Copula-based Estimation of Continuous Sources for a Class of Constrained Rate-Distortion Functions

Giuseppe Serra*, Photios A. Stavrou*, Marios Kountouris*†
*Communication Systems Department, EURECOM, Sophia-Antipolis, France
†Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Spain
{giuseppe.serra,fotios.stavrou,marios.kountouris}@eurecom.fr

***Abstract*—We present a new method to estimate the rate-distortion-perception function in the perfect realism regime (PR-RDPF), for multivariate continuous sources subject to a single-letter average distortion constraint. The proposed approach is not only able to solve the specific problem but also two related problems: the entropic optimal transport (EOT) and the output-constrained rate-distortion function (OC-RDF), of which the PR-RDPF represents a special case. Using copula distributions, we show that the OC-RDF can be cast as an I-projection problem on a convex set, based on which we develop a parametric solution of the optimal projection proving that its parameters can be estimated, up to an arbitrary precision, via the solution of a convex program. Subsequently, we propose an iterative scheme via gradient methods to estimate the convex program. Lastly, we characterize a Shannon lower bound (SLB) for the PR-RDPF under a mean squared error (MSE) distortion constraint. We support our theoretical findings with numerical examples by assessing the estimation performance of our iterative scheme using the PR-RDPF with the obtained SLB for various sources.**

I. Introduction

Rate-distortion-perception (RDP) theory, which provides a way to reconstruct complex data sources (e.g., audio, images, video) when perceptual quality is taken into account in addition to the distortion criterion, has recently attracted increasing interest within the information theory, computer vision, and machine learning communities. This framework, proposed by Blau and Michaeli [1] and Matsumoto [2], [3], generalizes the classical rate-distortion function (RDF) formulation by imposing a divergence constraint between the source distribution and its reconstruction. In RDP theory, the divergence constraint acts as a proxy for human perception, capturing the difference between the reconstructed samples and the source "natural statistic" [4]. It can also be used as a semantic quality metric measuring the relevance of the reconstructed source from the receiver's perspective [5].

Prior to the development of the RDP theory, a similar line of research in lossy compression has studied the link between the statistical properties of the distribution of the reconstructed samples and their perceptual quality, defining the so-called *output-constrained rate-distortion problem* [6]–[8]. In this class of constrained lossy compression problems, instead of restricting the maximal statistical divergence between the source distribution and its reconstruction, the focus is on constraining the reconstruction to belong to a specific distribution, which may differ from that of the source. The resulting problem is in close proximity to the EOT problem [9], [10]. Interestingly, in both problems, the source and the reconstruction distributions are assumed to be known *a priori*.

The mathematical formulation that quantifies the operational meaning in RDP theory is the RDPF, which, much like its classical RDF counterpart, is not generally available in analytical form. Despite the general complexity, closed-form expressions have been developed under different settings [1], [11]–[13]. The absence of a general analytic solution for the RDPF led to the research of computational methods for its estimation. However, dedicated algorithmic solutions have been developed so far only for discrete sources [14] or by discretizing certain classes of continuous sources [15]. For general sources, RDPF estimation methods often rely on data-driven solutions [1], [11], [16], which unfortunately do not have convergence guarantees.

A. Contributions

In this work, we propose a new copula-based estimation method for the computation of the PR-RDPF for multivariate continuous sources subject to a single-letter average distortion constraint. Our estimation method is quite general as it also allows the computation of the EOT and the OC-RDF for which the PR-RDPF is a particular case.

The main contributions of this paper are as follows. **(i)** We show that there exists a one-to-one correspondence between the feasible set of solutions of the OC-RDF and EOT (Theorem 1), making the two problems equivalent. **(ii)** Using properties of copula distributions, we demonstrate that the OC-RDF can be reformulated as a projection problem in the geometry induced by the Kullback–Leibler (KL)-divergence, i.e., I-projection, on a convex constraint set (Problem 1). However, although this class of projection has been extensively studied in [17], the existing parametric solution is not directly suitable for computational purposes. To bypass this technical issue, we introduce a relaxation of the constraint set of the I-projection, which results in a lower bound to the original optimization objective (Problem 2) that we subsequently show that it can be made arbitrarily tight (Theorem 4). **(iii)** We characterize the parametric closed-form solution of the relaxed I-projection, whose optimal parameters can be directly obtained as the solution of a strictly convex program (Theorem 5). **(iv)** We propose an algorithmic approach via a stochastic gradient descent method, to estimate the strictly convex optimization problem of Theorem 5 (see Alg. 1). **(v)** We derive a Shannon lower bound (SLB) for the PR-RDPF under MSE distortion (Theorem 6). We supplement our theoretical results with various numerical evaluations aiming to estimate the PR-RDPF under various sources and different distortion measures via Alg. 1, and to demonstrate the efficacy of our algorithmic approach compared to the obtained SLB.

B. Notation

We indicate with $\mathbb{R}$ the set of real numbers and with $\bar{\mathbb{R}}$ the extended set $\mathbb{R} \cup \{-\infty, +\infty\}$. For a set $\mathcal{X} \subseteq \mathbb{R}^d$, we denote with $\mathcal{P}(\mathcal{X})$ the set of distribution functions thereon defined. For a random variable (RV) X defined on $\mathcal{X}$, we denote with $F_X \in \mathcal{P}(\mathcal{X})$ its distribution function (shortly, d.f.) and with f_X its probability density function (shortly, pdf). Given two RVs X and Y, we will indicate their independent product d.f. as $F_X \otimes F_Y$, equivalent to the independent product pdf $f_{X,Y} = f_X f_Y$. Furthermore, given any joint pdf $f_{X,Y}$, we will indicate with $m_X(f_{X,Y})$ and $m_Y(f_{X,Y})$ the pdf associated with the marginal RV's X and Y, respectively. We will indicate with $\mathrm{D}_{\mathrm{KL}}(F_X||F_Y)$ the Kullback–Leibler (KL)-divergence between RV's X and Y, whereas $h(X)$ and $h(X|Y)$ will denote, respectively, the differential entropy of X and the conditional differential entropy of X given Y. Lastly, given a set $\mathcal{A} \subseteq \mathbb{R}^n$, we will denote with $l_p(\mathcal{A})$ the set of functions $g : \mathcal{A} \to \mathbb{R}$ such that $\int_{\mathcal{A}} |g(s)|^p ds < \infty$.

II. Preliminaries

A. OC-RDF - A link between PR-RDPF and EOT

We begin this section by providing the mathematical definition of PR-RDPF.

Definition 1. *(PR-RDPF) Let $f_X \in \mathcal{P}(\mathcal{X})$. Then, the PR-RDPF for the source $X \sim f_X$ under a distortion measure $\Delta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$ is given as follows*

$$R_{PR}(D) = \min_{\substack{f_{Y|X} \\ \mathbb{E}[\Delta(X,Y)] \leq D \\ X \overset{d}{=} Y}} I(X, Y)$$

where the minimization is on set of Markov kernels $f_{Y|X}$.

It should be noted that the *perfect realism* regime represents a limit case of the general problem of the RDPF [1], where one constrains the reconstruction Y to have the same distribution as the source X. Although PR-RDPF became quite popular through [1], similar ideas were previously explored by Li *et. al.* in [6], in the context of distribution-preserving quantization and distribution-preserving RDF. Multiple coding theorems have been developed for PR-RDPF. For instance, Chen *et. al.* in [18] proves the necessity of some form of randomness, either private or common, between the encoder and decoder, to achieve the *perfect realism* regime and derives the associated coding theorems. Wagner, in [19], provides a coding theorem for the RDPF trade-offs for the perfect and near-perfect realism cases, when only finite common randomness between the encoder and decoder is available.

Although our primary goal in this work is to study computational aspects of the PR-RDPF for continuous sources, we do it by also studying a generalization of this problem. In particular, we study the problem of OC-RDF that was formally introduced by Saldi *et al.* in [7] (see also [6]), for which the mathematical definition is stated next.

Definition 2. *(OC-RDF) Let $f_X \in \mathcal{P}(\mathcal{X})$. Then, the OC-RDF for the source $X \sim f_X$ under a distortion measure $\Delta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$ and a target reconstruction distribution $f_Y \in \mathcal{P}(\mathcal{Y})$ is given as follows*

$$R_{OC}(D) = \min_{\substack{f_{Y|X} \in \hat{\Pi}(f_X, f_Y) \\ \mathbb{E}[\Delta(X,Y)] \leq D}} I(X, Y) \quad (1)$$

where the minimization is on the convex set of Markov kernels $\hat{\Pi}(f_X, f_Y) \triangleq \{f_{X|Y} : m_Y(f_{Y|X} \cdot f_X) = f_Y\}$.

The main difference between the problems of PR-RDPF and OC-RDF lies in how the constraint on the reconstruction distribution f_Y is handled. While in the PR-RDPF case, we specifically constrain the reconstruction distribution and source distribution to be identical, in the OC-RDF we have an additional degree of freedom, allowing for the distribution of the reconstruction to be chosen freely. This results in the following observation.

Remark 1. *The problem of the OC-RDF particularizes to the problem of PR-RDPF by specifying the reconstruction distribution to be equal to the source distribution (i.e. $f_Y = f_X$).*

Additionally, the OC-RDF highlights an interesting connection to the EOT problem (see [9], [10]), of which the mathematical definition is stated as follows.

Definition 3. *(EOT) Let $f_X \in \mathcal{P}(\mathcal{X})$ and $f_Y \in \mathcal{P}(\mathcal{Y})$. Then, the EOT for $\epsilon > 0$ and distortion measure $\Delta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$, is given as follows*

$$D_{EOT}(\epsilon) = \min_{f_{X,Y} \in \bar{\Pi}(f_X, f_Y)} \mathbb{E}[\Delta(X, Y)] + \epsilon I(X, Y) \quad (2)$$

where the minimization is on the convex set of joint pdfs $\bar{\Pi}(f_X, f_Y) \triangleq \{f_{X,Y} : m_X(f_{X,Y}) = f_X, m_Y(f_{X,Y}) = f_Y\}$.

Notably, it can be shown that OC-RDF and EOT are closely related in the sense that for specific values of D and ϵ, there exists a one-to-one mapping between the sets of solutions of the two problems. In other words, we can find the solution to one problem based on the solution of the other. Although similar links have been observed for the classical RDF [20], [21], this observation has not been documented in this specific setting, hence we formalized it in the following theorem.

Theorem 1. *(Connection of OC-RDF and EOT) Let $f_X \in \mathcal{P}(\mathcal{X})$ and $f_Y \in \mathcal{P}(\mathcal{Y})$. Then, for any $D > 0$, there exists an $\epsilon > 0$ such that the problems of OC-RDF and EOT are equivalent.*

Proof: See [22, Theorem 1]. ■

In view of Theorem 1, we can treat the OC-RDF and EOT problems as equivalent problems. As a result, the computational schemes derived in Section III applicable to the OC-RDF problem, can be adapted *mutatis mutandis* to the EOT problem.

B. Copula distributions

In this subsection, we give some preliminaries to copulas distributions, as these have a central role in the derivation of the main results of this paper. The following definitions and theorems are taken from [23].

Definition 4. *(Copula distribution) For every $d \geq 2$, a d-dimensional copula d.f. is a d-variate d.f. on $[0,1]^d$ whose univariate marginals are uniformly distributed on $[0,1]$.*

The next theorem and the two companion corollaries, demonstrate that copulas are a powerful tool for the modeling and analysis of multivariate distributions.

Theorem 2. *(Sklar's Theorem) Let F be a d-dimensional d.f. with marginal d.f. $F_1, F_2, \dots, F_d$. Let A_j denote the range of F_j, $A_j \triangleq F_j(\bar{\mathbb{R}}) \quad (j = 1,2,\dots,d)$. Then, there exists a d-copula d.f. C such that for all $(x_1, x_2, \dots, x_d) \in \bar{\mathbb{R}}^d$,*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \tag{3}$$

Such a C is uniquely determined on $A_1 \times A_2 \times \cdots A_d$ and, hence, it is unique when $F_1, F_2, \dots, F_d$ are continuous.

Corollary 1. *Let $f : \bar{\mathbb{R}}^d \to \mathbb{R}^+$ be the pdf associated with* (3)*. Then, f can be uniquely decomposed as*

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{j=1}^{d} f_j(x_j) \tag{4}$$

where f_j is the pdf associated with the univariate marginal d.f. F_j and $c : [0,1]^d \to \mathbb{R}^+$ is the pdf associated with the copula d.f. C.

Corollary 2. *Let $F_1, F_2, \dots, F_d$ be univariate d.f.'s and C be a copula d.f.. Then, the function $F : \bar{\mathbb{R}}^d \to [0,1]$ defined in* (3) *is a d-dimensional d.f. with marginal $F_1, F_2, \dots, F_d$.*

It is worth noticing that Corollary 1 guarantees that the pdf of any multivariate distribution can be factorized as the product of the marginal densities and a unique copula distribution. This factorization can be effectively thought of as decoupling the correlation structure embedded in the joint distribution (represented by the copula distribution) from the information regarding each single marginal. On the other hand, Corollary 2 guarantees that, for a fixed set of marginals distributions, any copula distribution describes a proper joint distribution.

We conclude this subsection with the definition of the quantile function, which will also be of use in the derivation of our main results.

Definition 5. *(Quantile function) Let $X \sim F_X$ be a univariate RV on $\mathcal{X} \subseteq \mathbb{R}$. We define the quantile function $Q_X : [0,1] \to \mathbb{R}$ as $Q_X(u) \triangleq \sup\{x \in \mathcal{X} : F(x) \leq u\}$. If F_X is continuous and strictly increasing, then $Q_X = F_X^{-1}$. However, even if F_X may fail to have an inverse function, Q_X guaranties that $Q_X(F_X(X)) = X$ almost surely (a.s.).*

To ease the notation, in the sequel we denote by **uniform transformation** of an RV $X = (X_1, \dots, X_d)$ the function $\Phi_X : \mathcal{X} \to [0,1]^d$ defined as $\Phi_X(X) \triangleq (F_{X_1}(X_1), \dots, F_{X_d}(X_d))$. Moreover, we define the function $\Psi_X : [0,1]^d \to \mathcal{X}$ as $\Psi_X(U) \triangleq (Q_{X_1}(U_1), \dots, Q_{X_d}(U_d))$. By construction, Ψ_X is the a.s.-inverse of Φ_X, that is, $\Psi_X(\Phi_X(X)) = X$ a.s..

III. Main Results

In this section, we derive our main results.

A. Copula Lower Bound

First, we prove a lemma with which the functionals in the mathematical formulations of Definitions 2 and 3 can be redefined using copula distributions[1].

Lemma 1. *Let $(X,Y) \sim f_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a $2d$-variate RV with marginal pdfs $f_X \in \mathcal{P}(\mathcal{X})$ and $f_Y \in \mathcal{P}(\mathcal{Y})$. Then, the mutual information $I(X,Y)$ can be equivalently written as follows*

$$I(X,Y) = \mathrm{D_{KL}}(C_{X,Y} || C_X \otimes C_Y) \tag{5}$$

where $C_{X,Y}, C_X, C_Y$ are the copula d.f.'s associated with distributions $F_{X,Y}$, F_X, and F_Y, respectively. In addition, given a distortion function $\Delta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$, the following holds

$$\mathbb{E}_{F_{X,Y}}[\Delta(X,Y)] = \mathbb{E}_{C_{X,Y}}[\Delta(\Psi_X(U_X), \Psi_Y(U_Y))] \tag{6}$$

where $U = (U_X, U_Y) \sim C_{X,Y}$.

Proof: See [22, Lemma 1]. ■

Leveraging Lemma 1, we can provide an alternative formulation of the mathematical expression in (1), which will be the subject of our estimation analysis. This is stated next as Problem 1.

Problem 1. *(Copula-based OC-RDF) The mathematical expression* (1) *can be reformulated as follows*

$$R_{OC}(D) = \min_{C \in \mathcal{C}_{2d}} \mathrm{D_{KL}}(C || C_X \otimes C_Y) \tag{7}$$

$$\textit{s.t. } \mathbb{E}_C[\Delta(\Psi_X(U_X), \Psi_X(U_Y))] = D \tag{8}$$

where $\mathcal{C}_{2d}$ is the set of $2d$-copulas and $D \in [D_{\min}, D_{\max}]$.

Remark 2. *(On Problem 1) Problem 1 is a convex program in the space of copula d.f. Moreover, the problem is equivalent to finding the I-projection of $C_X \otimes C_Y$ on the set $\mathcal{B} \subset \mathcal{C}_{2d}$ of copula d.f. satisfying the modified distortion constraint* (8)*.*

Problem 1 represents a projection problem in information geometry, where the goal is to find the copula distribution C that minimizes the information divergence from the independent product copula $C_X \otimes C_Y$ while respecting a linear set of constraints. This class of projection problems has been thoroughly studied by Csiszár in [17], where the analytical form of the optimal projection for the considered case has been characterized. Using [17], we derive the following theorem.

Theorem 3. *(Analytical solution of Problem 1) Let $R = C_X \otimes C_Y$ and assume there exists a copula d.f. P such that*

[1]An alternative link between the mutual information $I(X,Y)$ and the associated copula entropy $h(C_{X,Y})$ can be found in [24].

$D_{KL}(P||R) < \infty$ *and* (8) *is satisfied. Then, Problem 1 admits a minimizing copula* Q *with Radon–Nikodym derivative with respect to the measure* R *of the form*

$$\frac{dC}{dR}(\mathbf{u}) = e^{\mu+\theta[\Delta(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))]} \prod_{i=1}^{2d} g_i(u_i) \tag{9}$$

for some constants (μ, θ), *and nonnegative uni-variate functions* g_i *such that* $\log(g_i(s)) \in l_1([0,1])$ *for* $i = 1, \dots, 2d$.

Proof: See [22, Theorem 3]. ■

Although Theorem 3 provides a characterization of the solution of Problem 1, the lack of an analytical form for the free functions $\{g_i(\cdot)\}_{i=1\dots,2d}$ poses a challenging problem in the computation of (9). Following an idea of [25], we circumvent this technical issue by introducing a relaxation on the constraint set of Problem 1, that results into a lower bound on OC-RDF. This is demonstrated next in Problem 2.

Problem 2. *(Lower bound to Problem 1) For any integer* N, *Problem 1 can be lower bounded as follows*

$$R_{OC}(D) \geq R_{OC}^{(N)} = \min_{\substack{Q\in\mathcal{P}([0,1]^{2d}) \\ \mathbb{E}[\Delta(\Psi_X(U_X),\Psi_Y(U_Y))]=D \\ \mathbb{E}_Q[u_i^n]=\alpha_n, \quad (i,n)\in I}} D_{KL}(Q||R)$$

where $R = C_X \otimes C_Y$, $I = (1, \dots, 2d) \times (1, \dots, N)$, $D \in [D_{\min}, D_{\max}]$, *and* α_n *is the* n^{th} *moment of a uniform distribution on* $[0, 1]$.

Remark 3. *(Problem 1 vs Problem 2) The main technical difference between Problems 1 and 2 concerns their constraint sets. Particularly, in Problem 1 we require that the minimizing distribution* Q^* *belongs to the set of copula distributions, which means that its marginals are uniformly distributed. On the other hand, the marginals of the minimizing distribution* $\hat{Q}_N^*$ *of Problem 2 only require to respect up to* N *moments of a uniform distribution. This in turn implies that the constraint set of Problem 1 is a proper subset of the constraint set of Problem 2, justifying the lower bound of the latter.*

In the following theorem, we show that, for $N \to \infty$, Problem 2 recovers the solution of Problem 1.

Theorem 4. *Let* Q^* *be the optimal solution of Problem 1 and* $\hat{Q}_N^*$ *be the optimal solution of Problem 2. Then, as* $N \to \infty$,

$$D_{KL}(\hat{Q}_N^*||Q^*) \to 0 \text{ and } R_{OC}^{(N)} \to R_{OC}.$$

Proof: See [22, Theorem 4]. ■

We now provide the analytical form of the solution of Problem 2. Unlike Theorem 3, the optimal solution does not depend on free functions $\{g_i(\cdot)\}_{i=1\dots,2d}$, but it depends only on the Lagrangian multipliers of Problem 2 obtained as result of its dual problem.

Theorem 5. *(Analytical solution of Problem 2) Let* $R = C_X \otimes C_Y$ *and assume there exists a d.f.* P *on* $[0,1]^{2d}$ *such that* $D_{KL}(P||R) < \infty$ *and* (8) *is satisfied. Then, Problem 2 admits minimizing copula* Q *with Radon–Nikodym derivative with respect to the measure* R *of the form*

$$\frac{dQ}{dR}(\mathbf{u}) = e^{\mu+\theta\Delta(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))} \prod_{i=1}^{2d} e^{\sum_{n=0}^{N}\nu_{i,n}u_i^n} \tag{10}$$

where the constants $(\mu, \theta, \{\nu_{i,n}\}_{(i,n)\in I})$ *are the Lagrangian multipliers of Problem 2 obtained as a result of the following dual program*

$$\min_{(\mu,\theta,\{\nu_{i,n}\}_{(i,n)\in I})} \mathbb{E}_R\left[\frac{dQ}{dR}\right] - \mu - \theta D - \sum_{(i,n)\in I} \nu_{i,n}\alpha_n \tag{11}$$

Proof: See [22, Theorem 5]. ■

The following result is a consequence of Theorem 5.

Corollary 3. *Let* Q *be the minimizing copula d.f. characterized in Theorem 5. Then, the mutual information* $I(X, Y)$ *of the joint distribution* (X, Y) *defined by marginals d.f.* $\{F_{X_i}\}_{i=1,\dots,d}$ *and* $\{F_{Y_i}\}_{i=1,\dots,d}$ *and copula* Q *is given by*

$$I(X, Y) = D_{KL}(Q||R) = -\mu - \theta D - \sum_{(i,n)\in I} \nu_{i,n}\alpha_n. \tag{12}$$

B. Copula Estimation

As anticipated in Theorem 5, the Lagrangian multipliers $(\mu, \theta, \{\nu_{i,n}\}_{(i,n)\in I})$ defining the optimal solution of Problem 2 can be obtained by solving (11). Although not available in closed form, the solution of (11) can be optimally computed using numerical methods, given the properties of the problem.

Lemma 2. *The optimization problem* (11) *is strictly convex, hence it has a unique solution.*

Proof: See [22, Lemma 2]. ■

To compute (11), we propose a low-complexity optimization scheme based on gradient methods. The main technical detail to clarify is related to the estimation of the integral present in (11), since numerically solving a possibly high dimensional integral could hinder the complexity of the algorithm. However, since its computation is required only for the estimation of the gradient and not for the computation of $I(X, Y)$ (as shown in (12)), we can approximate the integral using Monte Carlo method [26]. The resulting iterative scheme can be considered as a *mini-batch stochastic gradient descent algorithm* on a convex objective [27]. The algorithm is given in Alg. 1.

Algorithm 1 $R_{OC}(D)$ - Copula Estimation

Require: marginal distributions $\{F_{X_i}, F_{Y_i}\}_{i=1,\dots,d}$; distortion level D; number of iterations T; initial Lagrangian multipliers $\mathbf{l^{(0)}} = (\mu^{(0)}, \theta^{(0)}, \{\nu_{i,n}^{(0)}\}_{(i,n)\in I})$;

for i **do** = 1, …, T
    Sample $\{\mathbf{u}_i\}_{i=1\dots M}$ with $u_i \sim U([0,1]^{2d})$
    $f(\mathbf{l}) \approx (12) + \left(\frac{1}{M}\sum_{i=1}^{M} \frac{dQ}{dR}(\mathbf{l}, \mathbf{u}_i) dR(\mathbf{u}_i)\right)$
    $\mathbf{l^{(i)}} = \texttt{GradientMethod}(\mathbf{l}^{(i-1)}, f)$
end for

Ensure: Lagrangian multipliers $\mathbf{l}^{(T)}$; $I(X, Y) = (12)$.

C. SLB for PR-RDPF

In this subsection, we prove a generalization of the well-known SLB on the classical RDF with MSE distortion [28] to the case of PR-RDPF, denoted hereinafter by R_{PR}^{SLB}. The bound is stated in the following theorem.

Theorem 6. *(SLB for PR-RDPF) Let* $\mathcal{S} \triangleq \{f_X : \mathbb{E}_{f_X}\left[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T\right] \preceq \Sigma\}$ *be the set of source distribution with a fixed covariance matrix* Σ*. Then, for all* $X \sim f_X$ *with* $f_X \in \mathcal{S}$*, the PR-RDPF under MSE distortion constraint admits the following lower bound*

$$R_{PR}(D) \geq R_{PR}^{SLB}(D) = h(X) - h(X^*) + R_{PR}^{G}(D) \quad (13)$$

where $R_{PR}^{G}(D)$ *denotes the Gaussian PR-RDPF for a source* $X^* \sim N(0,\Sigma)$*.*

Proof: See [22, Theorem 6]. ■

We stress the following technical remark on Theorem 6.

Remark 4. *(On Theorem 6) For the scalar case of the PR-RDPF, let* $\mathcal{S} \triangleq \{f_X : \mathbb{E}_{f_X}\left[(X-\mathbb{E}[X])^2\right] \leq \sigma^2\}$ *for a finite variance value* σ^2*. Then,* (13) *can be further simplified to*

$$R_{PR}(D) \geq R_{PR}^{SLB}(D) = \frac{1}{2}\log\left(\frac{N(X)}{D - \frac{D^2}{4\sigma^2}}\right)$$

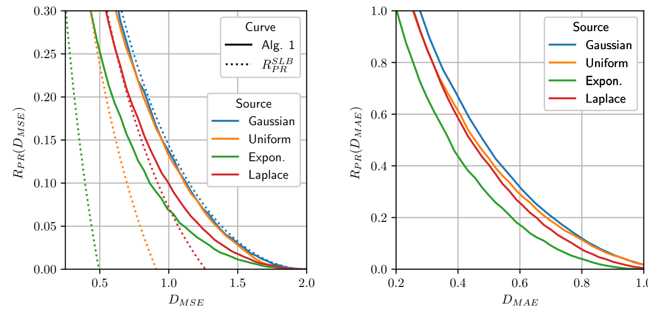
with $N(X)$ *denoting the entropy power of source* X*. For the general vector case, the lower bound depends on the vector Gaussian PR-RDPF,* R_{PR}^{G}*, which can be easily computed using the adaptive reverse-water-filling solution developed in [12, Corollary 3].*

IV. Numerical Results

In this section, we provide a numerical estimation of the PR-RDPF for both scalar and vector sources using Alg. 1.

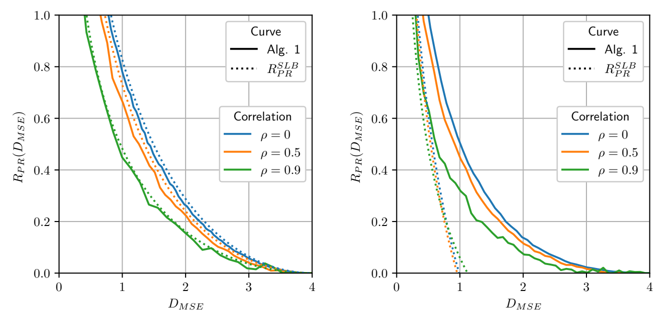
Scalar Case: We estimate the PR-RDPF for scalar sources under a single-letter constraint on the reconstruction error in terms of (a) the l_2 norm, i.e., the MSE distortion (see Fig. 1a), and (b) the l_1 norm i.e. the mean-absolute-error (MAE) distortion (see Fig. 1b). We compare the results for various source distributions, such as Gaussian, Laplace, exponential, and uniform, assuming that the source $X \sim (0,1)$, i.e., zero mean with variance $\sigma_X^2 = 1$. In Fig. 1a, we also compare the estimated result with the SLB derived in Theorem 6. In Fig. 1a, the Gaussian source case allows us to quantify the algorithm estimation accuracy by comparing it with the R_{PR}^{SLB}, which in this case represents the exact PR-RDPF. Regarding the other cases, the numerical results show that the bound R_{PR}^{SLB} behaves similarly to the SLB of the classical RDF, that is, being tight only in the low distortion (high resolution) regime, while becoming loose at the moderate to high distortion regimes.

Vector Case: We estimate the PR-RDPF under an MSE distortion metric for correlated bivariate sources, considering the cases where the source marginals are either Gaussian (see Fig. 2a) or exponentially (see Fig. 2b) distributed with zero mean and variance $\sigma^2 = 1$. In both cases, the multivariate distribution is constructed by imposing a Gaussian coupling [23] with variable correlation coefficient $\rho \in [0,1]$ on the considered marginal distributions. By changing ρ, we analyze the cases where the bivariate source presents independent ($\rho = 0$), mildly correlated ($\rho = 0.5$) and highly correlated ($\rho = 0.9$) marginals. In Fig. 2a, we demonstrate a comparison between the Gaussian PR-RDPF estimate obtained via Alg. 1 with the R_{PR}^{SLB} obtained in (13) with the term $R_{PR}^{G}(D)$ computed via the optimal adaptive reverse-water-filling solution of [12, Corollary 3], which results into a tight $R_{PR}^{SLB}(D)$. We observe that Alg. 1 provides a very good estimate of the Gaussian PR-RDPF for all the selected ρ. We also notice that the estimation error when using Alg. 1 remains stable in the low to moderate correlation cases while showing a slightly noisier behavior (fluctuations) in the high correlation case. Contrary to Fig. 2a, in Fig. 2b we observe that beyond high resolution (low distortion), the exponential PR-RDPF estimate obtained via Alg. 1 is much tighter compared to the R_{PR}^{SLB}. In fact, the latter demonstrates a similar behavior to the SLB of the classical RDF for the multivariate non-Gaussian case.



(a) (b)

Fig. 1: PR-RDPF for various source distributions under (a) MSE distortion metric and (b) MAE distortion metric.



(a) (b)

Fig. 2: PR-RDPF under MSE distortion metric for a (a) Gaussian, and (b) exponential bivariate source.

Acknowledgement

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 101003431).

REFERENCES

[1] Y. Blau and T. Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *International Conference on Machine Learning*. PMLR, 2019, pp. 675–685.

[2] 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.

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

[4] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[5] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 96–102, 2021.

[6] M. Li, J. Klejsa, and W. B. Kleijn, "On distribution preserving quantization," 2011. [Online]. Available: https://arxiv.org/abs/1108.3728

[7] N. Saldi, T. Linder, and S. Yüksel, "Randomized quantization and source coding with constrained output distribution," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 91–106, 2015.

[8] ——, "Output constrained lossy source coding with limited common randomness," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4984–4998, 2015.

[9] Y. Bai, X. Wu, and A. Özgür, "Information constrained optimal transport: From Talagrand, to Marton, to Cover," *IEEE Transactions on Information Theory*, vol. 69, no. 4, pp. 2059–2073, 2023.

[10] S. Wang, P. A. Stavrou, and M. Skoglund, "Generalizations of talagrand inequality for Sinkhorn distance using entropy power inequality," *Entropy*, vol. 24, no. 2, 2022.

[11] G. Zhang, J. Qian, J. Chen, and A. Khisti, "Universal rate-distortion-perception representations for lossy compression," *Advances in Neural Information Processing Systems*, vol. 34, pp. 11 517–11 529, 2021.

[12] G. Serra, P. A. Stavrou, and M. Kountouris, "On the computation of the Gaussian rate-distortion-perception function," *IEEE Journal on Selected Areas in Information Theory*, pp. 1–1, 2024.

[13] J. Qian, "On the rate-distortion-perception tradeoff for lossy compression," Ph.D. dissertation, McMaster University, October 2023, http://hdl.handle.net/11375/28976.

[14] G. Serra, P. A. Stavrou, and M. Kountouris, "Computation of rate-distortion-perception function under f-divergence perception constraints," in *Proc. IEEE Int. Symp. Inf. Theory*, 2023, pp. 531–536.

[15] C. Chen, X. Niu, W. Ye, S. Wu, B. Bai, W. Chen, and S.-J. Lin, "Computation of rate-distortion-perception functions with Wasserstein barycenter," *arXiv preprint arXiv:2304.14611*, 2023.

[16] O. Kirmemis and A. M. Tekalp, "A practical approach for rate-distortion-perception analysis in learned image compression," in *2021 Picture Coding Symposium (PCS)*, 2021, pp. 1–5.

[17] I. Csiszár, "I-Divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability*, vol. 3, no. 1, pp. 146 – 158, 1975.

[18] 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.

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

[20] D. Reshetova, Y. Bai, X. Wu, and A. Özgür, "Understanding entropic regularization in gans," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 825–830.

[21] E. Lei, H. Hassani, and S. S. Bidokhti, "On a relation between the rate-distortion function and optimal transport," *arXiv preprint arXiv:2307.00246*, 2023.

[22] G. Serra, P. A. Stavrou, and M. Kountouris, "Copula-based estimation of continuous sources for a class of constrained rate-distortion-functions," 2024. [Online]. Available: https://arxiv.org/abs/2401.17089

[23] F. Durante and C. Sempi, "Copula theory: An introduction," in *Copula Theory and Its Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 3–31.

[24] J. Ma and Z. Sun, "Mutual information is copula entropy," *Tsinghua Science and Technology*, vol. 16, no. 1, pp. 51–54, 2011.

[25] Y.-L. K. Samo, "Inductive mutual information estimation: A convex maximum-entropy copula approach," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 2242–2250.

[26] C. P. Robert, G. Casella, C. P. Robert, and G. Casella, "Monte carlo integration," *Monte Carlo statistical methods*, pp. 71–138, 1999.

[27] G. Garrigos and R. M. Gower, "Handbook of convergence theorems for (stochastic) gradient methods," 2023. [Online]. Available: https://arxiv.org/abs/2301.11235v2

[28] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, 1971.