Rate-Distortion-Perception Function of Gaussian Vector Sources

Jingjing Qian[1], Sadaf Salehkalaibar[2], Jun Chen[1], Ashish Khisti[2], Wei Yu[2], Wuxian Shi[3], Yiqun Ge[3] and Wen Tong[3]
[1]*ECE Department, McMaster University, Hamilton, Canada*, email: {qianj40, chenjun}@mcmaster.ca
[2]*ECE Department, University of Toronto, Toronto, Canada*, emails: {sadafs, akhisti, weiyu}@ece.utoronto.ca
[3]*Ottawa Research Center, Huawei Technologies, Ottawa, Canada*, emails: {wuxian.shi, yiqun.ge, tongwen}@huawei.com

***Abstract*—This paper studies the rate-distortion-perception (RDP) tradeoff for a Gaussian vector source coding problem where the goal is to compress the multi-component source subject to distortion and perception constraints. The purpose of imposing a perception constraint is to ensure visually pleasing reconstructions. Without the perception constraint, the traditional reverse water-filling solution for characterizing the rate-distortion (RD) tradeoff of a Gaussian vector source states that the optimal rate allocated to each component depends on a *constant*, called the water-level. If the variance of a specific component is below the water-level, it is assigned a *zero* compression rate. However, with active distortion and perception constraints, we show that the optimal rates allocated to the different components are always *positive*. Moreover, the water-levels that determine the optimal rate allocation for different components are *unequal*. We further treat the special case of perceptually perfect reconstruction and study its RDP function in the high-distortion and low-distortion regimes to obtain insight to the structure of the optimal solution.**

I. Introduction

The rate-distortion-perception (RDP) function is a generalization of Shannon's rate-distortion function that incorporates an additional perception loss function which measures the distance between the distributions of the source and the reconstruction. It has been observed that in the neural compression framework [1]–[4], improving realism in the reconstruction comes at the price of increased distortion. In this framework, realism is controlled by a perception loss function between the distributions of the source and the reconstruction, while distortion is controlled via a standard distortion loss function on the samples of the source and its reconstruction, e.g., in terms of mean squared error. The RDP function introduced in Blau and Michaeli [5] formalizes this tradeoff.

The extension of classical rate-distortion (RD) theory to incorporate constraints on the distribution of the reconstruction samples has been studied in various works in the information theory literature; see e.g. [6] and references therein. More recently, Theis and Wagner [7] present a one-shot coding theorem by means of the strong functional representation lemma (SFRL) [8] to establish the operational validity of the RDP function [5]. In [9], the authors establish analytic properties of the RDP function for the special case of (scalar) Gaussian sources, with a quadratic distortion function and a perception loss function of either Kullback–Leibler (KL) divergence or Wasserstein-2 distance between the source and the reconstruction distributions. The role of common randomness in the study of RDP function has been studied in [10], [11]. Furthermore the distortion-perception tradeoff with a squared error distortion and Wasserstein-2 perception loss, but without an explicit compression rate constraint, has been studied in [12], [13], where it is shown that the entire tradeoff curve can be achieved by interpolating the two extremal reconstructions based on a given representation. Other related works include [14], [15].

This paper studies the RDP function of a Gaussian vector source under a squared error distortion and either KL divergence or Wasserstein-2 distance as the perception loss metric. Our result is thus an extension of prior work [9] on scalar Gaussian source to the case of vector sources. We note that without the perception constraint, the rate-distortion function of a parallel Gaussian source model has a classical *reverse water-filling* characterization [16, Thm 10.3], where the optimal rate allocation across the components is computed according to a distortion dependent parameter called *water-level*. A positive rate is assigned to those components that have a variance above this parameter. Any component whose variance is below the water-level has a zero rate (see Fig. 1a).

In this work, we study the optimal rate allocation associated with the Gaussian vector source model and compare the result with the reserve water-filling solution. We observe a qualitatively different solution as shown in Fig. 1(b). First, unlike the case of reverse water-filling, the associated water-level for each component can be different and is characterized as a solution to a set of equations. Second, while reverse water-filling assigns zero rate to those source components whose variances are below the water-level, all components in the RDP setting are assigned a non-zero rate as long as both the distortion and perception constraints are active. We further consider the special case of zero perception loss (so the source and reconstruction distributions are identical) and establish analytical results in this case. Moreover, we present asymptotic results on high and low distortion cases with zero perception, and shed additional insights into the difference between the RDP function and the RD function.

II. System Model and Preliminaries

Let $X \sim P_X$ be an L-dimensional Gaussian vector source with mean 0 and covariance matrix $\Sigma_X \succ 0$. Consider the

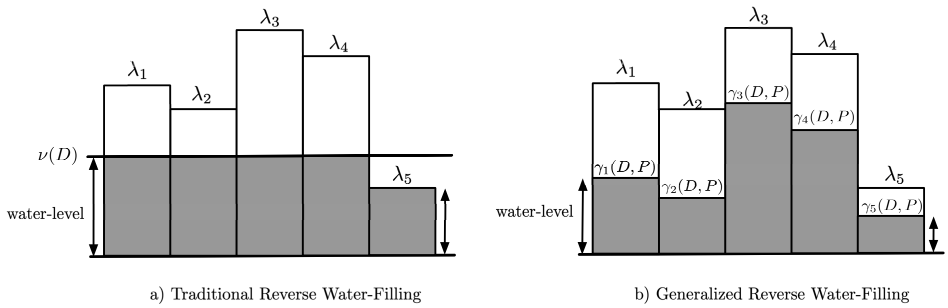

Fig. 1. (a) Without a perception constraint, the traditional reverse water-filling solution for a parallel Gaussian source fixes a constant water-level. When the variance of a specific component is less than the water-level, it is assigned zero rate. (b) With an active perception constraint, unequal water-levels are assigned to different components. The variance of each component is always greater than the corresponding water-level. Every component has a positive rate.

eigenvalue decomposition of Σ_X as follows:

$$\Sigma_X = \Theta^T \Lambda_X \Theta, \tag{1}$$

where Θ is unitary and Λ_X is a diagonal matrix of eigenvalues

$$\Lambda_X = \mathrm{diag}^L(\lambda_1, \dots, \lambda_L). \tag{2}$$

We assume that there is unlimited common randomness $K \in \mathcal{K}$ shared between the encoder and the decoder. Consider the following *one-shot* encoding and decoding functions where the source samples are encoded one at a time:

$$f\colon \mathbb{R}^L \times \mathcal{K} \to \mathcal{M}, \tag{3}$$
$$g\colon \mathcal{M} \times \mathcal{K} \to \mathbb{R}^L. \tag{4}$$

Here, $\mathcal{M}$ denotes the set of messages. Let $P_{\hat{X}}$ be the distribution of the reconstruction induced by the encoding and decoding mechanisms. In this paper, we measure distortion using a *squared-error* loss function $d\colon \mathbb{R}^L \times \mathbb{R}^L \to \mathbb{R}_{\geq 0}$ where $d(x,\hat{x}) := \|x-\hat{x}\|^2$. From a perceptual perspective, for given probability distributions P_X and $P_{\hat{X}}$, we use $\phi(P_X, P_{\hat{X}})$ to denote the perception loss function capturing the difference between the two distributions. Notice that $\phi(P_X, P_{\hat{X}}) = 0$ if and only if $P_X = P_{\hat{X}}$ almost surely.

The above framework is referred to as the one-shot setting, because it compresses one sample at a time. We can also define an *asymptotic setting* of encoding n independently and identically distributed (i.i.d.) samples $X^n = (X_1, \dots, X_n)$ and reconstructing $\hat{X}^n = (\hat{X}_1, \dots, \hat{X}_n)$, with $n \to \infty$.

Definition 1 (Operational RDP Functions): Let $X \sim P_X$. For given distortion-perception constraints (D, P), a rate R is said to be achievable if there exist encoder and decoder satisfying

$$\mathbb{E}[\ell(M)] \leq R, \tag{5}$$
$$\mathbb{E}[\|X - \hat{X}\|^2] \leq D, \tag{6}$$
$$\phi(P_X, P_{\hat{X}}) \leq P, \tag{7}$$

where $\ell(M)$ denotes the length of the message M for encoding one sample. The infimum of all achievable rates R is called *one-shot rate-distortion-perception (RDP) function*, denoted as $R^o(D,P)$.

For the asymptotic setting, a rate R is said to be achievable if there exist encoding and decoding functions such that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\|X_i - \hat{X}_i\|^2] \leq D, \tag{8}$$
$$\frac{1}{n}\sum_{i=1}^{n} \phi(P_{X_i}, P_{\hat{X}_i}) \leq P, \tag{9}$$

with the message M that encodes X^n satisfying

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\ell(M)] \leq R. \tag{10}$$

The infimum of all achievable rates is called *asymptotic RDP function*, denoted as $R^\infty(D,P)$.

Definition 2 (Information RDP Function): For given $X \sim P_X$, let $\mathcal{P}_{\hat{X}|X}(D,P)$ be the set of all conditional distributions $P_{\hat{X}|X}$ such that for a fixed (D,P), we have

$$\mathbb{E}[\|X - \hat{X}\|^2] \leq D, \qquad \phi(P_X, P_{\hat{X}}) \leq P. \tag{11}$$

The *rate-distortion-perception (RDP) function* is defined as

$$R(D,P) = \inf_{P_{\hat{X}|X} \in \mathcal{P}_{\hat{X}|X}(D,P)} I(X;\hat{X}), \tag{12}$$

Remark 1: Using the functional representation lemma as in [8] and following similar steps to Theorem 2 in Appendix A.2 of [9], one can show that

$$R^\infty(D,P) = R(D,P), \tag{13}$$

and

$$R(D,P) \leq R^o(D,P) \leq R(D,P) + \log(R(D,P)+1) + 5.$$

Consequently, the information RDP function $R(D,P)$ and the one-shot operational RDP function $R^o(D,P)$ are asymptotically close to each other at high rate.

In the rest of the paper, the perception metric $\phi(P_X, P_{\hat{X}})$ is assumed to be either the KL-divergence, i.e.,

$$D(P_{\hat{X}} \| P_X) = \int_x P_{\hat{X}}(x) \log \frac{P_{\hat{X}}(x)}{P_X(x)} dx, \tag{14}$$

or the Wasserstein-2 distance, i.e.,

$$W_2^2(P_X, P_{\hat{X}}) = \inf \mathbb{E}[\|X - \hat{X}\|^2], \tag{15}$$

where the infimum is taken over all joint distributions of $(X, \hat{X})$ with marginals P_X and $P_{\hat{X}}$.

III. Classical Reverse Water-Filling

The classical rate-distortion theory for a parallel Gaussian source states that the optimal rate allocated to each component depends on a constant parameter, called *water-level*. The water-level also represents the distortion allowed at those components whose variances are above the water-level. For a given distortion D, let $\nu(D)$ to be the solution to the equation

$$\sum_{\ell=1}^{L}[\lambda_\ell - \nu(D)]^+ = \left[\sum_{\ell=1}^{L}\lambda_\ell - D\right]^+, \tag{16}$$

where $[x]^+ := \max\{0, x\}$. Now, let

$$\gamma_\ell(D,\infty) = \begin{cases} \lambda_\ell & \text{if } \nu(D) \geq \lambda_\ell, \\ \nu(D) & \text{if } \nu(D) < \lambda_\ell. \end{cases} \tag{17}$$

The rate-distortion function for the Gaussian vector source with variance λ_ℓ on each of its component is as follows.

Theorem 1 (Thm 10.3 in [16]): For a Gaussian vector source, we have

$$R(D,\infty) = \frac{1}{2}\sum_{\ell=1}^{L}\log\frac{\lambda_\ell}{\gamma_\ell(D,\infty)}. \tag{18}$$

To simplify notation, we can redefine the water-level as $\gamma_\ell(D,\infty)$ in order account for components whose variances are below the water-level, If λ_ℓ is below $\nu(D)$ for some ℓ, then we set $\gamma_\ell(D,\infty) = \lambda_\ell$ and assign zero rate to this component. Two special cases of the above theorem are interesting.

Proposition 1 (High-Distortion Compression): In the high-distortion regime, we have that for sufficiently small $\epsilon > 0$

$$R\left(\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, \infty\right) = \frac{\epsilon}{2\lambda^{\max}} + O(\epsilon^2), \tag{19}$$

where $\lambda^{\max} = \max_\ell \lambda_\ell$. Let $L^{\max}$ denote the set of indices where their corresponding eigenvalues are equal to $\lambda^{\max}$. Then, the water-levels are given by

$$\gamma_\ell\left(\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, \infty\right) = \lambda_\ell, \qquad \forall \ell \in \{1,\ldots,L\}\backslash L^{\max}, \tag{20a}$$

$$\gamma_{\ell^{\max}}\left(\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, \infty\right) = \lambda^{\max} - \frac{\epsilon}{|L^{\max}|}, \;\; \forall \ell^{\max} \in L^{\max}. \tag{20b}$$

The above proposition states that in the high-distortion compression, a positive rate is only assigned to the components with the largest eigenvalue.

Proposition 2 (Low-Distortion Compression): In the low-distortion regime, we have that for a sufficiently small $\epsilon > 0$

$$R(\epsilon,\infty) = \frac{1}{2}\sum_{\ell=1}^{L}\log\frac{L\lambda_\ell}{\epsilon}, \tag{21}$$

where the water-levels are given by

$$\gamma_\ell(\epsilon,\infty) = \frac{\epsilon}{L}, \qquad \forall \ell \in \{1,\ldots,L\}. \tag{22}$$

For low-distortion compression, according to the above proposition, the same water-level is assigned to all components.

IV. Rate-Distortion-Perception Function

A. Generalized Reverse Water-Filling

We first state a result that says that for the two perception metrics considered in this paper and for a Gaussian vector source, jointly Gaussian reconstruction is optimal.

Theorem 2: If the perception metric is either the KL-divergence or the Wasserstein-2 distance, without loss of optimality, in the optimization problem (12), if the source is Gaussian, we can restrict to a jointly Gaussian reconstruction, i.e., the joint distribution of $(X, \hat{X})$ should be Gaussian.

Proof: See [17]. ■

A common property of the two perception metrics that enables the above theorem is that both metrics depend only on the second-order statistics of the source and the reconstruction, if they are jointly Gaussian distributed.

We now present the RDP function with the KL-divergence as the perception metric, i.e., $\phi(P_X, P_{\hat{X}}) = D(P_{\hat{X}}\|P_X)$. Similar results for the Wasserstein-2 distance as the perception metric can be found in [17] and are omitted here due to space limitations.

First, we investigate $R^o(D,P)$ and provide a one-shot coding strategy for achieving an (R,D,P) tuple. This allows an achievable (R,D,P) region to be characterized in terms of an optimization problem. The first step is to decompose the source using eigenvalue decomposition as in (1) and define

$$Z = \Theta X. \tag{23}$$

The main idea is to construct a new Gaussian random variable $\hat{Z}$ and to use the channel simulation result of [8] to communicate $\hat{Z}$ to the decoder at a rate of R. The new variable $\hat{Z}$ is designed to be correlated with Z in a very specific way in order to satisfy the distortion and perception constraints D and P, respectively. The correlation between Z and $\hat{Z}$ is controlled by two sets of parameters, $\{\gamma_\ell\}_{\ell=1}^L$ and $\{\hat{\lambda}_\ell\}_{\ell=1}^L$, such that $0 < \gamma_\ell \leq \lambda_\ell$ and $0 < \hat{\lambda}_\ell \leq \lambda_\ell$. The optimal values of these parameters will be determined later.

In effect, instead of the classical rate-distortion setting where $\hat{Z}$ is chosen to minimize the distortion subject to the rate constraint, here we choose $\hat{Z}$ to satisfy both distortion and perception constraints. We construct this noisy version of Z at the decoder by taking advantage of the availability of common randomness.

Specifically, $\hat{Z}$ is a zero-mean Gaussian random vector with independent components, jointly distributed with Z with

$$\mathrm{cov}(Z_\ell, \hat{Z}_\ell) = \begin{bmatrix} \lambda_\ell & \sqrt{\hat{\lambda}_\ell(\lambda_\ell - \gamma_\ell)} \\ \sqrt{\hat{\lambda}_\ell(\lambda_\ell - \gamma_\ell)} & \hat{\lambda}_\ell \end{bmatrix}. \tag{24}$$

With the above covariance structure, we can verify that γ_ℓ is the minimum mean-squared error (MMSE) of estimating Z_ℓ based on $\hat{Z}_\ell$, i.e.,

$$\gamma_\ell = \mathbb{E}[(Z_\ell - \mathbb{E}[Z_\ell|\hat{Z}_\ell])^2]. \tag{25}$$

We now use a consequence of the SFRL [8, Theorem 1] to show that when common randomness K is available at both the encoder and decoder, there exists a channel simulation scheme that allows $\hat{Z}_\ell$ to be reconstructed at the decoder at a communication rate of

$$I(Z_\ell;\hat{Z}_\ell)+\log(I(Z_\ell;\hat{Z}_\ell)+1)+5. \tag{26}$$

After the reconstruction of $\hat{Z}_\ell$ at the decoder, we use the same unitary matrix to transform it into $\hat{X}$, i.e.,

$$\hat{X}=\Theta^T\hat{Z}. \tag{27}$$

The above scheme leads to the rate, distortion, and perception loss for the ℓth component of Z as functions of λ_ℓ, $\hat{\lambda}_\ell$ and γ_ℓ as follows:

$$R_\ell(\lambda_\ell,\gamma_\ell)=\frac{1}{2}\log\left(\frac{\lambda_\ell}{\gamma_\ell}\right)+\log\left(\frac{1}{2}\log\left(\frac{\lambda_\ell}{\gamma_\ell}\right)+1\right)+5, \tag{28}$$

$$D_\ell(\lambda_\ell,\hat{\lambda}_\ell,\gamma_\ell)=\lambda_\ell-2\sqrt{\hat{\lambda}_\ell(\lambda_\ell-\gamma_\ell)}+\hat{\lambda}_\ell, \tag{29}$$

$$P_\ell(\lambda_\ell,\hat{\lambda}_\ell)=\frac{1}{2}\left(\frac{\hat{\lambda}_\ell}{\lambda_\ell}-1+\log\frac{\lambda_\ell}{\hat{\lambda}_\ell}\right). \tag{30}$$

This would allow a characterization of an achievable one-shot RDP function of a Gaussian vector source as an optimization problem over $\hat{\lambda}_\ell$ and γ_ℓ across its components.

For the asymptotic setting, the achievable scheme is identical, except that we compress a block of n samples together. As $n\to\infty$, the logarithm and the constant terms in (28) can be neglected. This leads to a characterization of an achievable region for $R^\infty(D,P)$, which is equal to $R(D,P)$. This achievable region turns out to be optimal, i.e., a converse can be proved. This gives the following characterization of $R(D,P)$.

Theorem 3: The rate-distortion-perception function $R(D,P)$ for a Gaussian vector source with parameters defined by (1) and (2), and with KL-divergence as the perception metric, is given by the solution to the following optimization problem:

$$R(D,P)=\min_{\{\hat{\lambda}_\ell,\gamma_\ell\}_{\ell=1}^L}\frac{1}{2}\sum_{\ell=1}^{L}\log\frac{\lambda_\ell}{\gamma_\ell} \tag{31a}$$
$$\text{s.t.}\quad 0<\gamma_\ell\le\lambda_\ell, \tag{31b}$$
$$0\le\hat{\lambda}_\ell\le\lambda_\ell, \tag{31c}$$
$$\sum_{\ell=1}^{L}D_\ell(\gamma_\ell,\hat{\lambda}_\ell)\le D, \tag{31d}$$
$$\sum_{\ell=1}^{L}P_\ell(\lambda_\ell,\hat{\lambda}_\ell)\le P. \tag{31e}$$

Proof: See [17]. ■

An interpretation of the above is as follows. For given (D,P), let $\gamma_\ell^*(D,P)$ and $\hat{\lambda}_\ell^*(D,P)$, $\ell\in\{1,\dots,L\}$, be the optimal solution to (31). Comparing this with (18), it can be seen that $\gamma_\ell^*(D,P)$ can be interpreted as the water-level for the ℓ-th component, which determines the rate allocated to that component according to (31a) (see Fig. 1b).

We now proceed to analyze the solution to the optimization program in Theorem 3. It can be shown that the optimization problem (31) is convex [17]. Let ν_1, ν_2, $\{\xi_\ell\}_{\ell=1}^L$, $\{\eta_\ell,\eta'_\ell\}_{\ell=1}^L$ be nonnegative Lagrange multipliers. For $\ell\in\{1,\dots,L\}$, we have the first-order conditions:

$$\frac{1}{2\gamma_\ell^*(D,P)}-\nu_1\sqrt{\frac{\hat{\lambda}_\ell^*(D,P)}{\lambda_\ell-\gamma_\ell^*(D,P)}}-\xi_\ell=0, \tag{32}$$

and

$$\nu_1\left(-\sqrt{\frac{\lambda_\ell-\gamma_\ell^*(D,P)}{\hat{\lambda}_\ell^*(D,P)}}+1\right)+\frac{\nu_2}{2}\left(\frac{1}{\lambda_\ell}-\frac{1}{\hat{\lambda}_\ell^*(D,P)}\right)+\eta_\ell-\eta'_\ell=0, \tag{33}$$

We focus on the most interesting regime where the distortion and the perception constraints are both active so $\nu_1,\nu_2>0$, and $\xi_\ell=\eta_\ell=\eta'_\ell=0$ for all $\ell\in\{1,\dots,L\}$. In this case, (32) and (33) yield the following solutions

$$\hat{\lambda}_\ell^*(D,P)=\frac{\lambda_\ell(-b_\ell+\sqrt{b_\ell^2+8\lambda_\ell\nu_1\nu_2(2\lambda_\ell\nu_1+\nu_2)}+2\nu_2^2)}{2(2\lambda_\ell\nu_1+\nu_2)^2}, \tag{34}$$

$$\gamma_\ell^*(D,P)=\frac{-2\lambda_\ell\nu_1(1+2\lambda_\ell\nu_1)-\nu_2+\sqrt{b_\ell^2+8\lambda_\ell\nu_1\nu_2(2\lambda_\ell\nu_1+\nu_2)}}{8\lambda_\ell\nu_1^2(-1+\nu_2)},$$

where

$$b_\ell=2\lambda_\ell\nu_1-4\lambda_\ell^2\nu_1^2+\nu_2-4\lambda_\ell\nu_1\nu_2. \tag{35}$$

The above expressions give us the following generalized reverse water-filling interpretaton of the optimal RDP solution. At given distortion constraint D and perception constraint P, each component of the source with variance λ_ℓ is reconstructed by $\hat{Z}_\ell$ having variance $\hat{\lambda}_\ell^*(D,P)$. Because $\gamma_\ell^*(D,P)$ is the variance of the MMSE error for estimating Z_ℓ given $\hat{Z}_\ell$, this requires a rate of $\frac{1}{2}\log\left(\frac{\lambda_\ell}{\gamma_\ell^*(D,P)}\right)$. The parameters $\hat{\lambda}_\ell^*(D,P)$ and $\gamma_\ell^*(D,P)$ are chosen to satisfy the distortion and perception constraints. As already mentioned, $\gamma_\ell^*(D,P)$ can be thought of as the water-level, cf. (18).

When the distortion and perception constraints are active[1], i.e., $\nu_1,\nu_2>0$, based on (33) we have that in the finite-rate regime (i.e., $\gamma_\ell^*(D,P)>0$), the following must hold

$$\lambda_\ell>\hat{\lambda}_\ell^*(D,P)>\lambda_\ell-\gamma_\ell^*(D,P). \tag{36}$$

Together with (32), this implies that $\lambda_\ell>\gamma_\ell^*(D,P)$, so every component of the source is always allocated a non-zero rate regardless of the distortion constraint—unlike the traditional reverse water-filling solution, where a component may be allocated zero rate if its variance is below the water-level.

B. Perceptually Perfect Reconstruction

In this section, we focus on the special case of perfect perceptual quality, and study the properties of the RDP function with $P=0$. The proofs are deferred to [17].

[1]It is possible to prove that under the KL-divergence perception metric, at a positive rate the perception constraint is always active.

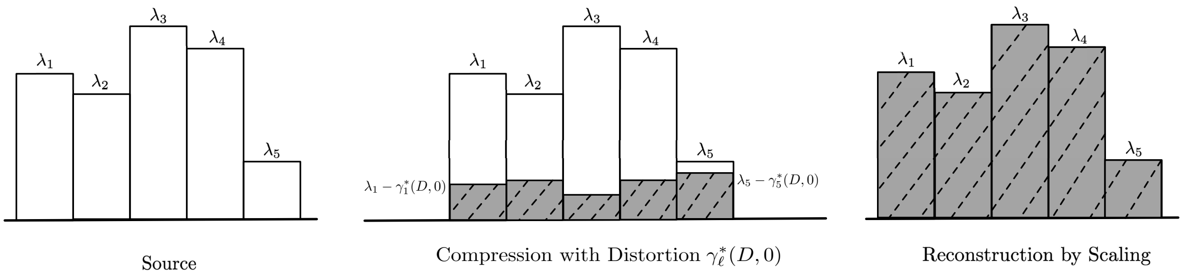

Fig. 2. Generalized reverse water-filling solution for the perceptually perfect reconstruction. The source is first compressed to get a representation whose components have distortion levels $\gamma_\ell^*(D,0)$, $\ell = 1, \cdots, L$. After compression, each component has a variance given by $\lambda_\ell - \gamma_\ell^*(D,0)$. Each component is then scaled to generate a reconstruction whose distribution matches that of the original source.

Corollary 1: The RDP function of a Gaussian vector source with $P = 0$ is

$$R(D,0) = \frac{1}{2}\sum_{\ell=1}^{L} \log \frac{1+\sqrt{1+16\nu_1^2\lambda_\ell^2}}{2}, \tag{37}$$

for some positive ν_1 that satisfies

$$D = \sum_{\ell=1}^{L}\left[2\lambda_\ell - 2\sqrt{\lambda_\ell(\lambda_\ell - \gamma_\ell^*(D,0))}\right], \tag{38}$$

where

$$\gamma_\ell^*(D,0) = \frac{2\lambda_\ell}{1+\sqrt{1+16\nu_1^2\lambda_\ell^2}}, \qquad \ell \in \{1,\ldots,L\}. \tag{39}$$

An interpretation of optimal rate allocation in this $P = 0$ case is as follows. By (37), the optimal rate allocated to the ℓ-th component is controlled by the expression $\frac{1+\sqrt{1+16\nu_1^2\lambda_\ell^2}}{2}$. So, if a component has a larger variance, it is compressed at a higher rate. Further, by (39) it also has a higher water-level.

The encoding and decoding scheme here can be thought of as first compressing each component of the source (using SFRL) at the individual rate given by (37) and distortion given by (39); then, the decoder simply scales each component of the compressed signal to match the variance of the source in order to ensure zero perception loss, as shown in Fig. 2. The resulting distortion after scaling is given by (38). Note that this interpretation also applies to the arbitrary P case with $\gamma_\ell^*(D,P)$ as the distortion after compression, then scaling to get a variance of $\hat{\lambda}_\ell^*(D,P)$. We further note that the compression in this case depends on P; it may not be universal, unlike the *universal representations* scenario presented in [9].

Proposition 3 (High-Distortion Compression): In the high-distortion and perfect perception regime, we have that for sufficiently small $\epsilon > 0$,

$$R\left(2\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, 0\right) = \frac{\epsilon^2}{8\sum_{\ell=1}^{L}\lambda_\ell^2} + O(\epsilon^4), \tag{40}$$

where the water-levels are given by

$$\gamma_\ell^*\left(2\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, 0\right) = \lambda_\ell - \frac{\epsilon^2\lambda_\ell^2}{4\left(\sum_{\ell=1}^{L}\lambda_\ell\right)^2} + O(\epsilon^4), \quad \ell \in \{1,\ldots,L\}. \tag{41}$$

Here, we express $R(D,0)$ in term of deviation from the maximum distortion at perfect perception at zero rate. This maximum distortion can be shown to be twice of the variance of the source [9], because at zero rate, the decoder should simply generate an independent random variable with the same variance. Comparing $R\Big(2\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, 0\Big)$ of Proposition 3 with $R\Big(\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, \infty\Big)$ in Proposition 1, it is interesting to see that the variances of the source enter $R\Big(2\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, 0\Big)$ as $\sum_{\ell=1}^{L}\lambda_\ell^2$ which is the sum of the variances over all the components. This is in contrast to the corresponding factor in $R\Big(\sum_{\ell=1}^{L}\lambda_\ell - \epsilon, \infty\Big)$ in the traditional reverse water-filling solution which is simply $\lambda^{\max}$. This is a consequence of the perfect perception constraint, which requires all components to be reconstructed with the same variances as the source at the decoder.

Proposition 4 (Low-Distortion Compression): In the low-distortion and perfect perception regime, we have that for sufficiently small $\epsilon > 0$,

$$R(\epsilon,0) = \frac{1}{2}\sum_{\ell=1}^{L}\log\frac{L\lambda_\ell}{\epsilon} + \sum_{\ell=1}^{L}\frac{\epsilon}{4L\lambda_\ell} + O(\epsilon^2), \tag{42}$$

where the water-levels are given by

$$\gamma_\ell^*(\epsilon,0) = \frac{\epsilon}{L} - \frac{\epsilon^2}{2L^2\lambda_\ell} + O(\epsilon^3), \quad \ell \in \{1,\ldots,L\}. \tag{43}$$

Comparing Proposition 4 with Proposition 2, we see that in this high-rate low-distortion regime, the extra rate required to satisfy zero-perception scales as

$$R(\epsilon,0) - R(\epsilon,\infty) = \sum_{\ell=1}^{L}\frac{\epsilon}{4L\lambda_\ell} + O(\epsilon^2), \tag{44}$$

$$\gamma_\ell^*(\epsilon,\infty) - \gamma_\ell^*(\epsilon,0) = \frac{\epsilon^2}{2L^2\lambda_\ell} + O(\epsilon^3), \quad \ell \in \{1,\ldots,L\}. \tag{45}$$

V. Conclusions

This paper characterizes the RDP function for a Gaussian vector source. In contrast to the traditional reverse water-filling solution, with a perception constraint, the water-levels assigned to different components should be unequal. This leads to a positive rate allocation to every component of the source.

REFERENCES

[1] 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. and Pattern Recognit. (CVPR)*, 2019, pp. 221–231.

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

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

[4] 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. and Pattern Recognit. (CVPR)*, 2018.

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

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

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

[8] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Trans. on Info. Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

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

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

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

[12] D. Freirich, T. Michaeli, and R. Meir, "A theory of the distortion-perception tradeoff in wasserstein space," *Proc. Adv. in Neural Info. Process. Sys. (NeurIPS)*, vol. 34, pp. 25 661–25 672, 2021.

[13] 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. ACM Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 11 682–11 692.

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

[15] S. Salehkalaibar, B. Phan, J. Chen, W. Yu, and A. Khisti, "On the choice of perception loss function for learned video compression," in *Proc. Adv. in Neural Info. Process. Sys. (NeurIPS)*, 2023.

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

[17] J. Qian, S. Salehkalaibar, J. Chen, A. Khisti, W. Yu, W. Shi, Y. Ge, and W. Tong, "Rate-distortion-perception tradeoff for vector gaussian sources," 2023. [Online]. Available: https://www.comm.utoronto.ca/~akhisti/rdv.pdf