Fixed-Length Lossy Compression with Distortion Risk Measure Constraints

Malcolm Egan

Fixed-Length Lossy Compression with Distortion Risk Measure Constraints

Malcolm Egan

Abstract

Storage or communication of data often requires lossy compression. As this data may be later utilized for decision making, compression introduces a source of uncertainty. In order to verify the quality of decisions, it is often desirable to quantify this uncertainty. This is often achieved via families of risk measures (such as distortion risk measures), rather than the mean. In this paper, we study fixed-length lossy compression subject to distortion risk measure constraints. We first establish conditions for the existence of a fixed-length lossy source code, which satisfies a distortion risk measure constraint. We then investigate the impact of quantifying uncertainty via a distortion risk measure, rather than standard expected distortion constraints. Finally, we quantify the impact of changing the source distribution on performance, measured in terms of a distortion risk measure.

Index Terms

Lossy compression, risk measures

I. Introduction

When dealing with high dimensional data, it is often undesirable or impractical to store or communicate an entire data set. To address this problem, one solution is to choose a finite number

M. Egan is with Inria, INSA Lyon, CITI, UR3720, 69621 Villeurbanne, France (Email: malcolm.egan@inria.fr).

M of representative examples, thereby compressing or quantizing the data. As the original data cannot be exactly reconstructed from a representative example, the compression is lossy.

Due to the importance of storing and communicating high dimensional data, the problem of designing and analyzing of methods to construct representative examples—also known as *fixed-length lossy source codes* or vector quantizers—has attracted significant attention. The basis of any method to design such source codes is a quality or *distortion* criterion. Moreover, a distortion criterion consists of two components:

(i) A distortion function $D(x, \hat{x})$ which measures the error between the original data x and the corresponding compressed data $\hat{x}$. For example, when $x, \hat{x} \in \mathbb{R}^n$, it is common to choose $D(x, \hat{x}) = \frac{1}{n}\|x - \hat{x}\|_2^2$.

(ii) Viewing the data X as a random variable drawn from the probability distribution P_X, the second component is a statistic of the distortion random variable $D(X, \hat{X})$, where $\hat{X}$ is the compressed version of the data X. A ubiquitous choice is the mean distortion given by $\mathbb{E}_X[D(X, \hat{X})]$.

The choice of a distortion function D is highly application-dependent; both the structure of the data (e.g., the support of P_X) and how the compressed data will be utilized need to be considered. For example, for speech data a range of specialized distortion functions have been proposed [1]. More recently, task-oriented compression has popularized the choice of a wider range of distortion function; e.g., value functions associated with optimization problems [2]. The case of multiple expectation constraints to model semantic aspects of communication has also been considered in [3].

While the distortion function D is often flexible, the choice of statistic for the distortion random variable has been more restricted. The standard choice is the mean distortion $\mathbb{E}_X[D(X, \hat{X})]$. Indeed, standard vector quantization algorithms to construct the source code aim to minimize $\mathbb{E}_X[D(X, \hat{X})]$ [1]. Similarly, the analysis of the required size of the source code, M, in rate-distortion theory typically imposes a constraint on $\mathbb{E}_X[D(X, \hat{X})]$ [4]. In particular, the distortion

in the rate-distortion tradeoff corresponds to $\mathbb{E}_X[D(X, \hat{X})]$.

On the other hand, compression is often only one step in a data processing chain. In this context, compression contributes to uncertainty in data available to users; e.g., for the purpose of decision making. In design or decision making under uncertainty [5], a much wider range of distortion statistics is common. These statistics may include the excess distortion probability $\mathbb{P}_X(D(X, \hat{X}) > \delta)$, the perception constraints in the rate-distortion-perception tradeoff [6], or a family of *risk measures* which are now ubiquitous in financial decision making [7], safe control [8], reliability engineering [9] and reinforcement learning [10]. Further applications of risk measures outside of finance are surveyed in [11]. Developing compression schemes which control more general statistics of the distortion is therefore desirable for design under uncertainty.

The design and analysis of compression schemes subject to constraints on general statistics of the distortion has seen limited attention. Beyond the mean distortion, the main focus has been on the excess distortion probability. Notably, finite blocklength analysis of fixed-length lossy source coding considers excess distortion probability constraints [12]. Nevertheless, to the best of the author's knowledge, there has not been an investigation of the design or analysis of fixed-length lossy source coding subject to constraints on a general family of distortion statistics, such as families of risk measures.

In this paper, we address this question. Namely, we establish conditions for the existence of fixed-length lossy source codes which satisfy general distortion statistics. In particular, we consider the family of *distortion risk measures*, which includes the mean distortion and a close relative of the excess distortion probability—the *value-at-risk* (VaR)—as special cases. Distortion risk measures are parameterized by a *risk distortion function* (not to be confused with the distortion function), which provides a flexible means of constructing distortion statistics.

Our main contributions are as follows:

(i) In the case the risk distortion function associated with a distortion risk measure is concave, we establish conditions for the existence of a fixed-length lossy source code subject to

distortion risk measure constraints. In particular, we establish achievability and converse bounds, which are applicable for general source distributions on arbitrary spaces, and general distortion functions. Notably, our converse argument does not rely on the hypothesis testing method [12]. We illustrate our analysis for the equiprobable memoryless binary source (EMBS) and show that our converse is tighter than the converse bound arising from the hypothesis testing method.

(ii) In practice, fixed-length lossy source codes are typically designed by optimizing the mean distortion, $\mathbb{E}_X[D(X, \hat{X})]$ (e.g., in [1]). Given that a code satisfies a mean distortion constraint, we establish a bound on the distortion risk measure associated with the code, which holds for all distortion risk measures parameterized by concave risk distortion functions.

(iii) We study the impact of varying the source distribution on the performance of a given fixed-length lossy source code. In particular, we establish a bound on changes of the distortion risk measure when the source distribution varies. Our bound depends on both the structure of the corresponding risk distortion function and the total variation distance between the source distributions.

II. Problem Formulation

A. Standard Operational Definitions

A fixed-length lossy source code of size M consists of:

(i) an encoder $P_{Z|X} : \mathcal{X} \to \{1, \ldots, M\}$;

(ii) and a decoder $P_{\hat{X}|Z} : \{1, \ldots, M\} \to \mathcal{X}$.

In general, the encoder and decoder may be probabilistic; i.e., $P_{Z|X}$ and $P_{\hat{X}|Z}$ are random transformations.

The performance of a fixed-length lossy source code is defined in terms of a distortion metric $D : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$. For example, when $\mathcal{X} = \mathbb{R}^n$, it is standard to choose $D(x, \hat{x}) = \frac{1}{n}\|x - \hat{x}\|_2^2$.

On the other hand, for $\mathcal{X} = \{0,1\}^n$, the standard choice is $D(x,\hat{x}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{x_i \neq \hat{x}_i\}$. Other distortion metrics can arise is specific applications; e.g., in speech compression [1], or in goal-oriented communications [2].

The performance of a lossy fixed-length source code is then defined as either:

(i) *Expected Distortion:* $\mathbb{E}[D(X,\hat{X})]$.

(ii) *Excess Distortion Probability:* $\mathbb{P}(D(X,\hat{X}) > \delta)$.

Definition 1. Given a source P_X on $\mathcal{X}$, a (M,δ) code for an expected distortion constraint is the pair $(P_{Z|X}, P_{\hat{X}|Z})$ such that

$$\mathbb{E}[D(X,\hat{X})] \leq \delta. \tag{1}$$

Similarly, a (M,δ,ϵ) code for an exceed distortion probability constraint is the pair $(P_{Z|X}, P_{\hat{X}|Z})$ such that

$$\mathbb{P}(D(X,\hat{X}) > \delta) \leq \epsilon. \tag{2}$$

For fixed δ, ϵ, a widely studied question is the largest M for which a (M,δ) code or (M,δ,ϵ) code exists. When $\mathcal{X} = \mathcal{S}^n$ where $\mathcal{S}$ is a given set, P_X is memoryless and stationary, and D is separable, Shannon rate-distortion theory yields tight bounds on M for large n [4]. General formulae for general sources and non-separable D have been given via information spectrum methods as $n \to \infty$ [13]; however, these results are difficult to evaluate outside of simple cases. For general sources and distortion metrics, and finite dimension n, a characterization of achievable (M,δ,ϵ) codes has been investigated in [12], which are amenable to numerical computation.

B. Distortion Risk Measures

The expected distortion and the excess distortion probability can be viewed as statistics of the distortion random variable $D(X,\hat{X})$. In applications involving design under uncertainty, it is

often desirable to consider more general statistics. One such family is the distortion risk measures [14].

Definition 2. Let $h : [0,1] \to [0,1]$ be a concave non-decreasing function. The distortion risk measure parameterized by h is defined as

$$\rho_h(D(X,\hat{X})) = \int_0^\infty h\left(\mathbb{P}(D(X,\hat{X}) > u)\right) \mathrm{d}u, \tag{3}$$

Example 1 (Expected Distortion)**.** Letting $h(w) = w,\ w \in [0,1]$, we have

$$\rho_h(D(X,\hat{X})) = \int_0^\infty \mathbb{P}(D(X,\hat{X}) > u)\mathrm{d}u = \mathbb{E}[D(X,\hat{X})], \tag{4}$$

yielding the expected distortion as a special case.

Example 2 (Value-at-Risk)**.** The value-at-risk (VaR) at level α [7] is a distortion risk measure with distortion risk function defined as

$$h(w) = \begin{cases} 0, & 0 \le w < 1-\alpha \\ 1, & 1-\alpha \le w \le 1. \end{cases} \tag{5}$$

In particular,

$$\begin{aligned} \rho_h(D(X,\hat{X})) &= \int_0^\infty h(\mathbb{P}(D(X,\hat{X}) > u))\mathrm{d}u \\ &= \int_{u:\mathbb{P}(D(X,\hat{X})>u)\ge 1-\alpha} \mathrm{d}u \\ &= F^{-1}(\alpha), \end{aligned} \tag{6}$$

where[1] $F^{-1}(\alpha) = \inf\{u : \mathbb{P}(D(X,\hat{X}) \le u) > \alpha\}$.

[1]Given a random variable Z, often the VaR is defined as $-\inf\{u : \mathbb{P}(Z \le u) > \alpha\}$ [7], which is due to the fact that negative values of Z are losses. In our case, $D(X,\hat{X}) \ge 0$ and hence large values can be viewed as large losses. As such, we use the definition in (6).

There is a strong link between the VaR and excess distortion probability constraints. Suppose $\rho_h(D(X,\hat{X})) \leq \delta$. Then,

$$\alpha \leq \mathbb{P}(D(X,\hat{X}) \leq \delta), \tag{7}$$

which implies

$$\mathbb{P}(D(X,\hat{X}) > \delta) \leq 1-\alpha. \tag{8}$$

Hence, the VaR constraint implies an excess distortion probability constraint. Similarly, if $\mathbb{P}(D(X,\hat{X}) > \delta) \leq 1-\alpha$, then $1-\mathbb{P}(D(X,\hat{X}) \leq \delta) \leq 1-\alpha$, and hence

$$\rho_h(D(X,\hat{X})) = F^{-1}(\alpha) \leq \delta. \tag{9}$$

As such, when the excess distortion probability constraint is satisfied, the VaR constraint in (9) is also satisfied.

Example 3 (Conditional Value-at-Risk)**.** The conditional value at risk (CVaR) at level $\alpha \in [0,1]$ [5] is a distortion risk measure with distortion risk function defined as

$$h(w) = \begin{cases} \frac{w}{1-\alpha}, & 0 \leq w < 1-\alpha \\ 1, & 1-\alpha \leq w \leq 1. \end{cases} \tag{10}$$

Note that h in (10) is concave, but non-differentiable.

In the case $D(X,\hat{X})$ is a continuous random variable with distribution function F, then

$$\begin{aligned}
\rho_h(D(X,\hat{X})) &= \int_0^\infty h(\mathbb{P}(D(X,\hat{X})>u))\mathrm{d}u \\
&= \int_{u:\mathbb{P}(D(X,\hat{X})>u)\geq 1-\alpha} \mathrm{d}u + \frac{1}{1-\alpha}\int_{u:\mathbb{P}(D(X,\hat{X})>u)\leq 1-\alpha} \mathbb{P}(D(X,\hat{X})>u)\mathrm{d}u \\
&= \int_0^{F^{-1}(\alpha)} \mathrm{d}u + \frac{1}{1-\alpha}\int_{F^{-1}(\alpha)}^\infty \mathbb{P}(D(X,\hat{X})>u)\mathrm{d}u \\
&= \int_0^{F^{-1}(\alpha)} \mathrm{d}u + \int_{F^{-1}(\alpha)}^\infty \frac{\mathbb{P}(D(X,\hat{X})>u)}{\mathbb{P}(D(X,\hat{X})>F^{-1}(\alpha))}\mathrm{d}u \\
&= \int_0^{F^{-1}(\alpha)} \mathrm{d}u + \int_0^\infty \frac{\mathbb{P}(D(X,\hat{X})>u, D(X,\hat{X})>F^{-1}(\alpha))}{\mathbb{P}(D(X,\hat{X})>F^{-1}(\alpha))}\mathrm{d}u \\
&\quad - \int_0^{F^{-1}(\alpha)} \frac{\mathbb{P}(D(X,\hat{X})>u, D(X,\hat{X})>F^{-1}(\alpha))}{\mathbb{P}(D(X,\hat{X})>F^{-1}(\alpha))}\mathrm{d}u \\
&= \int_0^\infty \frac{\mathbb{P}(D(X,\hat{X})>u, D(X,\hat{X})>F^{-1}(\alpha))}{\mathbb{P}(D(X,\hat{X})>F^{-1}(\alpha))}\mathrm{d}u \\
&= \mathbb{E}[D(X,\hat{X})|D(X,\hat{X})>F^{-1}(\alpha)]. \qquad (11)
\end{aligned}$$

Example 4 (Quadration Distortion Risk Measure)**.** The quadratic risk distortion function [14] is defined by

$$h(w) = w + k(w-w^2),\ k\in[0,1], \qquad (12)$$

which is concave and differentiable. Observe that in the case $k=0$, the corresponding risk measure is $\rho_h(D(X,\hat{X})) = \mathbb{E}[D(X,\hat{X})]$.

Example 5 (Power Distortion Risk Measure)**.** The power risk distortion function [14] is defined by

$$h(w) = w^k,\ k\in(0,1], \qquad (13)$$

which is concave and differentiable. Observe in that in the case $k=1$, the corresponding risk measure is $\rho_h(D(X,\hat{X})) = \mathbb{E}[D(X,\hat{X})]$.

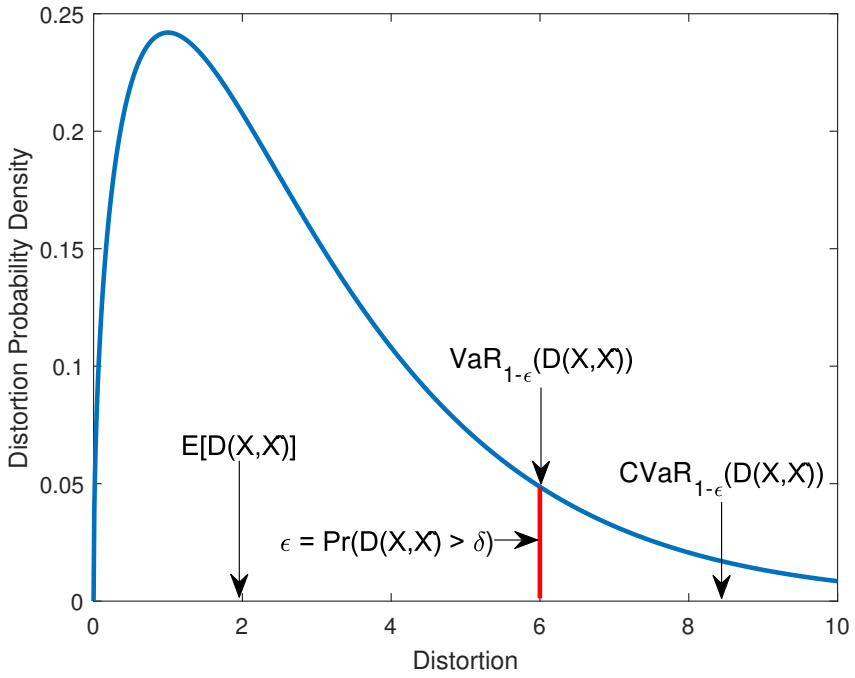

Fig. 1: Illustration of different distortion risk measures.

The risk measures in the previous examples (illustrated in Fig. 1), with the exception of the VaR, admit concave risk distortion functions. This family of distortion risk measures admits the following property.

Lemma 1. Let $h : [0,1] \to [0,1]$ be a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$. Then,

$$\rho_h(D(X,\hat{X})) \geq \mathbb{E}[D(X,\hat{X})]. \tag{14}$$

Proof: Let $w \in [0,1]$. Then, by concavity of h and the fact that $h(0) = 0,\ h(1) = 1$,

$$h(w) = h(w \cdot 1 + (1-w) \cdot 0) \geq wh(1) + (1-w)h(0) = w. \tag{15}$$

As a consequence,

$$\begin{aligned}\rho_h(D(X,\hat{X})) &= \int_0^\infty h\left(\mathbb{P}(D(X,\hat{X})>u)\right)\mathrm{d}u \\ &\geq \int_0^\infty \mathbb{P}(D(X,\hat{X})>u)\mathrm{d}u \\ &= \mathbb{E}[D(X,\hat{X})]. \end{aligned} \tag{16}$$

■

C. Risk-Constrained Compression

Replacing the expected distortion or the excess distortion probability with a risk distortion measure, we obtain a new family of codes containing (M,δ) codes as a special case.

Definition 3. Let $h:[0,1]\to[0,1]$ be a concave non-decreasing function. Given a source P_X on $\mathcal{X}$, a (M,h,δ) code for a distortion risk measure constraint is the pair $(P_{Z|X},P_{\hat{X}|Z})$ such that

$$\rho_h(D(X,\hat{X})) = \int_0^\infty h\left(\mathbb{P}(D(X,\hat{X})>u)\right)\mathrm{d}u \leq \delta. \tag{17}$$

In the following section, we provide conditions for the existence of (M,h,δ) codes. These conditions are analogous to achievability and converse bounds in rate-distortion theory established in [4], [12].

III. Existence of Codes with Distortion Risk Measure Constraints

A. Achievability

As noted in Example 2, there is a close relationship between the VaR and excess distortion probability constraints. As a consequence, it is straightforward to utilize [12, Theorem 10] in order to obtain a characterization of achievability for VaR (see Example 2) constraints.

Theorem 1. Let $\rho_h(D(X,\hat{X}))$ correspond to the VaR at level α (detailed in Example 2), and P_X be a source on $\mathcal{X}$. Then, there exists a (M,h,δ) code such that

$$\rho_h(D(X,\hat{X})) \leq \delta, \tag{18}$$

where

$$\alpha = 1 - \inf_{P_Y} \mathbb{E}_X \left[(\mathbb{P}(D(X,Y) > \delta | X))^M \right], \tag{19}$$

corresponding to

$$h(w) = \begin{cases} 0, & 0 \leq w < \inf_{P_Y} \mathbb{E}_X \left[(\mathbb{P}(D(X,Y) > \delta))^M \right] \\ 1, & \inf_{P_Y} \mathbb{E}_X \left[(\mathbb{P}(D(X,Y) > \delta | X))^M \right] \leq w \leq 1. \end{cases} \tag{20}$$

Proof: By [12, Theorem 10], a (M,δ,ϵ) code exists satisfying

$$\mathbb{P}(D(X,\hat{X}) > \delta) = \epsilon \leq \inf_{P_Y} \mathbb{E}_X \left[(\mathbb{P}(D(X,Y) > \delta | X))^M \right]. \tag{21}$$

As discussed in Example 2, this implies that

$$F^{-1}\left(1 - \inf_{P_Y} \mathbb{E}_X \left[(\mathbb{P}(D(X,Y) > \delta | X))^M \right]\right) = \rho_h(D(X,\hat{X})) \leq \delta, \tag{22}$$

with h defined as in (20). ■

We now consider a general family of risk measures, obtained via concave risk distortion functions (including Examples 1, 3, 4, 5). Note that the case $h(w) = w$, corresponding to $\rho_h(D(X,\hat{X})) = \mathbb{E}[D(X,\hat{X})]$, lies in this family.

Theorem 2. Let $h : [0,1] \to [0,1]$ be a concave non-decreasing function satisfying $h(0) = 0$ and $h(1) = 1$, and P_X be a source on $\mathcal{X}$. Then, there exists a (M,h,δ) code such that

$$\delta \leq \inf_{P_Y} \int_0^\infty h\left(\mathbb{E}_X \left[\left(\mathbb{P}\left(D(X,Y) > u | X \right) \right)^M \right] \right) \mathrm{d}u, \tag{23}$$

where Y is any random variable on $\mathcal{X}$, independent of X.

Proof: Consider the codewords $\{Y_1, \ldots, Y_M\}$ drawn i.i.d. from P_Y, and the encoding rule

$$j^* \in \arg\min_{j=1,\ldots,M} D(X,Y_j), \tag{24}$$

where ties are broken arbitrarily. Then,

$$\begin{aligned}
&\mathbb{E}_{Y_1,\ldots,Y_M}\left[\int_0^\infty h\left(\mathbb{P}_X\left(D(X,\hat{X})>u\right)\right)\mathrm{d}u\right]\\
&\overset{(a)}{\leq}\int_0^\infty h\left(\mathbb{E}_X\left[\mathbb{E}\left[\prod_{j=1}^{M}\mathbf{1}\{D(X,Y_i)>u\}|X\right]\right]\right)\mathrm{d}u\\
&\overset{(b)}{=}\int_0^\infty h\left(\mathbb{E}_X\left[\left(\mathbb{P}(D(X,Y)>u|X)\right)^M\right]\right)\mathrm{d}u,
\end{aligned}\tag{25}$$

where (a) follows from the concavity of h and Jensen's inequality, and (b) follows from the fact that $Y_1,\ldots,Y_M$ are i.i.d. and independent of X.

By Shannon's random coding argument, it follows there exists a (M,h,δ) code satisfying

$$\delta\leq\inf_{P_Y}\int_0^\infty h\left(\mathbb{E}_X\left[\left(\mathbb{P}_{Y|X}(D(X,Y)>u|X)\right)^M\right]\right)\mathrm{d}u,\tag{26}$$

as required. ■

B. Converse

We now turn to establishing converse bounds. As for the achievability bounds, we first exploit the connection between the VaR and excess distortion probability constraints to apply the results in [12] for the case of VaR constraints. This converse bound is a counterpart to the achievability bound in Theorem 1. The basis for this bound is the following hypothesis test, introduced in [12].

Let P_X be a source and Q_X be a comparison distribution on $\mathcal{X}$. Define the hypothesis test

$$\begin{aligned}
\mathcal{H}_0&: X\sim Q_X\\
\mathcal{H}_1&: X\sim P_X.
\end{aligned}\tag{27}$$

Let $W\in\{0,1\}$ be the decision variable. By the Neyman-Pearson lemma, the minimum type-II error probability subject to a type-I error constraint is given by

$$\beta_\alpha(P_X,Q_X)=\min_{\substack{P_{W|X}:\\ \mathbb{P}_X(W=1)\geq\alpha}}\mathbb{Q}_X(W=1).\tag{28}$$

Theorem 3. Let $\rho_h(D(X,\hat{X}))$ correspond the VaR at level α (detailed in Example 2), and P_X be a source on $\mathcal{X}$. Let $(P_{Z|X}, P_{\hat{X}|Z})$ be a (M,h,δ) code satisfying

$$\rho_h(D(X,\hat{X})) \leq \delta. \tag{29}$$

Then,

$$M \geq \sup_{Q_X} \inf_{y} \frac{\beta_\alpha(P_X, Q_X)}{\mathbb{Q}_X(D(X,y) \leq \delta)}, \tag{30}$$

where Q_X is any distribution on $\mathcal{X}$ and $\beta_\alpha(P_X, Q_X)$ is defined in (28).

Proof: By assumption, $\rho_h(D(X,\hat{X})) \leq \delta$. By the definition of the VaR (see Example 2), we have

$$\mathbb{P}(D(X,\hat{X}) > \delta) \leq 1 - \alpha. \tag{31}$$

It then follows from [12, Theorem 8] that

$$M \geq \sup_{Q_X} \inf_{y} \frac{\beta_\alpha(P_X, Q_X)}{\mathbb{Q}_X(D(X,y) \leq \delta)}, \tag{32}$$

as required. ■

We now establish a converse in the case of concave risk distortion functions. This converse bound corresponds to a counterpart of the achievability bound in Theorem 2, and applies the hypothesis testing method in [12, Theorem 8].

Theorem 4. Let $h : [0,1] \to [0,1]$ be a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$, P_X be a source on $\mathcal{X}$, and $(P_{Z|X}, P_{\hat{X}|Z})$ be a (M,h,δ) code. Then,

$$M \geq \sup_{w \in \mathbb{R}_+} \sup_{Q_X} \inf_{y} \frac{\beta_{1-\frac{\delta}{w}}(P_X, Q_X)}{\mathbb{Q}_X(D(X,y) \leq w)}, \tag{33}$$

where Q_X is any distribution on $\mathcal{X}$ and $\beta_{1-\frac{\delta}{w}}(P_X, Q_X)$ is defined in (35).

Proof: Let Q_X be a comparison distribution on $\mathcal{X}$ and $w \in \mathbb{R}_+$. Define the hypothesis test

$$\mathcal{H}_0 : X \sim Q_X$$

$$\mathcal{H}_1 : X \sim P_X. \tag{34}$$

Let $W \in \{0,1\}$ be the decision variable. By the Neyman-Pearson lemma, the minimum type-II error probability subject to a type-I error constraint is given by

$$\beta_{1-\frac{\delta}{w}}(P_X, Q_X) = \min_{\substack{P_{W|X}: \\ \mathbb{P}_X(W=1) \geq 1-\frac{\delta}{w}}} \mathbb{Q}_X(W=1). \tag{35}$$

Consider the sub-optimal decision rule

$$W = \mathbf{1}\left\{ \int_0^\infty h\left(\mathbf{1}\{D(X,\hat{X}) > u\} \right) \mathrm{d}u \leq w \right\}. \tag{36}$$

Then,

$$\begin{aligned}
\mathbb{P}_X(W=1) &= \mathbb{P}_X\left(\int_0^\infty h\left(\mathbf{1}\{D(X,\hat{X}) > u\} \right) \mathrm{d}u \leq w \right) \\
&\overset{(a)}{\geq} 1 - \frac{1}{w}\mathbb{E}\left[\int_0^\infty h\left(\mathbf{1}\{D(X,\hat{X}) > u\} \right) \mathrm{d}u \right] \\
&\overset{(b)}{\geq} 1 - \frac{1}{w}\int_0^\infty h(\mathbb{P}_X(D(X,\hat{X}) > u)) \mathrm{d}u \\
&\overset{(c)}{\geq} 1 - \frac{\delta}{w},
\end{aligned} \tag{37}$$

where (a) follows from the Markov inequality, (b) exploits concavity of h, and (c) utilizes the assumption that $(P_{Z|X}, P_{\hat{X}|Z})$ is a (M, h, δ) code. As such, this decision rule satisfies the feasibility constraint.

We then have [2]

$$\begin{aligned}
\beta_{1-\frac{\delta}{w}}(P_X, Q_X) &\leq \mathbb{Q}_X(W=1) \\
&= \sum_x \sum_{z=1}^{M} \sum_y Q_X(x) P_{Z|X}(z|x) P_{\hat{X}|Z}(y|z) \mathbf{1}\left\{ \int_0^\infty h\left(\mathbf{1}\{D(X,\hat{X}) > u\} \right) \mathrm{d}u \leq w \right\} \\
&\leq M \sup_y \mathbb{Q}_X \left(\int_0^\infty h\left(\mathbf{1}\{D(X,y) > u\} \right) \mathrm{d}u \leq w \right) \\
&= M \sup_y \mathbb{Q}_X \left(\int_0^\infty \mathbf{1}\{D(X,y) \leq u\} \mathrm{d}u \leq w \right) \\
&= M \sup_y \mathbb{Q}_X(D(X,y) \leq w). \qquad (38)
\end{aligned}$$

As the choice of Q_X and w is arbitrary, it follows that

$$M \geq \sup_{w \in \mathbb{R}_+} \sup_{Q_X} \inf_y \frac{\beta_{1-\frac{\delta}{w}}(P_X, Q_X)}{\mathbb{Q}_X(D(X,y) \leq w)}. \qquad (39)$$

■

Remark 1. An alternative method to prove Theorem 4 is to directly utilize the converse bound in [12, Theorem 8] for the case of excess distortion probability constraints. For all $w \in \mathbb{R}_+$, we have by Lemma 1

$$\rho_h(D(X,\hat{X})) \geq \mathbb{E}[D(X,\hat{X})] \geq w\mathbb{P}(D(X,\hat{X}) > w). \qquad (40)$$

In particular, this implies that if $\rho_h(D(X,\hat{X})) \leq \delta$, then $\mathbb{P}(D(X,\hat{X}) > w) \leq \frac{\delta}{w}$. The converse bound in Theorem 4 then follows immediately from [12, Theorem 8].

We note that the bound in Theorem 4 does not depend on h, which is due to the fact the proof implicitly utilizes the fact $\rho_h(D(X,\hat{X})) \geq w\mathbb{P}(D(X,\hat{X}) > w)$, as noted in Remark 1. We may expect that a tighter bound can be obtained by incorporating information about the function h

[2]Note that the same analysis applies for continuous sources, where the sums are replaced by integrals and densities replace the probability mass functions.

parametrizing the distortion risk measure. We provide such a bound in Theorem 5, which does not rely on the hypothesis testing method.

Theorem 5. Let $h : [0,1] \to [0,1]$ be a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$, P_X be a source on $\mathcal{X}$, and $(P_{Z|X}, P_{\hat{X}|Z})$ be a (M, h, δ) code. Then,

$$\delta \geq \sup_{u_{\max} \geq 0} \int_0^{u_{\max}} h\left(1 - M \sup_y \mathbb{P}_X(D(X,y) \leq u)\right) \mathrm{d}u. \tag{41}$$

Proof: By assumption,

$$\delta \geq \int_0^{\infty} h\left(\mathbb{P}(D(X,\hat{X}) > u)\right) \mathrm{d}u. \tag{42}$$

Let $u_{\max} \geq 0$. Then,

$$\begin{aligned}
\delta &\overset{(a)}{\geq} \int_0^{u_{\max}} h\left(\sum_{z=1}^{M} \frac{1}{M} - \sum_x \sum_{z=1}^{M} \sum_y P_X(x) P_{Z|X}(z|x) P_{\hat{X}|Z}(y|z) \mathbf{1}\{D(x,y) \leq u\}\right) \mathrm{d}u \\
&\overset{(b)}{\geq} \frac{1}{M} \sum_{z=1}^{M} \int_0^{u_{\max}} h\left(1 - M \sum_x \sum_y P_X(x) P_{Z|X}(z|x) P_{\hat{X}|Z}(y|z) \mathbf{1}\{D(x,y) \leq u\}\right) \mathrm{d}u \\
&\overset{(c)}{\geq} \frac{1}{M} \sum_{z=1}^{M} \int_0^{u_{\max}} h\left(1 - M \sum_y P_{\hat{X}|Z}(y|z) \sum_x P_X(x) \mathbf{1}\{D(x,y) \leq u\}\right) \mathrm{d}u \\
&\geq \frac{1}{M} \sum_{z=1}^{M} \int_0^{u_{\max}} h\left(1 - M \sup_y \mathbb{P}_X(D(X,y) \leq u)\right) \mathrm{d}u \\
&= \int_0^{u_{\max}} h\left(1 - M \sup_y \mathbb{P}_X(D(X,y) \leq u)\right) \mathrm{d}u,
\end{aligned} \tag{43}$$

where (a) follows from the definition of $\mathbb{P}(D(X,\hat{X}) > u)$, (b) follows from the concavity of h and Jensen's inequality, and (c) follows from the fact $P_{Z|X}(z|x) \leq 1$.

As the choice of $u_{\max}$ is arbitrary, it follows that

$$\delta \geq \sup_{u_{\max} \geq 0} \int_0^{u_{\max}} h\left(1 - M \sup_y \mathbb{P}_X(D(X,y) \leq u)\right) \mathrm{d}u, \tag{44}$$

as required. ■

C. Example

Consider the stationary equiprobable binary source P_{X^n} on $\{0,1\}^n$ with $\mathbb{P}(X^n = x^n) = 2^{-n}$ and distortion metric

$$D(x^n, y^n) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{x_i \neq y_i\}. \tag{45}$$

Let

$$\left\langle {n \atop k} \right\rangle = \sum_{j=0}^{k} \binom{n}{j} \tag{46}$$

where $\left\langle {n \atop k} \right\rangle = 0$ if $k < 0$ and $\left\langle {n \atop k} \right\rangle = \left\langle {n \atop n} \right\rangle$ if $k > n$.

Theorem 6. Let $P_X = P_{X^n}$ be the stationary equiprobable binary source, $h : [0,1] \to [0,1]$ be a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$, $(P_{Z|X}, P_{\hat{X}|Z})$ be a (M, h, δ) code and $w \in \mathbb{R}_+$. Then,

$$\delta \geq w\left(1 - 2^{-n} M \left\langle {n \atop \lfloor nu \rfloor} \right\rangle\right). \tag{47}$$

Proof: Applying Theorem 4, we have

$$\begin{aligned} M &\geq \sup_{Q_X} \inf_{y^n \in \{0,1\}^n} \frac{\beta_{1-\frac{\delta}{w}}(P_X, Q_X)}{\mathbb{Q}_X(D(X^n, y^n) \leq w)} \\ &\geq \inf_{y \in \{0,1\}^n} \frac{\beta_{1-\frac{\delta}{w}}(P_X, P_X)}{\mathbb{P}_X(D(X^n, y^n) \leq w)} \\ &= \frac{1 - \frac{\delta}{w}}{\mathbb{P}(D(X^n, 0) \leq w)} \\ &= \frac{1 - \frac{\delta}{w}}{2^{-n}\left\langle {n \atop \lfloor nu \rfloor} \right\rangle}. \end{aligned} \tag{48}$$

Re-arranging this bound then yields the desired result. ■

Theorem 7. Let $P_X = P_{X^n}$ be the stationary equiprobable binary source, $h : [0,1] \to [0,1]$ be a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$, and $(P_{Z|X}, P_{\hat{X}|Z})$ be a (M, h, δ) code. Then,

$$\delta \geq \sup_{u_{\max} \geq 0} \int_0^{u_{\max}} h\left(1 - 2^{-n} M \left\langle {n \atop \lfloor nu \rfloor} \right\rangle\right) \mathrm{d}u. \tag{49}$$

Proof: Observe that

$$\sup_{y\in\{0,1\}^n} \mathbb{P}(D(X^n, y^n) \leq u) = \mathbb{P}(D(X^n, 0) \leq u) = 2^{-n}\left\langle \begin{matrix} n \\ \lfloor nu \rfloor \end{matrix} \right\rangle. \tag{50}$$

Applying Theorem 5 then yield the desired result. ■

Lemma 2. Let $P_X = P_{X^n}$ be the stationary equiprobable binary source and $h : [0,1] \to [0,1]$ be a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$. Then,

$$\min_{P_Y} \int_0^\infty h\left(\mathbb{E}_X\left[\left(\mathbb{P}_Y\left(D(X^n, Y^n) > u\right)\right)^M\right]\right) \mathrm{d}u = \int_0^\infty h\left(\left(1 - 2^{-n}\right)^M\right) \mathrm{d}u. \tag{51}$$

Proof: By concavity of $(1-x)^M$ for $x \in [0,1]$, it follows that for any P_Y on $\{0,1\}^n$,

$$\int_0^\infty h\left(\mathbb{E}_X\left[(1 - \mathbb{P}_Y(D(X^n, Y^n) \leq u))^M\right]\right) \mathrm{d}u \geq \int_0^\infty h\left(\left(1 - \mathbb{E}_X\left[\mathbb{P}_Y\left(D(X^n, Y^n) \leq u\right)\right]\right)^M\right) \mathrm{d}u. \tag{52}$$

Observing that equality is achieved when Y^n is equiprobable yields the desired result. ■

Theorem 8. Let $P_X = P_{X^n}$ be the stationary equiprobable binary source and $h : [0,1] \to [0,1]$ be a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$. Then, there exists a (M, h, δ) code such that

$$\delta \leq \int_0^\infty h\left(\left(1 - 2^{-n}\left\langle \begin{matrix} n \\ \lfloor nu \rfloor \end{matrix} \right\rangle\right)^M\right) \mathrm{d}u. \tag{53}$$

Proof: The desired result follows immediately by combining Lemma 2 and Theorem 2. ■

Fig. 2 and Fig. 3 plot the bounds on the risk constraint δ for varying dimension of the source n. In Fig. 2, we consider $M = 10$ and a quadratic risk distortion constraint (defined in Example 4) with $k = 1$. In Fig. 3, we consider $M = 10$ and a quadratic risk distortion constraint (defined in Example 4) with $k = 0.5$. In the bounds, the quantities w and $u_{\max}$ are optimized via a grid search.

Observe that as the dimension of the source increases, the achievable distortion risk constraint δ also increases. This is due to the fact that compressing higher dimensional data at the same

level of risk requires larger values of M. This is true even for the standard rate-distortion model with an expectation constraint. Indeed, the rate is defined as $\frac{\log M}{n}$, which implies to achieve a given expected distortion criterion as n increases requires a corresponding increase in M.

We also observe that our converse bound in Theorem 5 is tighter than the converse in Theorem 4 based on the hypothesis testing method. As noted in Remark 1, this is in part due to the fact that the risk distortion function is not taken into account.

Finally, comparing Fig. 2 and Fig. 3 shows that increasing the parameter k in the quadratic risk distortion function (see Example 4) leads to an increase in all the bounds on the achievable constraint δ. This shows that the choice of risk measure has a significant impact on the achievable values of the constraint δ. Moreover, increasing the weight on $\mathbb{P}(D(X,\hat{X}) > u)$ (i.e., reducing the tolerance for risk), leads to an increase in the achievable values of the constraint δ.

IV. Impact of the Risk Distortion Function h

Popular algorithms for the design of fixed-length lossy source codes are designed to minimize the expected distortion $\mathbb{E}[D(X,\hat{X})]$. These algorithms include the Lloyd algorithms and their generalizations (e.g., the fixed point vector quantization algorithm [1] and [2]). As a consequence, it is useful to characterize the performance of codes designed to minimize $\mathbb{E}[D(X,\hat{X})]$ in terms of the risk measure $\rho_h(D(X,\hat{X}))$. We provide a bound of this form in the following theorem.

Theorem 9. Suppose $D : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ satisfies $D(x,\hat{x}) \leq B,\ \forall x,\hat{x} \in \mathcal{X}$, $h : [0,1] \to [0,1]$ is a concave non-decreasing function with $h(0) = 0$ and $h(1) = 1$. Let $(P_{Z|X}, P_{\hat{X}|Z})$ be a code satisfying

$$\mathbb{E}[D(X,\hat{X})] \leq \delta \leq B. \tag{54}$$

Then,

$$\rho_h(D(X,\hat{X})) \leq Bh\left(\frac{\delta}{B}\right) \leq B. \tag{55}$$

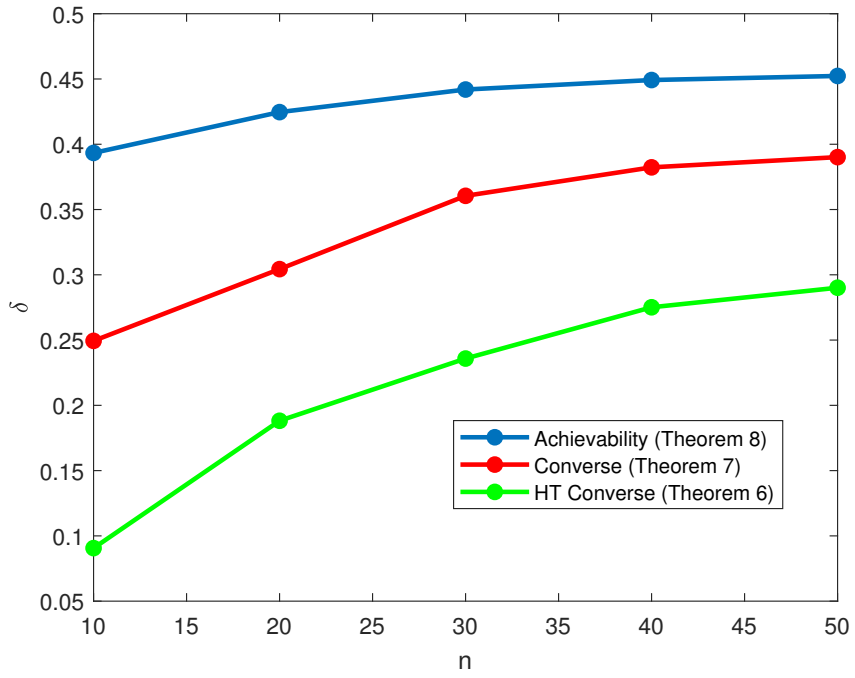

Fig. 2: Plot of bounds for the EMBS source with $M = 10$ and quadratic risk distortion function with $k = 1$.

Proof: As the distortion metric is bounded, it follows that $\mathbb{E}[D(X, \hat{X})] \leq B$. Moreover,

$$\begin{aligned}
\rho_h(D(X,\hat{X})) &= B \cdot \frac{1}{B} \int_0^B h\left(\mathbb{P}(D(X,\hat{X}) > u)\right) \mathrm{d}u \\
&\overset{(a)}{\leq} Bh\left(\frac{1}{B}\int_0^B \mathbb{P}(D(X,\hat{X}) > u)\mathrm{d}u\right) \\
&\overset{(b)}{\leq} Bh\left(\frac{\mathbb{E}[D(X,\hat{X})]}{B}\right) \\
&\overset{(c)}{\leq} Bh\left(\frac{\delta}{B}\right), \qquad (56)
\end{aligned}$$

where (a) follows from concavity of h and Jensen's inequality, (b) follows from the fact $\mathbb{E}[D(X, \hat{X})] = \int_0^\infty \mathbb{P}(D(X, \hat{X}) > u)\mathrm{d}u$, and (c) follows by the assumption in (54). ■

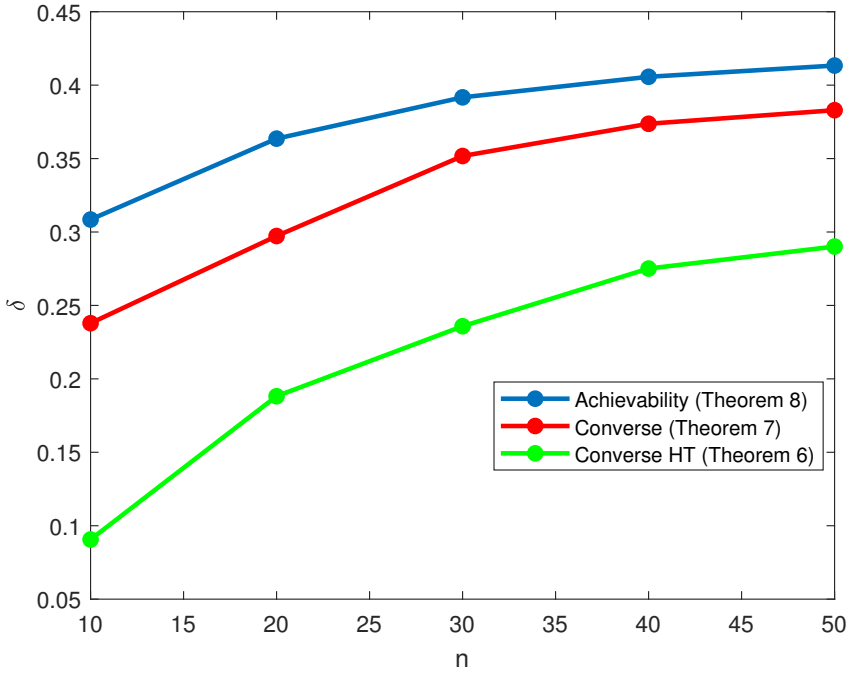

Fig. 3: Plot of bounds for the EMBS source with $M = 10$ and quadratic risk distortion function with $k = 0.5$.

Remark 2. The bound in Theorem 9 is tight in the sense that when $h(w) = w$,

$$\delta = \mathbb{E}[D(X, \hat{X})] = \rho_h(D(X, \hat{X})) \leq \delta. \tag{57}$$

Example 6. Let the assumptions in Theorem 9 hold and h be the quadratic risk distortion function (see Example 4) defined by

$$h(w) = w + k(w - w^2), \ \ k \in (0, 1]. \tag{58}$$

Then,

$$\begin{aligned} \rho_h(D(X, \hat{X})) &\leq B\left(\frac{\delta}{B} + k\left(\frac{\delta}{B} - \frac{\delta^2}{B^2}\right)\right) \\ &= \delta\left(1 + k - \frac{k\delta}{B}\right). \end{aligned} \tag{59}$$

Example 7. Let the assumptions in Theorem 9 hold and h be the power risk distortion function (see Example 5) defined by

$$h(w) = w^k, \ k \in (0,1). \tag{60}$$

Then,

$$\rho_h(D(X,\hat{X})) \leq B^{1-k}\delta^k. \tag{61}$$

V. Impact of Varying the Source Distribution

In many applications, the true source distribution is not perfectly known. In this case, it is desirable to characterize the impact of uncertainty in the source distribution on $\mathcal{X}$. For a bounded distortion metric $D(x,y) \leq B, \ x,y \in \mathcal{X}$, a useful fact is that

$$|\mathbb{E}_{P_1}[D(X,\hat{X})] - \mathbb{E}_{P_2}[D(X,\hat{X})]| \leq B\|P_1 - P_2\|_{\mathrm{TV}}, \tag{62}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. The following theorem gives an analogous bound for the case where the expectation is replaced with a risk measure.

Theorem 10. Let P_{X_1}, P_{X_2} be probability distributions on $\mathcal{X}$. Let h $h : [0,1] \to [0,1]$ be a differentiable non-decreasing function with $h(0) = 0$ and $h(1) = 1$ satisfying

$$\sup_{\xi \in (0,1)} \frac{\mathrm{d}h(\xi)}{\mathrm{d}w} < \infty. \tag{63}$$

Let $(P_{Z|X}, P_{\hat{X}|Z})$ be a fixed-length lossy source code, $D(X_1,\hat{X}_1)$, $D(X_2,\hat{X}_2)$ be the distortion random variables induced by the code and source distribution P_{X_1}, P_{X_2}, respectively. If $D(x,\hat{x}) \leq B, \ \forall x,\hat{x} \in \mathcal{X}$, then

$$|\rho_h(D(X_1,\hat{X}_1)) - \rho_h(D(X_2,\hat{X}_2))| \leq \sup_{\xi \in (0,1)} \frac{\mathrm{d}h(\xi)}{\mathrm{d}w} \cdot 2B\|P_{X_1} - P_{X_2}\|_{TV}, \tag{64}$$

where $\|\cdot\|_{TV}$ denotes the total variation norm.

Proof: Let ν be a probability measure ν such that P_{X_1} and P_{X_2} are absolutely continuous with respect to ν. We then have

$$
\begin{aligned}
&|\rho_h(D(X_1,\hat{X}_1)) - \rho_h(D(X_2,\hat{X}_2))| \\
&= \left| \int_0^\infty h(\mathbb{P}(D(X_1,\hat{X}_1) > u))\mathrm{d}u - \int_0^\infty h(\mathbb{P}(D(X_2,\hat{X}_2) > u))\mathrm{d}u \right| \\
&\overset{(a)}{=} \left| \int_0^\infty \frac{\mathrm{d}h(\xi(u))}{\mathrm{d}w} \left(\mathbb{P}(D(X_1,\hat{X}_1)) - \mathbb{P}(D(X_2,\hat{X}_2)) \right) \mathrm{d}u \right| \\
&= \left| \int_0^\infty \frac{\mathrm{d}h(\xi(u))}{\mathrm{d}w} \left(\mathbb{E}_{X_1}[\mathbb{E}[\mathbf{1}\{D(X_1,\hat{X}_1) > u\}|X_1]] - \mathbb{E}_{X_2}[\mathbb{E}[\mathbf{1}\{D(X_2,\hat{X}_2) > u\}|X_2]] \right) \mathrm{d}u \right| \\
&\overset{(b)}{\leq} \int_0^\infty \frac{\mathrm{d}h(\xi(u))}{\mathrm{d}w} \int_0^\infty \left| \mathbb{E}[\mathbf{1}\{D(X,\hat{X}) > u\}|X] \left(\frac{\mathrm{d}P_{X_1}}{\mathrm{d}\nu} - \frac{\mathrm{d}P_{X_2}}{\mathrm{d}\nu} \right) \right| \mathrm{d}\nu \\
&\leq \int_0^\infty \frac{\mathrm{d}h(\xi(u))}{\mathrm{d}w} \int_0^\infty \left| \frac{\mathrm{d}P_{X_1}}{\mathrm{d}\nu} - \frac{\mathrm{d}P_{X_2}}{\mathrm{d}\nu} \right| \mathrm{d}\nu \\
&\overset{(c)}{\leq} 2B \sup_\xi \frac{\mathrm{d}h(\xi)}{\mathrm{d}w} \|P_{X_1} - P_{X_2}\|_{\mathrm{TV}}.
\end{aligned} \tag{65}
$$

Here, (a) follows from the mean-value theorem. (b) follows from the Radon-Nikodym theorem. (c) follows from the definition of the total variation norm. ■

Remark 3. Observe that in the case $h(w) = w$, $\rho_h(D(X,\hat{X})) = \mathbb{E}[D(X,\hat{X})]$ and Theorem 10 reduces to (64).

A. Example

Theorem 9 can be applied when h is differentiable on $(0,1)$, and $\sup_\xi \frac{\mathrm{d}h(\xi)}{\mathrm{d}w} < \infty$. These conditions hold for the quadratic risk distortion function (see also Example 4)

$$h(w) = w + k(w - w^2), k \in [0,1], \tag{66}$$

where

$$\frac{\mathrm{d}h(\xi)}{\mathrm{d}w} = 1 + k - 2k\xi \leq 1 + k, \ \xi \in [0,1]. \tag{67}$$

Suppose $D(x,\hat{x}) \leq B,\ \forall x,\hat{x} \in \mathcal{X}$. In this case, Theorem 10 yields

$$\begin{aligned} &|\rho_h(D(X_1,\hat{X}_1)) - \rho_h(D(X_2,\hat{X}_2))| \\ &\leq 2(1+k)B\|P_{X_1} - P_{X_2}\|_{\mathrm{TV}},\ k \in [0,1], \end{aligned} \tag{68}$$

where $X_1 \sim P_{X_1}$ and $X_2 \sim P_{X_2}$ with common marginals. This suggests that as k is increased, there is a larger impact of variations in the probability distribution in terms of the total variation norm.

VI. Conclusions

In design under uncertainty, it is often desirable to taken into account tolerance for risk. In particular, the impact of uncertainty is quantified via risk measures (e.g., distortion risk measures) rather than expectations. A common source of uncertainty in the presence of storage or communication constraints is lossy compression.

In this paper, we have studied fixed-length lossy compression subject to distortion risk measure constraints. We have established conditions for a code satisfying a risk measure constraint to exist, which hold for general source distributions supported on general spaces, and general distortion functions. We also investigated the impact of design based on expected distortion constraints in terms of distortion risk measures, as well as the impact of variations in the source distribution.

References

[1] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.

[2] H. Zou *et al.*, "Goal-oriented quantization: analysis, design, and application to resource allocation," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 42–54, 2022.

[3] P. Stavrou and M. Kountouris, "The role of fidelity in goal-oriented semantic communication: a rate distortion approach," *IEEE Transactions on Communications*, 2023.

[4] T. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, NJ: John Wiley & Sons, 2006.

[5] R. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, pp. 21–42, 2000.

[6] Y. Blau and T. Michaeli, "Rethinking lossy compression: the rate-distortion-perception tradeoff," in *International Conference on Machine Learning*, 2019.

[7] P. Artzner *et al.*, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.

[8] A. Singletary, M. Ahmadi, and A. Ames, "Safe control for nonlinear system with stochastic uncertainty via risk control barrier functions," *IEEE Control Systems Letters*, vol. 7, pp. 349–254, 2022.

[9] R. Rockafellar and J. Royset, "Engineering decisions under risk averseness," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, vol. 1, no. 2, 2015.
[10] O. Mihatsch and R. Neuneier, "Risk-sensitive reinforcement learning," *Machine Learning*, vol. 49, pp. 267–290, 2002.
[11] C. Filippi, G. Guastaroba, and M. Speranza, "Conditional value-at-risk beyond finance: a survey," *International Transactions in Operational Research*, vol. 27, no. 3, pp. 1277–1319, 2020.
[12] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transcations on Information Theory*, vol. 58, no. 6, pp. 3309–3338, 2012.
[13] T. Han, *Information-Spectrum Methods in Information Theory*. Berlin: Springer-Verlag, 2003.
[14] E. Sereda *et al.*, "Distortion risk measures in portfolio optimization," *Handbook of Portfolio Construction*, pp. 649–673, 2010.