Rate-Distortion-Perception Tradeoff for Lossy Compression Using Conditional Perception Measure

Sadaf Salehkalaibar[1], Jun Chen[2], Ashish Khisti[1] and Wei Yu[1]
[1]*ECE Department, University of Toronto, Toronto, Canada*, emails: {sadafs, akhisti, weiyu}@ece.utoronto.ca
[2]*ECE Department, McMaster University, Hamilton, Canada*, email: chenjun@mcmaster.ca

***Abstract*—This paper studies the rate-distortion-perception (RDP) tradeoff for a memoryless source model in the asymptotic limit of large block-lengths. The perception measure is based on a divergence between the distributions of the source and reconstruction sequences *conditioned* on the encoder output, first proposed by Mentzer et al. We consider the case when there is no shared randomness between the encoder and the decoder. For the case of discrete memoryless sources we derive a single-letter characterization of the RDP function, in contrast to the marginal-distribution metric case (introduced by Blau and Michaeli), whose RDP characterization remains open when there is no shared randomness. The achievability scheme is based on lossy source coding with a posterior reference map. For the case of continuous valued sources under squared error distortion measure and squared quadratic Wasserstein perception measure we also derive a single-letter characterization and show that a noise-adding mechanism at the decoder suffices to achieve the optimal representation. Interestingly, the RDP function characterized for the case of zero perception loss coincides with that of the marginal metric and further zero perception loss can be achieved with a 3-dB penalty in minimum distortion. Finally we specialize to the case of Gaussian sources, and derive the RDP function for Gaussian vector case and propose a waterfilling like solution. We also partially characterize the RDP function for a mixture of Gaussian vector sources.**

I. Introduction

Rate-distortion-perception (RDP) tradeoff [1], a generalization of the classical rate-distortion function [2] to incorporate distribution constraints on the reconstruction, provides a theoretical framework for a variety of deep neural compression systems that exhibit an inherent tradeoff between reconstruction fidelity and realism [3]. In this framework, the perception loss is measured through a suitable divergence metric between the source and reconstruction distributions, with perfect realism corresponding to the case when the source and reconstruction distributions are identical. The work of Blau and Michaeli [1] establishes that when distortion loss is measured using mean squared error, perfect realism can be achieved with no more than 3-dB increase in the minimum distortion. The work of Theis and Wagner [4] establishes an operational interpretation of the rate-distortion-perception function. The special case of (scalar) Gaussian sources is studied in [5] where it is shown that Gaussian distributions attain the RDP function. Furthermore a natural notion of universality is established where any representation corresponding to a boundary point on RDP tradeoff curve can be converted to another representation associated with another boundary point. The case when there is limited or no shared randomness between the encoder and decoder is studied in [6]–[9] (see also [10]). To our knowledge, unlike the setting with (unlimited) shared randomness, a computable characterization of RDP function remains largely open in the limited shared randomness settings. The extension of RDP function to the case when correlated side information is available to either the encoder or the decoder is studied in [11], [12]. The applications of RDP function to neural compression are studied in e.g., [13]–[20] and references therein.

While the perception loss metric in prior works [3] is based on the divergence between the source and reconstruction distributions, a different choice is proposed in [13], [18], where it is empirically observed that a perception loss metric that measures the divergence between the source and reconstruction distributions *conditioned* on the output of the encoder results in higher perceptual quality in reconstructions. Intuitively, this metric forces the decoder to follow the conditional distribution of the source given, the reconstruction based on the encoder output (e.g., the MMSE reconstruction) and thus introduces adjustments in the fine details that improve the blurriness, while not deviating significantly from the reconstruction. In this work, we provide a theoretical study of the RDP function when the perception measure is based on such a conditional metric[1]. We make the assumption that there is no shared randomness between the encoder and decoder, and denote this setting as *conditional-distribution based perception measure*, while denoting the original setting of [3] as *marginal-distribution based perception measure*. The main contributions of this paper are as follows:

- We characterize the RDP tradeoff for finite alphabet sources (Theorem 1) and explicitly derive the tradeoff for the uniform Bernoulli source (Theorem 2). The achievable scheme uses some recent tools developed for lossy source coding with a posterior reference map [22]. It is interesting to note that a complete characterization of the RDP function for the conditional-distribution based perception measure is possible, while a similar characterization of the RDP function for the marginal-distribution based perception measure only exists for the case of zero perception loss (i.e., when the source and reconstruction distributions exactly match) [8], if there is no shared randomness.

[1]An earlier investigation of this setting from an information-theoretic perspective was reported in [21]. In this paper, we provide a more comprehensive and rigorous treatment.

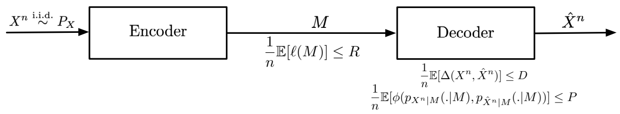

Fig. 1. System model with the perception measure based on the conditional distribution.

- The RDP tradeoff is further characterized for continuous alphabet sources (Theorem 3) under squared error distortion measure and squared quadratic Wasserstein perception measure. For the case of zero perception loss, it is shown that the rate-distortion tradeoff coincides with that of the marginal-distribution based perception measure (Corollary 1). As has been previously observed for the latter measure [9], [23], [24], for a fixed encoder, the distortion of reconstruction satisfying zero perception loss is exactly twice that of the MMSE representation. Furthermore, the MMSE represenation can be transformed to representations for other operating points on the RDP tradeoff by adding noise at the decoder.
- Shannon's lower bound [2, Eq. (13.159)] for the RD tradeoff is extended to the RDP setting (Theorem 4). Using this lower bound, we are able to partially characterize the RDP tradeoff for a Gaussian-mixture source (Corollary 2) and completely characterize the tradeoff for a Gaussian vector source (Corollary 3). For the latter case, a water-filling type solution is also derived.

Notation: We use $[a:b]$ to represent the set of integers from a to b for any two integers $a \leq b$, and use $x \wedge y$ to represent the minimum of two real numbers x and y . Throughout this paper, the base of the logarithm function is e.

II. Problem Formulation

A. System Model

Let the source $\{X(t)\}_{t=1}^{\infty}$ be a stationary and memoryless process with marginal distribution p_X over alphabet $\mathcal{X}$ (see Fig. 1). A stochastic encoder $f^{(n)}: \mathcal{X}^n \to \mathcal{M}$ maps a length-n source sequence X^n to a codeword M in a binary prefix code $\mathcal{M}$ according to some conditional distribution $p_{M|X^n}$. A stochastic decoder $g^{(n)}: \mathcal{M} \to \mathcal{X}^n$ then generates a length-n reconstruction sequence $\hat{X}^n$ based on M according to some conditional distribution $p_{\hat{X}^n|M}$. Note that this coding system induces the following joint distribution:

$$p_{X^n M \hat{X}^n} = p_{X^n} p_{M|X^n} p_{\hat{X}^n|M}.$$

Let $\Delta: \mathcal{X} \times \mathcal{X} \to [0, \infty)$ be a distortion measure with $\Delta(x, \hat{x}) = 0$ if and only if $x = \hat{x}$. Define $\Delta(x^n, \hat{x}^n) := \sum_{t=1}^{n} \Delta(x(t), \hat{x}(t))$ for $x^n, \hat{x}^n \in \mathcal{X}^n$. Let $\phi: \mathcal{P} \times \mathcal{P} \to [0, \infty]$ be a divergence with $\phi(p, p') = 0$ if and only if $p = p'$ a.s., where $\mathcal{P}$ denotes the set of probability distributions. Throughout this paper, we focus on a special class of divergences that arise from the theory of optimal transport. Specifically, for any two probability distributions $p_{\tilde{X}^n}$ and $p_{\bar{X}^n}$ over $\mathcal{X}^n$, let

$$\phi(p_{\tilde{X}^n}, p_{\bar{X}^n}) := \inf_{p_{\tilde{X}^n \bar{X}^n} \in \Pi(p_{\tilde{X}^n}, p_{\bar{X}^n})} \sum_{t=1}^{n} \mathbb{E}[c(\tilde{X}(t), \bar{X}(t))], \tag{1}$$

where $\Pi(p_{\tilde{X}^n}, p_{\bar{X}^n})$ denotes the set of couplings of $p_{\tilde{X}^n}$ and $p_{\bar{X}^n}$, and $c: \mathcal{X} \times \mathcal{X} \to [0, \infty)$ is a cost function with $c(\tilde{x}, \bar{x}) = 0$ if and only if $\tilde{x} = \bar{x}$. In this work, ϕ serves the role of perception measure.

Proposition 1: ϕ defined in (1) has the following properties.

(a) Tensorizability:

$$\phi(p_{\tilde{X}^n}, p_{\bar{X}^n}) \geq \sum_{t=1}^{n} \phi(p_{\tilde{X}(t)}, p_{\bar{X}(t)})$$

and the equality holds if $p_{\tilde{X}^n} = \prod_{i=1}^{n} p_{\tilde{X}(t)}$ and $p_{\bar{X}^n} = \prod_{i=1}^{n} p_{\bar{X}(t)}$.

(b) Convexity:

$$\begin{aligned}&\phi((1-\lambda)p_{\tilde{X}^n} + \lambda p'_{\tilde{X}^n}, (1-\lambda)p_{\bar{X}^n} + \lambda p'_{\bar{X}^n}) \\ &\qquad \leq (1-\lambda)\phi(p_{\tilde{X}^n}, p_{\bar{X}^n}) + \lambda \phi(p'_{\tilde{X}^n}, p'_{\bar{X}^n})\end{aligned}$$

for $\lambda \in [0, 1]$.

(c) Continuity:

$$\begin{aligned}&|\phi(p_{\tilde{X}^n}, p_{\bar{X}^n}) - \phi(p_{\tilde{Y}^n}, p_{\bar{Y}^n})| \\ &\qquad \leq n c_{\max}(d_{\mathrm{TV}}(p_{\tilde{X}^n}, p_{\tilde{Y}^n}) + d_{\mathrm{TV}}(p_{\bar{X}^n}, p_{\bar{Y}^n})),\end{aligned}$$

where d_{TV} is the total variation distance, and $c_{\max} := \sup_{x, x' \in \mathcal{X}} c(x, x')$.

Proof: See [25, Appendix A]. ■

B. Rate-Distortion-Perception Function

Definition 1: We say rate R is achievable subject to distortion and percetion constraints D and P if for some n, there exist encoder $f^{(n)}$ and decoder $g^{(n)}$ such that (see Fig. 1)

$$\begin{aligned}&\frac{1}{n}\mathbb{E}[\ell(M)] \leq R, \\ &\frac{1}{n}\mathbb{E}[\Delta(X^n, \hat{X}^n)] \leq D, \\ &\frac{1}{n}\mathbb{E}[\phi(p_{X^n|M}(\cdot|M), p_{\hat{X}^n|M}(\cdot|M))] \leq P,\end{aligned} \tag{2}$$

where $\ell(M)$ denotes the length of M. The infimum of all such R is denoted by $R_{\mathrm{C}}(D, P)$, is referred to as the operational rate-distortion-perception function for the conditional-distribution based perception measure.

We also introduce the following informational rate-distortion-perception function:

$$R(D, P) := \inf_{p_{U\hat{X}|X}} I(X; U)$$

$$\text{s.t. } X \leftrightarrow U \leftrightarrow \hat{X} \text{ form a Markov chain}, \tag{3}$$

$$\mathbb{E}[\Delta(X, \hat{X})] \leq D, \tag{4}$$

$$\mathbb{E}[\phi(p_{X|U}(\cdot|U), p_{\hat{X}|U}(\cdot|U))] \leq P. \tag{5}$$

Proposition 2: $R(D,P)$ is convex in (D,P).

Proof: See [25, Appendix B]. ■

Proposition 3: If $|\mathcal{X}| < \infty$, then there is no loss of generality in assuming that the alphabet of U, denoted by $\mathcal{U}$, satisfies $|\mathcal{U}| \leq |\mathcal{X}|+2$; moreover, the infimum in the definition of $R(D,P)$ can be attained, thus is a minimum.

Proof: Note that $|\mathcal{X}| < \infty$ implies $c_{\max} < \infty$, which in light of part 3) of Proposition 1 further implies the continuity of $\phi(p,p')$ in (p,p'). Therefore, we can invoke the support lemma [26, p. 631] to establish the desired cardinality bound. Moreover, the continuity of $\phi(p,p')$, together with the cardinality bound, implies that the feasible domain for $p_{U\hat{X}|X}$ is compact. As a consequence, the objective function $I(X;U)$, which is continuous in $p_{U\hat{X}|X}$, has a minimum value over this domain. ■

III. Finite Alphabet Sources

We focus on finite alphabet sources in this section. Our first main result shows that the operational rate-distortion-perception function coincides with its informational counterpart for such sources.

Theorem 1: Assume $|\mathcal{X}| < \infty$. For $D \geq 0$ and $P \geq 0$,

$$R_{\mathrm{C}}(D,P) = R(D,P).$$

Proof: See [25, Appendix C]. The achievability part of the proof relies on a recent development in information theory known as lossy source coding with a posterior reference map [22]. ■

Remark 1: The proof actually indicates that if (2) is replaced with the following stronger constraint

$$\frac{1}{n}\phi(p_{X^n|M}(\cdot|m), p_{\hat{X}^n|M}(\cdot|m)) \leq P, \quad m \in \mathcal{M},$$

Theorem 1 continues to hold.

The next result provides an explicit characterization of $R_{\mathrm{C}}(D,P)$ for the uniform Bernoulli source (i.e., $X \sim \mathrm{Ber}(\frac{1}{2})$) under Hamming distortion measure (i.e., $\Delta(x,\hat{x}) = d_H(x,\hat{x})$) and divergence induced by Hamming cost function (i.e., $c(x,\hat{x}) = d_H(x,\hat{x})$). For $D \geq 0$ and $P \geq 0$, let

$$\hbar(D,P) = \begin{cases} H_b\left(\frac{1+(D\wedge P)-\sqrt{1+(D\wedge P)^2-2D}}{2}\right), & D \in [0,\frac{1}{2}), \\ \log 2, & D \in [\frac{1}{2},\infty), \end{cases}$$

where H_b denotes the binary entropy function. Moreover, let $\overline{\hbar}$ be the upper concave envelope of $\hbar$ over $[0,\infty)^2$.

Theorem 2: Assume $X \sim \mathrm{Ber}(\frac{1}{2})$, $\Delta(x,\hat{x}) = d_H(x,\hat{x})$, and $c(x,\hat{x}) = d_H(x,\hat{x})$. For $D \geq 0$ and $P \geq 0$,

$$R_{\mathrm{C}}(D,P) = \log 2 - \overline{\hbar}(D,P).$$

Proof: See [25, Appendix D]. ■

Remark 2: The upper concave envelope operation is necessary as $\hbar$ itself is not concave in (D,P). See [25, Appendix E] for some relevant analysis.

Remark 3: Theorem 2 implies that for $D \geq 0$ and $P \geq D$,

$$\begin{aligned} R_{\mathrm{C}}(D,P) &\leq \log 2 - \hbar(D,P) \\ &= \begin{cases} \log 2 - H_b(D), & D \in [0,\frac{1}{2}), \\ 0, & D \in [\frac{1}{2},\infty). \end{cases} \end{aligned}$$

This upper bound is tight because it coincides with the rate-distortion function of the uniform Bernoulli source under Hamming distortion measure [2, Theorem 13.3.1], which is the infimum of achievable rates when the perception constraint is ignored.

The operational RDP function for the marginal-distribution based perception measure, denoted as $R_{\mathrm{M}}(D,P)$, can be defined similarly by replacing (2) with

$$\frac{1}{n}\phi(p_{X^n}, p_{\hat{X}^n}) \leq P. \tag{6}$$

It follows by part 2) of Proposition 1 that (2) implies (6). As a consequence, we must have

$$R_{\mathrm{M}}(D,P) \leq R_{\mathrm{C}}(D,P). \tag{7}$$

Different from $R_{\mathrm{C}}(D,P)$, a single-letter characterization of $R_{\mathrm{M}}(D,P)$ is unavailable except for the special case $P=0$, for which it is known [10, Section III.B], [8, Corollary 1], [9, Eq. (16)] that

$$R_{\mathrm{M}}(D,0) = \inf_{p_{U\hat{X}|X}} \max\{I(X;U), I(\hat{X};U)\} \tag{8}$$

$$\text{s.t. } X \leftrightarrow U \leftrightarrow \hat{X} \text{ form a Markov chain}, \tag{9}$$

$$\mathbb{E}[\Delta(X,\hat{X})] \leq D, \tag{10}$$

$$p_{\hat{X}} = p_X. \tag{11}$$

The difficulty in characterizing $R_{\mathrm{M}}(D,P)$ arises from the fact that the i.i.d. form of the reconstruction sequence $\hat{X}^n$ favored by the perception constraint (6) (see part (a) and part (b) of Proposition 1) is not necessarily desirable from the rate perspective. This tension disappears when $P=0$ as $\hat{X}^n$ is forced to be an i.i.d. sequence. In contrast, under constraint (2), the conditional i.i.d. form of the source sequence X^n and the reconstruction sequence $\hat{X}^n$ given the codeword M is desirable from both the perception and the rate perspectives. This explains why $R_{\mathrm{C}}(D,P)$ is more amenable to single-letterization as compared to $R_{\mathrm{M}}(D,P)$.

In light of Theorem 2, when $X \sim \mathrm{Ber}(\frac{1}{2})$ and $\Delta(x,\hat{x}) = d_H(x,\hat{x})$,

$$\begin{aligned} R_{\mathrm{C}}(D,0) &\leq \log 2 - \hbar(D,0) \\ &= \begin{cases} \log 2 - H_b(\frac{1-\sqrt{1-2D}}{2}), & D \in [0,\frac{1}{2}), \\ 0, & D \in [\frac{1}{2},\infty). \end{cases} \end{aligned}$$

Interestingly, the corresponding $R_{\mathrm{M}}(D,0)$ is given by [9, Theorem 7]

$$R_{\mathrm{M}}(D,0) = \begin{cases} \log 2 - H_b(\frac{1-\sqrt{1-2D}}{2}), & D \in [0,\frac{1}{2}), \\ 0, & D \in [\frac{1}{2},\infty). \end{cases}$$

In view of (7), we must have

$$R_{\mathrm{C}}(D,0) = R_{\mathrm{M}}(D,0) \tag{12}$$

for this special case. As shown in the next section (see Corollary 1), this is a general phenomenon rather than a coincidence.

IV. Continuous Alphabet Sources

In this section, we consider continuous alphabet sources, more specifically, the case where $X := (X_1, X_2, \ldots, X_L)^T$ is a random vector with $\mathcal{X} = \mathbb{R}^L$. The following result indicates that Theorem 1 continues to hold for square-integrable sources (i.e., $\mathbb{E}[\|X\|^2] < \infty$) under squared error distortion measure (i.e, $\Delta(x,\hat{x}) = \|x-\hat{x}\|^2$) and squared quadratic Wasserstein distance (i.e., $\phi(p,p') = W_2^2(p,p')$, resulting from choosing $c(x,\hat{x}) = \|x-\hat{x}\|^2$). As $D=0$ corresponds to lossless source coding, which is generally impossible for continuous alphabet sources (unless p_X has a discrete support), we focus on the case $D>0$ throughout this section.

Theorem 3: Assume $\mathbb{E}[\|X\|^2] < \infty$, $\Delta(x,\hat{x}) = \|x-\hat{x}\|^2$, and $\phi(p,p') = W_2^2(p,p')$. For $D>0$ and $P \geq 0$,

$$R_{\mathrm{C}}(D,P) = R(D,P). \tag{13}$$

Moreover, in this case,

$$R(D,P) = R'(D,P), \tag{14}$$

where

$$\begin{aligned} R'(D,P) = \inf_{p_{U'\hat{X}'|X}} \ & I(X;U') \\ \text{s.t. } & X \leftrightarrow U' \leftrightarrow \hat{X}' \text{ form a Markov chain}, \quad (15) \\ & U' = \mathbb{E}[X|U'] = \mathbb{E}[\hat{X}'|U'] \text{ almost surely}, \quad (16) \\ & \mathbb{E}[\|V\|^2] + \mathbb{E}[\|\hat{V}\|^2] \leq D, \quad (17) \\ & \mathbb{E}[W_2^2(p_{V|U'}(\cdot|U'), p_{\hat{V}|U'}(\cdot|U'))] \leq P, \quad (18) \end{aligned}$$

with $V := X - U'$ and $\hat{V} = \hat{X}' - U'$.

Proof: See [25, Appendix F]. ∎

Remark 4: In some scenarios, the source only takes on values from a strict subset $\mathcal{X}$ of $\mathbb{R}^L$ and the reconstruction is also confined to $\mathcal{X}$. The proof of (13) is not directly applicable to such scenarios as the output of quantizer ξ may live outside $\mathcal{X}$ (except for the special case $P=0$ where the reconstruction is forced to have the same distribution as the source). Nevertheless, it can be shown (see [25, Appendix G]) using a more delicate argument that (13) continues to hold in the aforementioned scenarios (correspondingly, $R(D,P)$ is defined with $\hat{X}$ restricted to $\mathcal{X}$). On the other hand, except for the special case $P=0$, the proof of (14) relies critically on the fact that $\hat{X}$ and $\hat{X}'$ in the definition of $R(D,P)$ and $R'(D,P)$ have the freedom to take on values from $\mathbb{R}^L$.

Remark 5: The equivalent characterization in (14) suggests that it suffices to consider MMSE representation U' since any other optimal representation X' can be generated from U' through a simple noise-adding mechanism. This is closely related to the universality of MMSE representation observed in the setting with (unlimited) shared randomness [5].

The following result indicates that $R_{\mathrm{C}}(D,0)$ is always equal to $R_{\mathrm{M}}(D,0)$ under squared error distortion measure, and connects them to the classical rate-distortion function.

Corollary 1: Assume $\mathbb{E}[\|X\|^2] < \infty$ and $\Delta(x,\hat{x}) = \|x-\hat{x}\|^2$. For $D>0$,

$$R_{\mathrm{C}}(D,0) = R_{\mathrm{M}}(D,0) = R\left(\frac{D}{2}\right), \tag{19}$$

where

$$\begin{aligned} R\left(\frac{D}{2}\right) = \inf_{p_{\bar{U}|X}} \ & I(X;\bar{U}) \\ \text{s.t. } & \mathbb{E}[\|X-\bar{U}\|^2] \leq \frac{D}{2}. \end{aligned} \tag{20}$$

Proof: See [25, Appendix H]. ∎

Remark 6: The second equality in (19) is known [9, Eq. (20)] (see also [23, Theorem 2], [24, Theorem 2]).

Remark 7: Note that (12) can be viewed as a special case of Corollary 1, because $d_H(x,\hat{x}) = \|x-\hat{x}\|^2$ for $x,\hat{x} \in \{0,1\}$.

Remark 8: In the definition of $R(\frac{D}{2})$, we allow $\bar{U}$ to have the freedom to take on values from $\mathbb{R}^L$ even if X is only defined over a strict subset of $\mathbb{R}^L$. Otherwise, the second equality in (19) might not hold. For example, when $X \sim \mathrm{Ber}(\frac{1}{2})$ and $\Delta(x,\hat{x}) = \|x-\hat{x}\|^2$, we have $R_{\mathrm{C}}(\frac{1}{2},0) = R_{\mathrm{M}}(\frac{1}{2},0) = 0$; on the other hand, $R(\frac{1}{4}) = 0$ only if $\bar{U}$ is allowed to be equal to $\frac{1}{2}$, which does not belong to $\{0,1\}$. See Remark 4 for a related discussion.

The next result extends the Shannon lower bound [2, Eq. (13.159)] to the RDP setting.

Theorem 4: Assume $\sigma_\ell^2 = \mathbb{E}[(X_\ell - \mathbb{E}[X_\ell])^2] \in (0,\infty)$, $\ell \in [1:L]$, $\Delta(x,\hat{x}) = \|x-\hat{x}\|^2$, and $\phi(p,p') = W_2^2(p,p')$. For $D>0$ and $P \geq 0$,

$$R_{\mathrm{C}}(D,P) \geq h(X) - \sum_{\ell=1}^{L} \frac{1}{2}\log(2\pi e\omega_\ell),$$

where

$$\omega_\ell = \begin{cases} \omega \wedge \sigma_\ell^2, & D + \sqrt{(2D-(D\wedge P))(D\wedge P)} < 2\sum_{\ell'=1}^{L}\sigma_{\ell'}^2, \\ \sigma_\ell^2, & D + \sqrt{(2D-(D\wedge P))(D\wedge P)} \geq 2\sum_{\ell'=1}^{L}\sigma_{\ell'}^2, \end{cases} \quad \ell \in [1:L], \tag{21}$$

with ω being the unique solution to

$$\sum_{\ell=1}^{L}(\omega \wedge \sigma_\ell^2) = \frac{D + \sqrt{(2D-(D\wedge P))(D\wedge P)}}{2}.$$

Proof: See [25, Appendix I]. ∎

The following result provides a partial characterization of $R_{\mathrm{C}}(D,P)$ for Gaussian mixture sources. Let $\sum_{k=1}^{K}\beta_k\mathcal{N}(\mu_k,\Sigma_k)$ be a mixture of K Gaussian distributions, $\mathcal{N}(\mu_k,\Sigma_k)$, $k \in [1:K]$, with $\beta_k > 0$, $k \in [1:K]$, and $\sum_{k=1}^{K}\beta_k = 1$. We assume $\Sigma_k \succ 0$ and consequently $\lambda_{\min}(\Sigma_k) > 0$, $k \in [1:K]$, where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of symmetric matrix A.

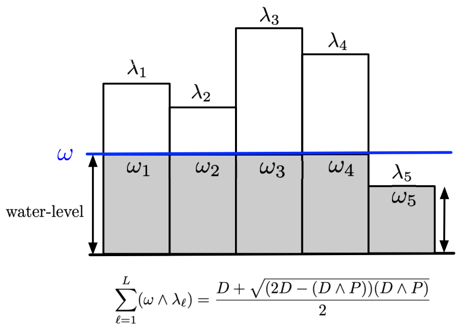

Fig. 2. Water-filling solution for the Gaussian vector source.

Corollary 2: Assume $X \sim \sum_{k=1}^{K} \beta_k \mathcal{N}(\mu_k, \Sigma_k)$, $\Delta(x, \hat{x}) = \|x - \hat{x}\|^2$, and $\phi(p, p') = W_2^2(p, p')$. For $D > 0$ and $P \geq 0$ satisfying

$$\frac{D + \sqrt{(2D - (D \wedge P))(D \wedge P)}}{2L} \leq \min\{\lambda_{\min}(\Sigma_k)\}_{k=1}^{K},$$

we have

$$R_{\mathrm{C}}(D, P) = h(X) - \frac{L}{2} \log\left(\frac{2\pi e(D + \sqrt{(2D - (D \wedge P))(D \wedge P)})}{2L}\right).$$

Proof: See [25, Appendix J]. ∎

A complete characterization of $R_{\mathrm{C}}(D, P)$ can be obtained for Gaussian sources, namely, $X \sim \mathcal{N}(\mu, \Sigma)$. Let $\Sigma = \Theta^T \Lambda \Theta$ be the eigenvalue decomposition of Σ, where Θ is a unitary matrix and Λ is a diagonal matrix with the ℓ-th diagonal entry denoted by λ_ℓ, $\ell \in [1 : L]$. We assume $\Sigma \succ 0$ and consequently $\lambda_\ell > 0$, $\ell \in [1 : L]$.

Corollary 3: Assume $X \sim \mathcal{N}(\mu, \Sigma)$, $\Delta(x, \hat{x}) = \|x - \hat{x}\|^2$, and $\phi(p, p') = W_2^2(p, p')$. For $D > 0$ and $P \geq 0$,

$$R_{\mathrm{C}}(D, P) = \sum_{\ell=1}^{L} \frac{1}{2} \log\left(\frac{\lambda_\ell}{\omega_\ell}\right),$$

where ω_ℓ is defined in (21) with σ_ℓ^2 replaced by λ_ℓ, $\ell \in [1 : L]$ (see Fig. 2).

Proof: See [25, Appendix K]. ∎

Remark 9: Note that $R_{\mathrm{C}}(D, P)$ degenerates to the rate-distortion function $R(D)$ of the Gaussian vector source with quadratic distortion when $P \geq D$, where $R(D)$ is given by the conventional reverse waterfilling formula [2, Theorem 13.3.3]

$$R(D) = \begin{cases} \sum_{\ell=1}^{L} \frac{1}{2} \log\left(\frac{\lambda_\ell}{\omega \wedge \lambda_\ell}\right), & D < \sum_{\ell=1}^{L} \lambda_\ell, \\ 0, & D \geq \sum_{\ell=1}^{L} \lambda_\ell, \end{cases}$$

with ω being the unique solution to

$$\sum_{\ell=1}^{L} (\omega \wedge \lambda_\ell) = D.$$

It is also easy to verify that $R_{\mathrm{C}}(D, 0) = R(\frac{D}{2})$, which is consistent with Corollary 1.

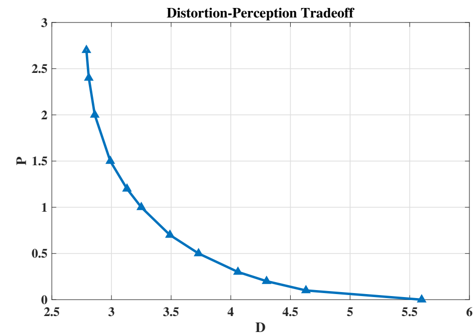

Fig. 3. Distortion-Perception tradeoff for $R = 1$, $\lambda_1 = 1$, $\lambda_2 = 2$ and $\lambda_3 = 3$.

Remark 10: Note that ω_ℓ can be interpreted as the water level in the subspace associated with eigenvalue λ_ℓ, $\ell \in [1 : L]$ (see Fig. 2). Different from the conventional waterfilling formula where the water level coincides with the distortion loss in each subspace when the distortion constraint is active, the situation is more complex here due to the presence of the perception constraint. Specifically, for $\ell \in [1 : L]$, the distortion loss D_ℓ and the perception loss P_ℓ in the subspace associated with eigenvalue λ_ℓ are given respectively by[2]

$$D_\ell = \begin{cases} \left(\frac{2D^2 - 2D\sqrt{(2D-P)P}}{(D-P)^2}\right)\omega_\ell, & D > P, \\ \omega_\ell, & D \leq P, \end{cases}$$

$$P_\ell = \begin{cases} \left(\frac{\sqrt{(2D-P)P} - P}{D-P}\right)^2 \omega_\ell, & D > P, \\ \omega_\ell, & D \leq P. \end{cases}$$

In Fig. 3, based on Corollary 3 and for a given rate, we plot the tradeoff between distortion and perception for some values of parameters.

V. Conclusion

This paper characterizes the RDP tradeoff for both finite and continuous alphabet sources when the perception measure is based on the divergence between the distributions of the source and reconstruction sequences conditioned on the encoder output. For the Gaussian vector source, a novel waterfilling type solution is obtained under squared error distortion measure and squared quadratic Wasserstein perception measure. In contrast to the conventional reverse waterfilling solution, here the water level depends on both distortion and perception losses. Throughout this work, we have focused on the setting when no shared randomness is assumed between the encoder and the decoder. When shared randomness is available, the analysis of the proposed conditional-distribution based perception measure appears significantly harder and is left for future research.

[2]Actually D_ℓ and P_ℓ are not uniquely defined when $D + \sqrt{(2D - (D \wedge P))(D \wedge P)} > 2\sum_{\ell=1}^{L} \lambda_\ell$ (which implies $R_{\mathrm{C}}(D, P) = 0$).

References

[1] Y. Blau and T. Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *Proc. Int. Conf. on Mach. Learn. (ICML)*. PMLR, 2019, pp. 675–685.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed*. Wiley, 2006.

[3] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 6228–6237.

[4] L. Theis and A. Wagner, "A coding theorem for the rate-distortion-perception function," in *Neural Compression Workshop of Int. Conf. Learn. Represent. (ICLR)*, 2021.

[5] G. Zhang, J. Qian, J. Chen, and A. Khisti, "Universal rate-distortion-perception representations for lossy compression," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2021, pp. 11 517–11 529.

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

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

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

[9] J. Chen, L. Yu, J. Wang, W. Shi, Y. Ge, and W. Tong, "On the rate-distortion-perception function," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 4, pp. 664–673, 2022.

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

[11] X. Niu, D. Gündüz, B. Bai, and W. Han, "Conditional rate-distortion-perception trade-off," *arXiv:2305.09318*, 2023.

[12] Y. Hamdi and D. Gündüz, "The rate-distortion-perception trade-off with side information," *arXiv:2305.13116*, 2023.

[13] F. Mentzer, E. Agustsson, J. Ballé, D. Minnen, N. Johnston, and G. Toderici, "Neural video compression using gans for detail synthesis and propagation," in *Proc. Europ. Conf. Comput Vis. (ECCV)*, 2022.

[14] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool, "Generative adversarial networks for extreme learned image compression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 221–231.

[15] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," in *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2017.

[16] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017.

[17] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, "Conditional probability models for deep image compression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018.

[18] F. Mentzer, G. Toderici, M. Tschannen, and E. Agustsson, "High-fidelity generative image compression," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2020.

[19] A. Golinski, R. Pourreza, Y. Yang, G. Sautiere, and T. S. Cohen, "Feedback recurrent autoencoder for video compression," in *Proc. Asian Conf. Comput. Vis.*, 2020.

[20] H. Liu, G. Zhang, J. Chen, and A. Khisti, "Lossy compression with distribution shift as entropy constrained optimal transport," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022.

[21] S. Salehkalaibar, B. Phan, A. Khisti, and W. Yu, "Rate-distortion-perception tradeoff based on the conditional perception measure," in *Proc. IEEE Biennial Symp. Commun. (BSC)*, 2023, pp. 31–37.

[22] T. A. Atif, M. A. Sohail, and S. S. Pradhan, "Lossy quantum source coding with a global error criterion based on a posterior reference map," *arXiv: 2302.00625*, 2023.

[23] Z. Yan, F. Wen, R. Ying, C. Ma, and P. Liu, "On perceptual lossy compression: The cost of perceptual reconstruction and an optimal training framework," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021.

[24] L. Theis and E. Agustsson, "On the advantages of stochastic encoders," in *Neural Compression Workshop of Int. Conf. Learn. Represent. (ICLR)*, 2021.

[25] S. Salehkalaibar, J. Chen, A. Khisti, and W. Yu, "Rate-distortion-perception tradeoff based on the conditional-distribution perception measure," *arXiv: 2401.12207*, 2024.

[26] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.