligned}
\mathbb{P}\left\{W_{M} \leq w\right\} & =1-\mathbb{P}\left\{\min _{i}\left\{U_{i}\right\}>w\right\} \\
& =1-\prod_{i=1}^{M} \mathbb{P}\left\{U_{i}>w\right\} \\
& =1-(1-w)^{M}
\end{aligned}
$$

Thus, the p.d.f. of $W_{M}$ is $f_{M}(w)=M \cdot(1-u)^{M-1}$.

## REFERENCES

[1] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," Information Theory, IEEE Transactions on, vol. 59, no. 5, pp. 2576-2595, 2013.
[2] N. Elkayam and M. Feder, "Achievable and converse bounds over a general channel and general decoding metric," arXiv preprint arXiv:1411.0319 2014. [Online]. Available: http://www.eng.tau.ac.il/~elkayam/FiniteBlockLen.pdf
[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," Information Theory, IEEE Transactions on, vol. 56, no. 5, pp. 2307-2359, 2010.
[4] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 2545-2575, 2013.
[5] H. Koga et al., Information-spectrum methods in information theory. Springer, 2002, vol. 50.
[6] T. Matsuta and T. Uyematsu, "Non-asymptotic bounds for fixed-length lossy compression," in 2015 IEEE International Symposium on Information Theory (ISIT). IEEE, 2015, pp. 1811-1815.
[7] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," Information Theory, IEEE Transactions on, vol. 58, no. 6, pp. 3309-3338, 2012.
[8] L. Palzer and R. Timo, "A converse for lossy source coding in the finite blocklength regime," in 24th International Zurich Seminar on Communications (IZS). ETH-Zürich, 2016
[9] S. Verdú and T. S. Han, "A general formula for channel capacity", IEEE Transactions on Information Theory, vol. 40, no. 4, pp. 1147-1157, 1994.
[10] F. Topsok, "Some bounds for the logarithmic function," Inequality theory and applications, vol. 4, p. 137.

