The Rate-Distortion-Perception Trade-off with Side Information

Yassine Hamdi*, Deniz Gündüz*
*Department of Electrical and Electronic Engineering, Imperial College London, UK,
{y.hamdi, d.gunduz}@imperial.ac.uk

***Abstract*—In image compression, with recent advances in generative modeling, the existence of a trade-off between the rate and the perceptual quality has been brought to light, where the perception is measured by the closeness of the output distribution to the source. This leads to the question: how does a perception constraint impact the trade-off between the rate and traditional distortion constraints, typically quantified by a single-letter distortion measure? We consider the compression of a memoryless source X in the presence of memoryless side information Z, studied by Wyner and Ziv, but elucidate the impact of a perfect realism constraint, which requires the output distribution to match the source distribution. We consider two cases: when Z is available only at the decoder or at both the encoder and the decoder. The rate-distortion trade-off with perfect realism is characterized for sources on general alphabets when infinite common randomness is available between the encoder and the decoder. We show that, similarly to traditional source coding with side information, the two cases are equivalent when X and Z are jointly Gaussian under the squared error distortion measure. We also provide a general inner bound in the case of limited common randomness.**

I. Introduction

In conventional rate-distortion theory, the goal is to enable the decoder to reconstruct a representation $Y^n \triangleq (Y_1, ..., Y_n)$ of the source signal $X^n = (X_1, ..., X_n)$ that is close to the latter for some distortion measure $d(X^n, Y^n)$. Shannon characterized the optimal rate-distortion trade-off under an additive distortion measure, i.e. $d(x^n, y^n) = (1/n)\sum_{i=1}^n d(x_i, y_i)$. In [1], Wyner and Ziv generalized this result to the case where a side information Z^n, correlated with X^n, is available either only at the decoder or at both the encoder and the decoder.

Recently, there has been a renewed interest in compression algorithms since methods based on deep neural networks (DNNs) have been shown to outperform traditional image and video compression codecs [2]–[14] under different distortion measures. In [15], the authors used generative adversarial networks (GANs) to push the limits of image compression in very low bit-rates by synthesizing image content, such as facades of buildings, using a reference image database. This allows the receiver to generate images that resemble the source image semantically, although they may not match perfectly in details, providing visually pleasing reconstructions even at very low bit-rates. It has been observed ([15]–[18]) that at such bit-rates the increase in perceptive quality comes at the cost of increased distortion, and the *rate-distortion-perception trade-off* was formalized in [19]–[21].

Motivated by successful results in generative modeling, where the generated images would exhibit the same statistical properties of the images in the dataset, the formalism of distribution-preserving image compression [16] was adopted in [18], and extended in [19]–[21]. The problem is then to characterize the optimal rate for which both the distortion $d(X^n, Y^n) \leq \Delta$, and the perception $\delta(P_{X^n}, P_{Y^n}) \leq \lambda$, where δ is a similarity measure, e.g., the total variation distance or a divergence. This *strong perception constraint* can be replaced by weaker variants [22]. In conventional rate-distortion theory, it is known that deterministic encoders are sufficient to achieve the optimal rate-distortion performance. This simplifies both the analysis and implementation of rate-distortion optimal codes. However, for the rate-distortion-perception trade-off, it has been shown in [23] that stochastic encoders can be strictly better. This is extensively studied in [22].

The characterization of the optimal rate-distortion-perception trade-off for variable-rate codes and arbitrary perception measures is given in [24]. See also [20] in the case of a deterministic decoder and for general information sources and [22] in the case of weaker perception constraints. The characterization for fixed-rate codes for the perfect realism case, i.e., $\lambda = 0$, is given in [25] and generalized in [26] to a larger family of distortion measures and alphabets. The impact of the amount of available common randomness on the achievable trade-off is explicitly shown. The optimal rate-distortion trade-off for perfect perception is also explicitly characterized [26] for a Gaussian source and mean-squared error distortion measure. When sufficient common randomness is available, this result boils down to the one in [16].

First used in another context, a random coding technique to construct distribution-matching stochastic encoder-decoder pairs was developed in [27] and [28]. It involves the *soft covering lemma*, and is central in [26] and the present paper.

Here, we further push the understanding of how perception constraints affect traditional information-theoretic results by studying the impact of the near-perfect realism constraint, i.e. $\lambda = 0$, on the traditional problem of lossy compression in the presence of side information. We consider two cases: when the side information Z^n is available only at the decoder or at both

The present work has received funding from the European Union's Horizon 2020 Marie Skłodowska Curie Innovative Training Network Greenedge (GA. No. 953775). It has also received funding from UKRI (EP/X030806/1) for the project AIR (ERC-CoG).

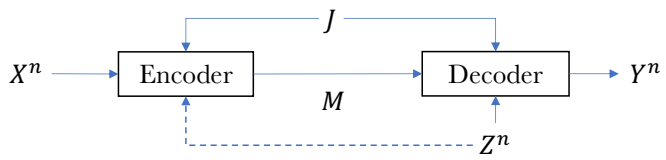

Fig. 1. The system model. Side information Z^n is always available at the decoder, but not necessarily at the encoder.

the encoder and the decoder. In Section II, after formalizing the problem, we prove the equivalence between achievability between two notions of asymptotically perfect realism, similarly to [26]. This result, interesting in itself, is key to handling general alphabets in the rest of this paper. In Section III, we state our main result: a single-letter characterization of the rate-distortion region for fixed-rate codes when infinite common randomness is available between the encoder and the decoder, and an inner bound for limited common randomness. The proof is the subject of Sections IV and V. In Section VI, we show that, similarly to traditional source coding with side information, the two cases are equivalent when X and Z are jointly Gaussian.

II. Problem formulation and a first property

A. Notation

Calligraphic letters such as $\mathcal{X}$ denote sets, except in $p_{\mathcal{J}}^{\mathcal{U}}$, which denotes the uniform distribution over alphabet $\mathcal{J}$. We denote by $[a]$ the set $\{1, ..., \lfloor a \rfloor\}$ and by x^n the finite sequence $(x_1, ..., x_n)$. We denote by $\|p - q\|_{TV}$ the total variation distance between distributions p and q. We use $I(X;Y)$ to denote the mutual information X and Y, defined for general alphabets as in [29].

B. Definitions

In this section we formulate the information-theoretical problem for general alphabets. To have the existence of conditional distribution given a joint distribution, we assume all alphabets are Polish spaces. This includes discrete spaces and real vector spaces. We omit measure-theoretic justifications where reasonable.

Definition 1: Given an alphabet $\mathcal{X}$, a distortion measure is a function $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ extending to sequences as

$$d(x^n, y^n) = \tfrac{1}{n}\textstyle\sum_{i=1}^n d(x_i, y_i).$$

As shown in Figure 1, we consider two cases: side information available at the decoder only or available at both the encoder and the decoder.

Definition 2: Given two measurable spaces with respective *source* alphabet $\mathcal{X}$ and *side-information* alphabet $\mathcal{Z}$, a (n, R, R_c) D-code (resp. E-D-code) is a privately randomized encoder and decoder couple $(F^{(n)}, G^{(n)})$ consisting of a mapping $F^{(n)}_{M|X^n,J}$ (resp. $F^{(n)}_{M|X^n,Z^n,J}$) from $\mathcal{X}^n \times [2^{nR_c}]$ (resp. $\mathcal{X}^n \times \mathcal{Z}^n \times [2^{nR_c}]$) to $[2^{nR}]$ and a mapping $G^{(n)}_{Y^n|Z^n,J,M}$ from $\mathcal{Z}^n \times [2^{nR_c}] \times [2^{nR}]$ to $\mathcal{X}^n$.

We choose the total variation distance as similarity measure, which allows us to use existing proof techniques for general alphabets, including the soft covering lemma.

Definition 3: Given source and side-information alphabets $\mathcal{X}$ and $\mathcal{Z}$ and a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$, a triplet (R, R_c, Δ) is said to be D-achievable (resp. E-D-achievable) with *near-perfect realism* if there exists a sequence of (n, R, R_c) D-codes (resp. E-D-codes) $(F^{(n)}, G^{(n)})_n$ such that

$$\limsup_{n\to\infty} \mathbb{E}_P[d(X^n, Y^n)] \leq \Delta \quad \text{and} \tag{1}$$

$$\|P_{Y^n} - p_X^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0, \quad \text{where} \tag{2}$$

$$P_{X^n,Z^n,J,M,Y^n} = p_{X,Z}^{\otimes n} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \cdot F^{(n)}_{M|X^n,Z^n,J} \cdot G^{(n)}_{Y^|Z^n,J,M}.$$

If in addition there exists an integer N such that for all $n \geq N$, $P_{Y^n} \equiv p_X^{\otimes n}$, then (R, R_c, Δ) is said to be D-achievable (resp. E-D-achievable) with *perfect realism.*

The following notion, appearing in [26], is key to handling general alphabets and is satisfied by finite ones and by the normal distribution with squared-error distortion (Section VI).

Definition 4: Given a probability space $(\Omega, \mathcal{B}, \mathbb{P})$, an alphabet $\mathcal{X}$, a probability distribution p on $\mathcal{X}$ and a distortion measure d, we say that (d, p) is uniformly integrable iff for every $\varepsilon > 0$ there is a $\tau > 0$ such that

$$\sup_{X,Y,B} \mathbb{E}[d(X, Y)\mathbf{1}_B] \leq \varepsilon,$$

where X and Y represent all variables on $(\Omega, \mathcal{B})$ with law $\mathbb{P}_X \equiv \mathbb{P}_Y \equiv p$ and B represents all events in $\mathcal{B}$ s.t. $\mathbb{P}(B) \leq \tau$.

C. Equivalence of the perfect / near-perfect realism problems

To prove inner and outer bounds we use (Sections IV, V) the uniform integrability in conjunction with the following results.

Theorem 5: Given Polish source and side-information alphabets $\mathcal{X}$ and $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d on $\mathcal{X}$ such that (d, p_X) is uniformly integrable, then a triplet (R, R_c, Δ) is D-achievable (resp. E-D-achievable) with near-perfect realism if and only if it is D-achievable (resp. E-D-achievable) with perfect realism.

Remark 6: In Section V, we use a slightly stronger result: achievability with perfect realism is implied by the existence of codes satisfying (2) and at vanishing total variation distance from a sequence of distributions satisfying (1).

The proof of Theorem 5 and its variant in Remark 6 is the same as that of [26] in the absence of side information.

III. Main result

Our main result is the full characterization of the region of D-achievable triplets assuming infinite common randomness. We also prove an inner bound for finite common randomness.

Theorem 7: Consider Polish source and side-information alphabets $\mathcal{X}$ and $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d such that (d, p_X) is uniformly integrable. Then, assuming infinite common randomness, the closure of

the set $\mathcal{A}_{D,\infty}$ of D-achievable tuples with perfect or near-perfect realism is the closure of the following set $\mathcal{S}_{D,\infty}$:

$$\left\{ \begin{array}{rcl} (R,\Delta) \in \mathbb{R}^2_{\geq 0} & : & \exists\, p_{X,Z,V,Y} \in \mathcal{D}_D \text{ s.t.} \\ R & \geq & I_p(X;V|Z) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \quad (3)$$

where $\mathcal{D}_D$ is defined as

$$\left\{ \begin{array}{c} p_{X,Z,V,Y} : (X,Z) \sim p_{X,Z},\ p_Y \equiv p_X \\ Z - X - V, \quad X - (Z,V) - Y \\ I_p(Z;V) < \infty \end{array} \right\}, \quad (4)$$

where the alphabet of V is constrained to be Polish. Moreover, in the case of finite common randomness, the closure of the set $\mathcal{A}_D$ of D-achievable triplets (R, R_c, Δ) with perfect or near-perfect realism contains the closure of the following $\mathcal{S}_D$:

$$\left\{ \begin{array}{rcl} (R,R_c,\Delta) \in \mathbb{R}^3_{\geq 0} & : & \exists\, p_{X,Z,V,Y} \in \mathcal{D}_D \text{ s.t.} \\ R & \geq & I_p(X;V|Z) \\ R + R_c & \geq & I_p(Y;V) - I_p(Z;V) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}. \quad (5)$$

When the side information Z is independent from the source X, it is independent of (X,V) due to the first Markov chain property. Therefore $I(X;V|Z) = I(X;V)$ and the second Markov property implies $X - V - Y$. Thus, we recover the result of [24] where no side information is present.

Corollary 8: The same result applies for E-D-achievability, where $\mathcal{D}_D$ is replaced by $\mathcal{D}_{E\text{-}D}$, defined as

$$\left\{ \begin{array}{c} p_{X,Z,V,Y} : (X,Z) \sim p_{X,Z},\ p_Y \equiv p_X \\ X - (Z,V) - Y \\ I_p(Z;V) < \infty \end{array} \right\}. \quad (6)$$

Remark 9: The region of [26], taken with $R_c = \infty$, includes the translation of $\mathcal{S}_{E\text{-}D,\infty}$ by a rate $+I(X;Z)$.

Corollary 8 follows from applying Theorem 7 with source (X^n, Z^n) instead of X^n and choosing a suitable distortion measure.

We prove Theorem 7 in Sections IV and V.

IV. Converse

We prove that $\overline{\mathcal{A}}_{D,\infty} \subset \overline{\mathcal{S}}_{D,\infty}$ by proving that $\mathcal{A}_{D,\infty} \subset \overline{\mathcal{S}}_{D,\infty}$. This converse proof builds on the approach of [30], where the same Markov chains as in (4) are studied. Let (R,Δ) be D-achievable with near-perfect realism with infinite common randomness. Then, by Theorem 5, it is D-achievable with perfect realism. Fix $\varepsilon > 0$. Then there exists a (n, R, ∞) code inducing a joint distribution P such that $\mathbb{E}_P[d(X^n,Y^n)] < \Delta + \varepsilon$ and $P_{Y^n} \equiv p_X^{\otimes n}$. Let T be a uniform random variable over $[n]$. Define V=$(M, J, X_{T+1:n}, Z_{1:T-1}, T)$. For an i.i.d. vector such as $(X^n, Z^n) \sim p_{X,Z}^{\otimes n}$, the distribution of (X_T, Z_T) is $p_{X,Z}$ Similarly, we have $p_{Y_T} \equiv p_X$. Markov chains in $\mathcal{D}_D$ hold as proved in [30]. Moreover, we have

$$\begin{aligned} I(V;Z_T) &= I(M,J,X_{T+1:n},Z_{1:T-1};Z_T|T) \\ &= \tfrac{1}{n}\textstyle\sum_{t=1}^n I(M,J,X_{t+1:n},Z_{1:t-1};Z_t) \\ &= \tfrac{1}{n}\textstyle\sum_{t=1}^n I(M;Z_t|J,X_{t+1:n},Z_{1:t-1}) < \infty, \end{aligned}$$

where the first two equalities use the independence of T from all other variables and are true for discrete alphabets. A quantization argument based on [29] yields the same result for general alphabets. Since the product of Polish spaces is Polish, the alphabet of V is. Thus $p_{X_T,Y_T,Z_T,V} \in \mathcal{D}_D$. Using the independence of T from all other variables, we have

$$\mathbb{E}[d(X_T,Y_T)] = \sum_{t=1}^n \mathbb{E}[\mathbf{1}_{T=t} d(X_t,Y_t)] = \sum_{t=1}^n \mathbb{P}_T(t)\mathbb{E}[d(X_t,Y_t)].$$

Therefore, we have $\Delta + \varepsilon \geq \mathbb{E}[d(X_T,Y_T)]$. Moreover,

$$\begin{aligned} nR \geq H(M) &\geq I(M;X^n|Z^n,J) \\ &= I(M,J;X^n|Z^n) \quad (7) \\ &= \textstyle\sum_{t=1}^n I(M,J;X_t|Z^n,X_{t+1:n}) \\ &= \textstyle\sum_{t=1}^n I(M,J,X_{t+1:n},Z_{[n]\setminus t};X_t|Z_t) \\ &\geq \textstyle\sum_{t=1}^n I(M,J,X_{t+1:n},Z_{1:t-1};X_t|Z_t) \\ &= nI(M,J,X_{T+1:n},Z_{1:T-1};X_T|Z_T,T) \quad (8) \\ &= nI(V;X_T|Z_T), \quad (9) \end{aligned}$$

where (7) follows from the independence between the common randomness and the sources and equations (8) and (9) hold similarly to the above computation of $I(V;Z_T)$. Hence $(R,\Delta+\varepsilon) \in \mathcal{S}_{D,\infty}$, which concludes the proof.

V. Achievability

A. Informal outline

We introduce a virtual message M' with rate R' generated and used by the encoder, but not be transmitted. The decoder will then guess it. We start with a distribution $Q^{(1)}$, analyzed using the soft covering lemma of [27], then change it little by little to obtain intermediate distribution $Q^{(2)}$ and a distribution $P^{(1)}$ corresponding to a coding scheme. In order to obtain a final distortion bound, we use the uniform integrability in conjunction with the equivalence between achievability with near-perfect and perfect realism. For joint distributions, we use, with abuse of notation, $Q_{X,Y}(x,y) = Q_X(x)\rho(y|x)$, which defines $Q_{X,Y}$ by marginal distribution Q_X and conditional probability kernel ρ.

B. Random codebook indexed by a virtual message

Here, we prove that $\overline{\mathcal{S}}_D \subset \overline{\mathcal{A}}_D$ by proving that $\mathcal{S}_D \subset \overline{\mathcal{A}}_D$. This will also yield that $\mathcal{S}_{D,\infty}$ is contained in the projection -along the two coordinates R and Δ- of $\overline{\mathcal{A}}_D$, which is contained in $\overline{\mathcal{A}}_{D,\infty}$. Let (R,R_c,Δ) be a triplet in $\mathcal{S}_D$. Fix some $\varepsilon > 0$. Let $p_{X,Y,Z,V}$ be a corresponding distribution from the definition of $\mathcal{S}_D$. Then, we have $I_p(Z;V) < \infty$, and

$$R \geq I_p(X;V|Z) \quad (10)$$

$$R + R_c \geq I_p(Y;V) - I_p(Z;V) \quad (11)$$

$$\Delta \geq \mathbb{E}_p[d(X,Y)]. \quad (12)$$

By (4), we have

$$I_p(X;V|Z) = I_p(X;V) - I_p(Z;V). \quad (13)$$

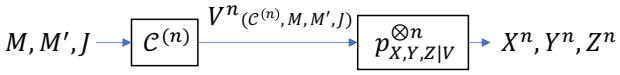

Fig. 2. Graphical model for $Q^{(1)}$.

We introduce a rate R' corresponding to a virtual message M', as follows. If $I_p(Z;V) = 0$ then set $R' = 0$. Otherwise, fix R' in $\big(I_p(Z;V) - \varepsilon, I_p(Z;V)\big) \cap \big(0, +\infty\big)$, which is possible since $I(Z;V) < \infty$. Then, by (10), (13), (11), we have:

$$(R+\varepsilon) + R' > I_p(X;V) = I_p(X,Z;V), \tag{14}$$
$$(R+\varepsilon) + R' + R_c > I_p(Y;V). \tag{15}$$

For every $n \geq 1$ we generate a random codebook $\mathcal{C}^{(n)}$ with $\lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR'} \rfloor \times \lfloor 2^{nR_c} \rfloor$ i.i.d. codewords sampled from $p_V^{\otimes n}$. The codewords are indexed by triples (m, m', j). We denote this random codebook distribution by $\mathbb{Q}_{\mathcal{C}^{(n)}}$.

C. Distribution $Q^{(1)}$ and soft-covering lemma

For every positive integer n, we define the distribution $Q^{(1)}$, described in Figure 2, by the following density:

$$Q^{(1)}\left(c^{(n)}, m, m', j, v^n, x^n, y^n, z^n\right) = \mathbb{Q}_{\mathcal{C}^{(n)}}(c^{(n)}) \cdot \frac{1}{\lfloor 2^{n(R+\varepsilon)} \rfloor \lfloor 2^{nR'} \rfloor \lfloor 2^{nR_c} \rfloor} \mathbf{1}_{v^n = v^n(c^{(n)}, m, m', j)} \prod_{t=1}^{n} p_{X,Y,Z|V=v_t}(x_t, y_t, z_t) \tag{16}$$

Then, by the definition of $\mathbb{Q}_{\mathcal{C}^{(n)}}$, we have $Q^{(1)}_{V^n} \equiv p_V^{\otimes n}$, and therefore, from (16) we have $Q^{(1)}_{X^n,Y^n,Z^n} \equiv p_{X,Y,Z}^{\otimes n}$. Hence, from (12) and the additivity of d, we have

$$\mathbb{E}_{Q^{(1)}}[d(X^n, Y^n)] \leq \Delta. \tag{17}$$

We can use the Markov property $Z - X - V$ to show that

$$(M, M') - (X^n, J) - Z^n \text{ under } Q^{(1)}. \tag{18}$$

1) Marginals of Y^n and (X^n, Z^n) knowing a codebook:

We start by stating a standard lemma regarding the total variation distance.

Lemma 10: Let Π be two distributions on the product of two Polish spaces $\mathcal{W}$ and $\mathcal{L}$, and let $\Pi_{L|W}, \Gamma_{L|W}$ be two channels. Then, we have

$$\|\Pi_W \Pi_{L|W} - \Pi_W \Gamma_{L|W}\|_{TV} = \mathbb{E}_{\Pi_W}\big[\|\Pi_{L|W} - \Gamma_{L|W}\|_{TV}\big].$$

We use the *soft covering lemma* for general alphabets [27, Corollary VII.4] in its memoryless case, and get, as in [27] :

$$\mathbb{E}_{\mathcal{C}^{(n)}}\big[\|Q^{(1)}_{Y^n|\mathcal{C}^{(n)}} - p_X^{\otimes n}\|_{TV}\big] \underset{n\to\infty}{\longrightarrow} 0 \tag{19}$$
$$\mathbb{E}_{\mathcal{C}^{(n)}}\big[\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]} p_{X,Z}^{\otimes n}\|_{TV}\big] \underset{n\to\infty}{\longrightarrow} 0. \tag{20}$$

2) Decoding of M':

The virtual message M' is to be decoded as $\hat{M}'$, conditionally independent from M', X^n, Y^n knowing $M, J, Z^n, \mathcal{C}^{(n)}$:

$$Q^{(1)}_{\hat{M}'|\mathcal{C}^{(n)}=c^{(n)}, M=m, J=j, Z^n=z^n}(\hat{m}')$$

is a joint $p_{V,Z}$-typicality decoder for subcodebook $(v^n(c^{(n)}, m, a, j))_a$. By the typical random coding argument for channel coding (see, e.g., [31]), we get

$$Q^{(1)}_{\hat{M}', M'}(\hat{M}' \neq M') \underset{n\to\infty}{\longrightarrow} 0 \tag{21}$$

in the case of finite alphabets. It can be proved that the same holds for general sources, using a joint typicality decoder with respect to a quantized distribution $p_{[V],[Z]}$ of mutual information close enough to $I_p(Z;V)$. The probability in (21) can be rewritten as an expectation over $\mathcal{C}^{(n)}$. Since the convergence towards zero in expectation implies a convergence in probability for a non-negative variable, we get:

$$Q^{(1)}\big(M' \neq \hat{M}' \big|\, \mathcal{C}^{(n)}\big) \underset{n\to\infty}{\overset{\mathcal{P}}{\longrightarrow}} 0. \tag{22}$$

3) Choosing a codebook:

From (17), Lemma 10, and the Markov inequality we get

$$\mathbb{Q}_{\mathcal{C}^{(n)}}\Big(\mathbb{E}[d(X^n, Y^n)|\mathcal{C}^{(n)}]\Big) \leq \Delta + \varepsilon) \geq \varepsilon/(\Delta + \varepsilon). \tag{23}$$

In addition, similarly to (22), we get convergence in probability from (19) and (20). Combining this with (23) gives that for a certain N_0, $\forall n \geq N_0$ there is a codebook $c_*^{(n)}$ such that

$$\mathbb{E}_{Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}=c_*^{(n)}}}[d(X^n, Y^n)] \leq \Delta + \varepsilon, \tag{24}$$
$$\|Q^{(1)}_{Y^n|\mathcal{C}^{(n)}=c_*^{(n)}} - p_X^{\otimes n}\|_{TV} \leq \varepsilon. \tag{25}$$
$$\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}=c_*^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]} p_{X,Z}^{\otimes n}\|_{TV} \leq \varepsilon. \tag{26}$$
$$Q^{(1)}_{M', \hat{M}'|\mathcal{C}^{(n)}=c_*^{(n)}}(\hat{M}' \neq M') \leq \varepsilon. \tag{27}$$

In the following sections we shall omit the conditioning on $\mathcal{C}^{(n)} = c_*^{(n)}$, which will be implicit.

D. Construction of a code

In this subsection, we show how distribution $Q^{(1)}$ can be modified, in a way that is minor in terms of total variation distance, to lead to a code. The latter inherits the realism for Y^n and the small error for decoding M'. Regarding expected distortion, we use the uniform integrability in conjunction with the equivalence between near-perfect and perfect realism.

1) Some lemmas on the total variation distance:

We start by citing some lemmas from [27] and [28].

Lemma 11: [27, Lemma V.1] Let Π and Γ be two distributions on an alphabet $\mathcal{W} \times \mathcal{L}$. Then

$$\|\Pi_W - \Gamma_W\|_{TV} \leq \|\Pi_{W,L} - \Gamma_{W,L}\|_{TV}.$$

Lemma 12: [27, Lemma V.2] Let Π and Γ be two distributions on an alphabet $\mathcal{W} \times \mathcal{L}$. Then when using the same channel $\Pi_{L|W}$ we have

$$\|\Pi_W \Pi_{L|W} - \Gamma_W \Pi_{L|W}\|_{TV} = \|\Pi_W - \Gamma_W\|_{TV}.$$

Lemma 13: [28, Lemma 2] Let P_{UWL} be a distribution on an alphabet of the form $\mathcal{U} \times \mathcal{U} \times \mathcal{L}$ and let $\eta \in (0,1)$. If $P(U \neq W) \leq \eta$ we have $\|P_{UL} - P_{WL}\|_{TV} \leq \eta$.

2) Construction of $Q^{(2)}$ and comparison to $Q^{(1)}$:

Using definition (16) of $Q^{(1)}$ and the Markov property $X -$

$Z, V-Y$ from (4) we get:

$$Q^{(1)}_{X^n,Y^n,Z^n|V^n=v^n}(x^n,y^n,z^n) = \prod_{t=1}^{n} p_{X,Z|V=v_t}(x_t,z_t)\, p_{Y|V=v_t,\,Z=z_t}(y_t).$$

With this in mind, for every positive integer n we define the following distribution which differs from $Q^{(1)}$ in that Y^n is sampled using $\hat{M}'$ instead of M' :

$$\begin{aligned} &Q^{(2)}(m,m',j,v^n,x^n,y^n,z^n,\hat{m}') \\ &\quad = Q^{(1)}(m,m',j,v^n,x^n,z^n,\hat{m}')\, p_{Y^n|Z^n=z^n,V^n=v^n}(y^n)(m,\hat{m}',j) \end{aligned}$$

By construction, we have

$$Q^{(2)}_{M',\hat{M}',M,J,V^n,X^n,Z^n} \equiv Q^{(1)}_{M',\hat{M}',M,J,V^n,X^n,Z^n}.$$

Using Lemma 13 on this joint distribution with $U=M', W=\hat{M}', L=(M,J,V^n,X^n,Z^n)$ we have by (27)

$$\left\|Q^{(1)}_{M',M,J,V^n,X^n,Z^n} - Q^{(2)}_{\hat{M}',M,J,V^n,X^n,Z^n}\right\|_{TV} \le \varepsilon.$$

As a consequence and by construction of $Q^{(2)}$ and its similarity to that of $Q^{(1)}$ we have by Lemma 12 with $L=Y^n$:

$$\left\|Q^{(1)}_{M',M,J,V^n,X^n,Y^n,Z^n} - Q^{(2)}_{\hat{M}',M,J,V^n,X^n,Y^n,Z^n}\right\|_{TV} \le \varepsilon. \quad (28)$$

Since $Q^{(2)}_{J,X^n,Z^n} \equiv Q^{(1)}_{J,X^n,Z^n}$ we also have by (26):

$$\|Q^{(2)}_{J,X^n,Z^n} - p^{\mathcal{U}}_{[2^{nR_c}]} p^{\otimes n}_{X,Z}\|_{TV} \le \varepsilon. \quad (29)$$

3) Finalizing the code construction:
We define the distribution $P^{(1)}$ achieving near-perfect realism and from which a distribution $P^{(2)}$ with perfect realism will be derived. The former differs from $Q^{(2)}$ in having the correct marginal for (X^n,Z^n) as follows

$$\begin{aligned} &P^{(1)}(m,m',j,v^n,x^n,y^n,z^n,\hat{m}') \\ &= \tfrac{1}{\lfloor 2^{nR_c}\rfloor} \prod_{t=1}^{n} p_{X,Z}(x_t,z_t)\; Q^{(2)}_{M,M',V^n,\hat{M}',Y^n|J=j,X^n=x^n,Z^n=z^n}(m,m',v^n,\hat{m}',y^n) \end{aligned}$$

Then by Lemma 12 comparing $P^{(1)}$ with $Q^{(2)}$ reduces to comparing marginals, i.e. to equation (29):

$$\begin{aligned} &\left\|P^{(1)}_{M,M',J,V^n,X^n,Z^n,\hat{M}',Y^n} - Q^{(2)}_{M,M',J,V^n,X^n,Z^n,\hat{M}',Y^n}\right\|_{TV} \\ &\quad = \left\|P^{(1)}_{J,X^n,Z^n} - Q^{(2)}_{J,X^n,Z^n}\right\|_{TV} \le \varepsilon. \end{aligned} \quad (30)$$

Therefore by Lemma 11 with $W=(X^n,Y^n)$ and the triangle inequality and (28) we get

$$\left\|P^{(1)}_{X^n,Y^n} - Q^{(1)}_{X^n,Y^n}\right\|_{TV} \le \varepsilon. \quad (31)$$

Due to (18), it can be easily checked that $P^{(1)}$ defines a $(n, R+\varepsilon, R_c)$ D-code. The entire construction layed out in this Section V is valid for any $\varepsilon>0$ (which was fixed in Section b). The blocklength N_0 after which (24) and (25) hold depends on $\varepsilon>0$. Consider a sequence of codes associated to some vanishing sequence $(\varepsilon_k)_k$. For any $\varepsilon'>0$, the achievability of $(R+\varepsilon', R_c, \Delta+\varepsilon')$ then follows from Remark 9 with $Q^{(1)}$ and $P^{(1)}$ and from (24), (25) and (31). Hence, $(R,R_c,\Delta)\in\overline{\mathcal{A}}_D$ as desired.

VI. The Gaussian case

Interestingly, similarly to standard source coding with side information, when the source and side information form a bi-dimensional Gaussian then D-achievability is equivalent to E-D-achievability in the following sense.

Theorem 14: Consider the setting of Theorem 7 in the case of infinite common randomness with $d:(x,y)\mapsto(x-y)^2$

$$\text{and } p_{X,Z} = \mathcal{N}\Big(0, \begin{pmatrix} 1 & \eta \\ \eta & 1 \end{pmatrix}\Big).$$

For any Δ in $(0, 2-2\eta]$, and denoting $\rho = 1-\Delta/2$, the infimum of rates such that (R,Δ) is D- or E-D-achievable with (near-)perfect realism is:

$$R_D(\Delta) = R_{E\text{-}D}(\Delta) = \frac{1}{2}\log\Big(\frac{1-\eta^2}{1-\rho^2}\Big), \quad (32)$$

The numerator is the same as in standard source coding with side information (see e.g. [31]) and the denominator is the same as in [16], which we recover when the side-information Z is independent from X. We now prove Theorem 14.
We know that $R_D(\Delta) \ge R_{E\text{-}D}(\Delta)$. Moreover, by Remark 9 the region $\mathcal{S}_{E\text{-}D,\infty}$ translated by a rate $+I(X;Z) = -\frac{1}{2}\log((1-\eta^2))$ is included in the region of [16], [24]. Therefore, by [16, Proposition 2] we have

$$R_{E\text{-}D}(\Delta) \ge \tfrac{1}{2}\log((1-\eta^2)/(1-\rho^2)). \quad (33)$$

We upper bound $R_D(\Delta)$. Fix a Δ in $(0, 2-2\eta]$. Let

$$(Z,X,V) = \big(\eta\tilde{X} + \sqrt{1-\eta^2}\tilde{Z}, \tilde{X}, b\tilde{X} + \sqrt{1-b^2}\tilde{V}\big), \quad (34)$$

where $(\tilde{Z},\tilde{X},\tilde{V})$ is standard Gaussian. Since $\rho\ge\eta$, with

$$b = \sqrt{(\rho^2-\eta^2)/(1+\eta^2\rho^2-2\eta^2)} \text{ we find} \quad (35)$$

$$\mathbb{E}\big[\mathbb{E}[X|Z,V]^2\big] = \rho^2, \text{ and therefore} \quad (36)$$

$$I_p(X;V|Z) = h(X|Z) - h(X|Z,V) = \tfrac{1}{2}\log((1\text{-}\eta^2)/(1\text{-}\rho^2)). \quad (37)$$

Define $Y = \rho^{-1}\mathbb{E}[X|Z,V]$. One can check that $p_{X,Z,V,Z}\in\mathcal{D}_D$, that $\mathbb{E}[d(X,Y)]=\Delta$ and (d,p_X) is uniformly integrable, then by Theorem 7, we have

$$R_D(\Delta) \le \tfrac{1}{2}\log((1-\eta^2)/(1-\rho^2)).$$

VII. Conclusion

We have considered the traditional problem of source coding of a memoryless source X^n in the presence of correlated side information Z^n, studied by Wyner and Ziv, with the additional requirement of perfect realism on the reconstruction. We have characterized the rate-distortion-perception trade-off for sources on general alphabets when infinite common randomness is available between the encoder and the decoder, in two cases: when Z^n is available only at the decoder or at both the encoder and the decoder. We showed that, similarly to traditional source coding with side information, the two cases are equivalent when X^n and Z^n are jointly Gaussian. We also provided a general inner bound in the case of limited common randomness.

REFERENCES

[1] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[2] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," in *International Conference on Learning Representations, 2017*.

[3] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, "Conditional probability models for deep image compression," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018*, pp. 4394–4402.

[4] D. Minnen and S. Singh, "Channel-wise autoregressive entropy models for learned image compression," in *IEEE International Conference on Image Processing, 2020*, pp. 3339–3343.

[5] S. Kankanahalli, "End-to-end optimized speech coding with deep neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2018*, pp. 2521–2525.

[6] W. B. Kleijn, F. S. C. Lim, A. Luebs, J. Skoglund, F. Stimberg, Q. Wang, and T. C. Walters, "Wavenet based low rate speech coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2018*, pp. 676–680.

[7] D. Petermann, S. Beack, and M. Kim, "Harp-net: Hyper-autoencoded reconstruction propagation for scalable neural audio coding," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2021*, pp. 316–320.

[8] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi, "Soundstream: An end-to-end neural audio codec," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 495–507, 2022.

[9] C.-Y. Wu, N. Singhal, and P. Krähenbühl, "Video compression through image interpolation," in *ECCV, 2018*, pp. 425–440.

[10] O. Rippel, S. Nair, C. Lew, S. Branson, A. Anderson, and L. Bourdev, "Learned video compression," in *IEEE/CVF International Conference on Computer Vision, 2019*, pp. 3453–3462.

[11] A. Djelouah, J. Campos, S. Schaub-Meyer, and C. Schroers, "Neural inter-frame compression for video coding," in *IEEE/CVF International Conference on Computer Vision, 2019*, pp. 6420–6428.

[12] A. Habibian, T. V. Rozendaal, J. Tomczak, and T. Cohen, "Video compression with rate-distortion autoencoders," in *IEEE/CVF International Conference on Computer Vision, 2019*, pp. 7032–7041.

[13] Z. Hu, G. Lu, and D. Xu, "Fvc: A new framework towards deep video compression in feature space," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021*, pp. 1502–1511.

[14] B. Liu, Y. Chen, S. Liu, and H. S. Kim, "Deep learning in latent space for video prediction and compression," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021*, pp. 701–710.

[15] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, "Generative adversarial networks for extreme learned image compression," in *IEEE/CVF International Conference on Computer Vision, 2019*, pp. 221–231.

[16] M. Li, J. Klejsa, and W. Kleijn, "On distribution preserving quantization," August 2011, arXiv:1108.3728.

[17] S. Santurkar, D. Budden, and N. Shavit, "Generative compression," in *Picture Coding Symposium, 2018*, pp. 258–262.

[18] M. Tschannen, E. Agustsson, and M. Lucic, "Deep generative models for distribution-preserving lossy compression," in *NeurIPS, 2018*, vol. 31.

[19] R. Matsumoto, "Introducing the perception-distortion tradeoff into the rate-distortion theory of general information sources," *IEICE Communications Express*, vol. 7, no. 11, pp. 427–431, 2018.

[20] ——, "Rate-distortion-perception tradeoff of variable-length source coding for general information sources," *IEICE Communications Express*, vol. 8, no. 2, pp. 38–42, 2019.

[21] Y. Blau and T. Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 675–685.

[22] J. Chen, L. Yu, J. Wang, W. Shi, Y. Ge, and W. Tong, "On the rate-distortion-perception function," *IEEE Journal on Selected Areas in Information Theory*, pp. 1–1, 2022.

[23] L. Theis and E. Agustsson, "On the advantages of stochastic encoders," in *Neural Compression: From Information Theory to Applications – Workshop @ ICLR*, 2021.

[24] L. Theis and A. B. Wagner, "A coding theorem for the rate-distortion-perception function," in *Neural Compression: From Information Theory to Applications – Workshop @ ICLR 2021*.

[25] N. Saldi, T. Linder, and S. Yüksel, "Output constrained lossy source coding with limited common randomness," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4984–4998, 2015.

[26] A. B. Wagner, "The rate-distortion-perception tradeoff: The role of common randomness," February 2022, arXiv:2202.04147.

[27] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[28] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1836–1849, 2016.

[29] A. D. Wyner, "A definition of conditional mutual information for arbitrary ensembles," *Information and Control*, vol. 38, no. 1, pp. 51–59, 1978.

[30] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," in *IEEE International Symposium on Information Theory, 2012*, pp. 3053–3057.

[31] T. Cover and J. Thomas, *Elements of Information Theory*, ser. Wiley-Interscience. Wiley, 2006.