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

Giuseppe Serra
Communication Systems Department
EURECOM, France
giuseppe.serra@eurecom.fr

Photios A. Stavrou
Communication Systems Department
EURECOM, France
fotios.stavrou@eurecom.fr

Marios Kountouris
DaSCI, Department of Computer Science and AI, University of Granada, Spain
Communication Systems Department, EURECOM, Sophia-Antipolis, France
mariosk@ugr.es, marios.kountouris@eurecom.fr

Abstract

We propose a novel method for estimating the rate-distortion-perception function in perfect realism regime (PR-RDPF) for a multivariate continuous source subject to a single-letter average distortion constraint. Our approach leads to a general computation scheme able to solve 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 is equivalent to an I-projection problem on a convex set, which allows us to recover the parametric solution of the optimal projection whose 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 support our theoretical findings with numerical examples by assessing the estimation performance of our scheme.

1 Introduction

Rate-distortion-perception (RDP) theory, proposed by Blau and Michaeli [3] and Matsumoto [13, 14], is a novel generalization of the classical rate-distortion framework [23], tailored to consider the compression of complex data sources (e.g., audio, images, video) when perceptual quality is taken into account. This generalization considers the classical rate-distortion function (RDF) formulation and imposes an additional divergence constraint between the source distribution and its reconstruction. The divergence constraint acts as a proxy for human perception, capturing the difference between the reconstructed samples and the source "natural statistic" [15]. Alternatively, it can also be interpreted as a semantic quality metric measuring the relevance of the reconstructed source from the receiver's perspective [11].

Previous studies on the connection between the statistical properties of the reconstructed samples and their perceived quality led to the formulation of the so-called *output-constrained rate-distortion problem*. 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 shows close similarities with the EOT problem [1, 26], as, in both problems, the source and the reconstruction distributions are assumed 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

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

the general complexity, closed-form expressions have been developed under different settings [3, 27, 22, 16]. 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 [21] or by discretizing certain classes of continuous sources [4]. For general sources, RDPF estimation methods often rely on data-driven solutions [3, 27, 9], which unfortunately do not have convergence guarantees.

Contributions With this work, our objective is the development of a novel estimation method for the PR-RDPF for multivariate continuous sources subject to a single-letter average distortion constraint. As a result, our approach leads to a generic computational scheme for the EOT and the OC-RDF, of which the PR-RDPF represents a particular case. The main contributions of this paper are the following. **(i)** We establish the existence of 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 cast the OC-RDF 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 [5], 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 Algorithm 1). We supplement our theoretical results with various numerical evaluations aiming to estimate the PR-RDPF under various sources and different distortion measures via Algorithm 1.

Notation We indicate with $\mathbb{R}$ the set of real numbers, with $\bar{\mathbb{R}}$ the extended set $\mathbb{R} \cup \{-\infty, +\infty\}$. Given a Euclidean space (possibly finite-dimensional) $\mathcal{X}$, we denote by $(\mathcal{X}, \mathbb{B}(\mathcal{X}))$ the Borel measurable space induced by the metric, with $\mathcal{P}(\mathcal{X})$ denoting the set of distribution functions defined thereon. For a random variable (RV) X defined on $(\mathcal{X}, \mathbb{B}(\mathcal{X}))$, 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 $\mathrm{D_{KL}}(F_X||F_Y)$ 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 $\mathcal{A} \in \mathbb{R}^n$, we will denote with $l_p(\mathcal{A})$ the set of functions $g : \mathcal{A} \to \mathbb{R}$ such that $\int_{\mathcal{A}} |g(s)|^p ds < \infty$.

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

We begin this subsection by providing the mathematical definition of OC-RDPF, formally introduced by Saldi *et al.* in [19].

Definition 1. *(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*

$$R_{OC}(D) = \min_{f_{Y|X} \in \hat{\Pi}(f_X, f_Y)} I(X,Y) \quad \textit{s.t.} \quad \mathbb{E}\left[\Delta(X,Y)\right] \leq D \tag{1}$$

where $\hat{\Pi}$ is 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 R_{OC} particularizes to the problem of PR-RDPF R_{PR} by specifying the reconstruction distribution to be equal to the source distribution (i.e. $f_Y = f_X$).*

It should be noted that the PR-RDPF represents a limit case of the general problem of the RDPF [3]. Although it became quite popular through [3], similar ideas were previously explored by Li *et. al.* in [12], in the context of distribution-preserving quantization and distribution-preserving RDF. Multiple coding theorems have been developed for PR-RDPF. Under the assumption of infinite common randomness between the encoder and decoder, Theis and Wagner in [25] prove a coding theorem for stochastic variable-length codes in both one-shot and asymptotic regimes. Saldi *et. al.*, in [19], provide coding theorems focusing on the case where only finite common randomness between encoder and decoder is available.

Additionally, the OC-RDF highlights an interesting connection to the EOT problem (see [1, 26]), of which the mathematical definition is stated as follows.

Definition 2. *(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*

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

where $\bar{\Pi}$ is 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 bijection 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. We formalize this observation 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 OE-RDF and EOT are equivalent.*

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 2 applicable to the OC-RDF problem, can be adapted *mutatis mutandis* to the EOT problem.

1.2 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 [6].

Definition 3. *(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. $F_1, F_2, \ldots, F_d$. Then, there exists a d-copula d.f. C such that for all $(x_1, x_2, \ldots, x_d) \in \bar{\mathbb{R}}^d$,*

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right) \tag{3}$$

with C being uniquely determined on $[0, 1]^d$ iff $F_1, F_2, \ldots, F_d$ are continuous.

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

$$f(x_1, \ldots, x_d) = c\left(F_1(x_1), \ldots, F_d(x_d)\right) \prod_{j=1}^{d} f_j(x_j)$$

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 $F_1, F_2, \ldots, F_d$ 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 $F_1, F_2, \ldots, F_d$.*

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 inducing a decoupling between the correlation structure embedded in the joint distribution (represented by the copula distribution) and the information regarding each single marginal. On the other hand, Corollary 2 shows the modeling capabilities of copulas, where, given a set of marginal 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 4. *(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.

2 Copula Lower Bound

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

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*

$$I(X, Y) = \mathrm{D_{KL}}(C_{X,Y} || C_X \otimes C_Y) \tag{4}$$

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

$$\mathbb{E}_{F_{X,Y}}[\Delta(X, Y)] = \mathbb{E}_{C_{X,Y}}[\Delta(\Psi_X(U_X), \Psi_Y(U_Y))] \tag{5}$$

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

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*

$$R_{OC}(D) = \min_{C \in \mathcal{C}_{2d}} \mathrm{D_{KL}}(C || C_X \otimes C_Y) \tag{6}$$

$$\text{s.t. } \mathbb{E}_C[\Delta(\Psi_X(U_X), \Psi_Y(U_Y))] = D \tag{7}$$

where $\mathcal{C}_{2d}$ is the set of $2d$-copula distributions 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* (7).

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 [5], where the analytical form of the optimal projection for the considered case has been characterized. Using [5], 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 $\mathrm{D_{KL}}(P||R) < \infty$ and* (7) *is satisfied. Then, Problem 1 admits a minimizing copula Q with Radon–Nikodym derivative with respect to the measure R of the form*

$$\frac{dC}{dR}(\mathbf{u}) = e^{\mu + \theta[\Delta(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y))]} \prod_{i=1}^{2d} g_i(u_i) \tag{8}$$

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

Although Theorem 3 provides a characterization of the solution of Problem 1, the lack of an analytical form for the free functions $\{g_i(\cdot)\}_{i=1\ldots,2d}$ poses a challenging problem in the computation of (8). Following an idea of [20], we circumvent this technical issue via 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 by*

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

where $R = C_X \otimes C_Y$, $I = (1,\ldots,2d)\times(1,\ldots,N)$, $D \in [D_{\min}, D_{\max}]$, 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 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 $N \to \infty$,$\mathrm{D_{KL}}(\hat{Q}_N^*||Q^*) \to 0$ and $R_{OC}^{(N)} \to R_{OC}$.*

We now provide the analytical form of the solution of Problem 2. Unlike Theorem 3, the optimal solution does not depend on free functions $\{g_i(\cdot)\}_{i=1\ldots,2d}$, 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 $\mathrm{D_{KL}}(P||R) < \infty$ and (7) is satisfied. Then, Problem 2 admits minimizing copula Q with Radon–Nikodym derivative with respect to the measure R of the form*

$$\frac{dQ}{dR}(\mathbf{u}) = e^{\mu+\theta\Delta(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))}\prod_{i=1}^{2d} e^{\sum_{n=0}^{N}\nu_{i,n}u_i^n} \tag{9}$$

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

$$\min_{(\mu,\theta,\{\nu_{i,n}\}_{(i,n)\in I})} -\mu - \theta D - \sum_{(i,n)\in I}\nu_{i,n}\alpha_n + \left(\int_{[0,1]^{2d}}\frac{dQ}{dR}(\mathbf{u})dR(\mathbf{u}) - 1\right). \tag{10}$$

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. $\{F_{X_i}\}_{i=1,\ldots,d}$ and $\{F_{Y_i}\}_{i=1,\ldots,d}$ and copula Q is given by*

$$I(X,Y) = \mathrm{D_{KL}}(Q||R) = -\mu - \theta D - \sum_{(i,n)\in I}\nu_{i,n}\alpha_n. \tag{11}$$

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 (10). Although not available in closed form, the solution of (10) can be optimally computed using numerical methods, given the properties of the problem.

Lemma 2. *The optimization problem* (10) *is strictly convex, hence it has a unique solution.*

To compute (10), 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 (10), 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 (11)), we can approximate the integral using Monte Carlo method [17]. The resulting iterative scheme can be considered as a *mini-batch stochastic gradient descent algorithm* on a convex objective [7]. The algorithm is given in Algorithm 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 $\mathbf{l}^{(\mathbf{0})} = (\mu^{(0)}, \theta^{(0)}, \{\nu_{i,n}^{(0)}\}_{(i,n)\in I})$;

for i **do** $= 1, \dots, T$

  Sample $\{\mathbf{u}_i\}_{i=1\dots M}$ with $u_i \sim U([0,1]^{2d})$

  $f(\mathbf{l}) \approx (11) + \left(\frac{1}{M}\sum_{i=1}^{M} \frac{dQ}{dR}(\mathbf{l}, \mathbf{u}_i) dR(\mathbf{u}_i)\right)$

  $\mathbf{l}^{(\mathbf{i})} = \texttt{GradientMethod}(\mathbf{l}^{(i-1)}, f)$

end for

Ensure: Lagrangian multipliers $\mathbf{l}^{(T)}$; $I(X,Y) = (11)$.

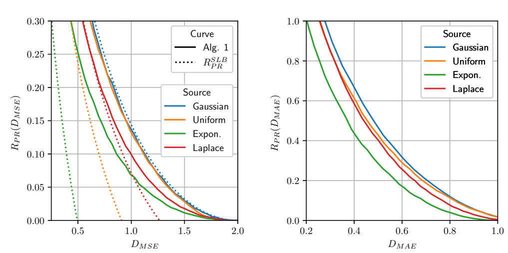
---

3 Numerical Results

In this section, we provide numerical estimation of the PR-RDPF for both scalar and vector sources using Algorithm 1. In the cases where the MSE distortion metric is considered, we compare the estimated result with the analytically available Shannon Lower Bound R_{PR}^{SLB} derived in Appendix G.

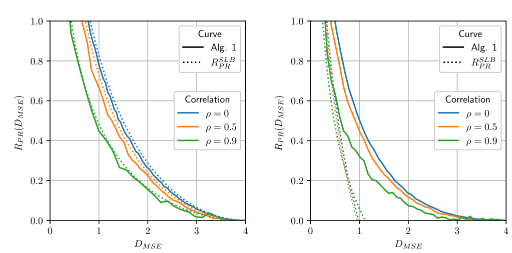
Scalar Case We estimate the PR-RDPF for scalar sources under a single-letter constraint on the reconstruction error in terms of (a) the l_2 norm, i.e., the MSE distortion (see Fig. 1a), and (b) the l_1 norm i.e. the mean-absolute-error (MAE) distortion (see Fig. 1b). We consider various source distributions, such as Gaussian, Laplace, exponential, and uniform, assuming that the source $X \sim \mathcal{G}(0,1)$ is normalized, i.e., zero mean with unitary variance. 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, being tight only in the low distortion (high resolution) regime.

Vector Case We estimate the PR-RDPF under an MSE distortion metric for correlated bivariate sources, considering the cases where the source marginals are normalized and either Gaussian (see Fig. 2a) or exponentially (see Fig. 2b) distributed. In both cases, the multivariate distribution is constructed by imposing a Gaussian coupling[1] with variable correlation coefficient $\rho \in [0,1]$. 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. Similarly to the scalar case, in Fig. 2a we compare the Gaussian PR-RDPF estimate obtained via Alg. 1 with the R_{PR}^{SLB}, which in this case represents the exact PR-RDPF, observing very good estimate of the Gaussian PR-RDPF for all the selected ρ. 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}.



(a) (b)

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



(a) (b)

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

[1] For more details on parametric copula models, we refer the reader to [6].

4 Acknowledgment

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. Bai, X. Wu, and A. Özgür. Information constrained optimal transport: From Talagrand, to Marton, to Cover. *IEEE Transactions on Information Theory*, 69(4):2059–2073, 2023.

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

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

[4] 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.

[5] I. Csiszár. I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, 3(1):146 – 158, 1975.

[6] F. Durante and C. Sempi. Copula theory: An introduction. In *Copula Theory and Its Applications*, pages 3–31, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

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

[8] I. M. Gelfand, R. A. Silverman, et al. *Calculus of variations*. Courier Corporation, 2000.

[9] O. Kirmemis and A. M. Tekalp. A practical approach for rate-distortion-perception analysis in learned image compression. In *2021 Picture Coding Symposium (PCS)*, pages 1–5, 2021. doi: 10.1109/PCS50896.2021.9477479.

[10] A. Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

[11] M. Kountouris and N. Pappas. Semantics-empowered communication for networked intelligent systems. *IEEE Commun. Mag.*, 59(6):96–102, 2021.

[12] M. Li, J. Klejsa, and W. B. Kleijn. On distribution preserving quantization, 2011. URL `https://arxiv.org/abs/1108.3728`.

[13] R. Matsumoto. Introducing the perception-distortion tradeoff into the rate-distortion theory of general information sources. *IEICE Comm. Express*, 7(11):427–431, 2018.

[14] R. Matsumoto. Rate-distortion-perception tradeoff of variable-length source coding for general information sources. *IEICE Comm. Express*, 8(2):38–42, 2019.

[15] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. doi: 10.1109/LSP.2012.2227726.

[16] J. Qian. *On the Rate-Distortion-Perception Tradeoff for Lossy Compression*. PhD thesis, McMaster University, October 2023. http://hdl.handle.net/11375/28976.

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

[18] R. T. Rockafellar. *Convex analysis*, volume 18. Princeton university press, 1970.

[19] N. Saldi, T. Linder, and S. Yüksel. Randomized quantization and source coding with constrained output distribution. *IEEE Transactions on Information Theory*, 61(1):91–106, 2015. doi: 10.1109/TIT.2014.2373382.

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

[21] 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*, pages 531–536, 2023.

[22] 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*, pages 1–1, 2024.

[23] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *Institute of Radio Engineers, National Convention Record*, 4:142–163, 1993.

[24] J. A. Shohat and J. D. Tamarkin. *The problem of moments*, volume 1. American Mathematical Society (RI), 1950.

[25] L. Theis and A. B. Wagner. A coding theorem for the rate-distortion-perception function. In *International Conference of Learning Representations (ICLR)*, pages 1–5, 2021.

[26] S. Wang, P. A. Stavrou, and M. Skoglund. Generalizations of talagrand inequality for Sinkhorn distance using entropy power inequality. *Entropy*, 24(2), 2022.

[27] G. Zhang, J. Qian, J. Chen, and A. Khisti. Universal rate-distortion-perception representations for lossy compression. *Advances in Neural Information Processing Systems*, 34:11517–11529, 2021.

A Proof of Theorem 1

We start by showing that $\hat{\Pi}$ and $\bar{\Pi}$ in Definitions 1 and 2, define essentially the same set, i.e., there exists a bijection between two sets. Assuming f_X to be the source distribution in OC-RDF, then for any Markov kernel $f_{Y|X} \in \hat{\Pi}$, the joint pdf $f_{Y|X} \cdot f_X$ lies in $\bar{\Pi}$. Conversely, for any joint distribution $f_{X,Y} \in \bar{\Pi}$ the Markov kernel $f_{Y|X} = \frac{f_{X,Y}}{f_X}$ belongs to $\hat{\Pi}$. Hence, there is a one-to-one mapping between the optimization variables of Definitions 1 and 2.
Let (D, λ) be the pair composed by the distortion level in the constraint of (1) and the associated Lagrangian multiplier. Then, the Lagrangian functional of Definition 1 for distortion level D is defined as

$$\mathcal{L}_{RD}(f_{Y|X}, \lambda) \triangleq I(X,Y) + \lambda \mathbb{E}_{f_{Y|X} f_X} \left[\Delta(X,Y)\right]. \tag{12}$$

Similarly, the Lagrangian functional associated with Definition 2 is defined as

$$\mathcal{L}_{EOT}(f_{X,Y}, \epsilon) \triangleq \mathbb{E}_{f_{X,Y}} \left[\Delta(X,Y)\right] + \epsilon I(X,Y). \tag{13}$$

Based on (12), (13), we observe that the following relation holds

$$\mathcal{L}_{RD}\left(\frac{f_{X,Y}}{f_X}, \lambda\right) = \mathcal{L}_{EOT}\left(f_{X,Y}, \frac{1}{\lambda}\right)$$

hence

$$\arg\min_{f_{X,Y} \in \bar{\Pi}} \mathcal{L}_{EOT}\left(f_{X,Y}, \frac{1}{\lambda}\right) = \arg\min_{f_{X,Y} \in \bar{\Pi}} \mathcal{L}_{RD}\left(\frac{f_{X,Y}}{f_X}, \lambda\right) = f_X \cdot \arg\min_{f_{Y|X} \in \hat{\Pi}} \mathcal{L}_{RD}(f_{Y|X}, \lambda).$$

As a result, (14) shows that the solution of Definition 2 for $\epsilon = \frac{1}{\lambda}$ is uniquely determined by the solution of Definition 1 for the pair (D, λ). This completes the proof.

B Proof of Lemma 1

From the definitions of $I(X,Y)$ and $\mathbb{E}_{F_{X,Y}}\left[\Delta(X,Y)\right]$, (4) can be derived as

$$\begin{aligned}
I(X,Y) &= \int_{\mathbb{R}^{2d}} f_{XY}(\mathbf{x},\mathbf{y}) \log\left(\frac{f_{XY}(\mathbf{x},\mathbf{y})}{f_X(\mathbf{x}) f_Y(\mathbf{y})}\right) d\mathbf{x}d\mathbf{y} \\
&\overset{(a)}{=} \int_{\mathbb{R}^{2d}} c_{X,Y}(\Phi_X(\mathbf{x}), \Psi_Y(\mathbf{y})) \log\left(\frac{c_{X,Y}(\Phi_X(\mathbf{x}), \Psi_Y(\mathbf{y}))}{c_X(\Phi_X(\mathbf{x})) c_Y(\Psi_Y(\mathbf{y}))}\right) \prod_{i=1}^{d} dF_{X_i}(x_i) dF_{Y_i}(y_i) \\
&\overset{(b)}{=} \int_{[0,1]^{2d}} c_{X,Y}(\mathbf{u}_x, \mathbf{u}_y) \log\left(\frac{c_{X,Y}(\mathbf{u}_x, \mathbf{u}_y)}{c_X(\mathbf{u}_x) c_Y(\mathbf{u}_y)}\right) d\mathbf{u}_x d\mathbf{u}_y \\
&= \mathrm{D_{KL}}(C_{X,Y} || C_X \otimes C_Y)
\end{aligned}$$

and (5) is similarly obtained by

$$\begin{aligned}
\mathbb{E}_{F_{X,Y}}\left[\Delta(X,Y)\right] &= \int_{\mathbb{R}^{2d}} \Delta(\mathbf{x},\mathbf{y}) f_{XY}(\mathbf{x},\mathbf{y}) d\mathbf{x}d\mathbf{y} \\
&\overset{(a)}{=} \int_{\mathbb{R}^{2d}} \Delta(\mathbf{x},\mathbf{y}) c_{X,Y}(\Phi_X(\mathbf{x}), \Phi_Y(\mathbf{y})) \prod_{i=1}^{d} dF_{X_i}(x_i) dF_{Y_i}(y_i) \\
&\overset{(b)}{=} \int_{[0,1]^2} \Delta\left(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y)\right) c_{X,Y}(\mathbf{u}_x, \mathbf{u}_y) d\mathbf{u}_x d\mathbf{u}_y \\
&= \mathbb{E}_{c_{X,Y}}\left[\Delta\left(\Psi_X(U_X), \Psi_Y(U_Y)\right)\right]
\end{aligned}$$

where, in both derivations, (a) follows from the application of Corollary 1 on $f_{X,Y}$,f_X, and f_Y, and (b) follows from the change of variables $\mathbf{u}_x = \Phi_X(\mathbf{x})$ and $\mathbf{u}_y = \Phi_Y(\mathbf{y})$.

C Proof of Theorem 3

The proof follows similar steps to the proof of [20, Theorem 3.1] with some technical differences, hence at certain points we skip the heavy mathematical details for ease of readability. In particular, we project on the product copula d.f. R, instead of the I-projection of the uniform distribution U on $[0,1]^{2d}$, which is considered in [20, Theorem 3.1]. First, we inquire about the existence of the projection.

Existence and uniqueness Under the assumption of our Theorem that there exist $P \in \mathcal{B}$ with $\mathrm{D}_{\mathrm{KL}}(P||R) < \infty$, if the convex set $\mathcal{B}$ is variation closed, i.e., closed in the topology induced by the total variation distance [10, Corollary 7.45], then there exists a unique Q being the I-projection of R on $\mathcal{B}$. This property of the set $\mathcal{B}$ can be proved using [20, Lemma B.1].

Parametric form of the density of the projection The projection task can be facilitated by defining an intermediate projection step onto the set $\mathcal{A}$ that corresponds to the set of d.f. on $[0,1]^d$ satisfying the distortion constraint (7). Clearly, in this case $\mathcal{A}$ is convex and $\mathcal{B} \subset \mathcal{A}$.

Since the set $\mathcal{A}$ is defined by linear constraints, then following [5, Theorem 3.1, (Case A)], we obtain that $R_{\mathcal{A}}$ is the unique I-projection of R onto $\mathcal{A}$ with density

$$\frac{dR_{\mathcal{A}}}{dR}(\mathbf{u}) = e^{\mu+\theta\Delta(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))} \tag{14}$$

and, for all $P \in \mathcal{A}$, holds that

$$\mathrm{D}_{\mathrm{KL}}(P||R) = \mathrm{D}_{\mathrm{KL}}(P||R_{\mathcal{A}}) + \mathrm{D}_{\mathrm{KL}}(R_{\mathcal{A}}||R). \tag{15}$$

Moreover, under the result of [5, Theorem 2.3], if R has I-projection $R_{\mathcal{A}}$ on $\mathcal{A}$, and I-projection $Q_{\mathcal{B}}$ on $\mathcal{B}$, and if (15) holds for all $P \in \mathcal{A}$, then $Q_{\mathcal{B}}$ is the unique I-projection of $Q_{\mathcal{A}}$ onto $\mathcal{B}$.

Since the set $\mathcal{B}$ is defined by imposing a constraint on the marginals of the measure, and $Q_{\mathcal{A}}$ has I-projection on $\mathcal{B}$, then using [5, Theorem 3.1, (Case B)], there exist nonnegative scalar functions g_i with $\log(g_i) \in l_1([0,1])$ for $i = 1, \ldots, 2d$ such that

$$\frac{dQ}{dR_{\mathcal{A}}}(\mathbf{u}) = \prod_{i=1}^{2d} g_i(u_i).$$

Therefore, Q has density with respect to R given by $\frac{dQ}{dR} = \frac{dQ}{dR_{\mathcal{A}}}\frac{dR_{\mathcal{A}}}{dR} =$ (8). This completes the proof.

D Proof of Theorem 4

Let $\hat{Q}_i$ and $\mathcal{A}_i$ be, respectively, the copula distribution solution of Problem 2 and its constraint set, for a number i of constraints on the moments of each marginal. Furthermore, let the constraint set and optimal solution of Problem 1 be denoted with $\mathcal{B}$ and Q^*, respectively. Our goal is to prove that the sequence $\{\hat{Q}_i\}_{i=0,1,\ldots}$ convergences to Q^*.

By construction, for all $i = 0, 1, \ldots$, it holds that $\mathcal{B} \subseteq \mathcal{A}_{i+1} \subseteq \mathcal{A}_i$. As a consequence of [5, Theorem 2.3], we can characterize $\hat{Q}_{i+1}$ and Q^* as the I-projections of $\hat{Q}_i$ onto the sets $\mathcal{A}_{i+1}$ and $\mathcal{B}$, respectively. Then, for all $i = 0, 1, \ldots$, the following geometric relation holds (see [5, Equation 3.1])

$$\mathrm{D}_{\mathrm{KL}}(\hat{Q}_i||Q^*) = \mathrm{D}_{\mathrm{KL}}(\hat{Q}_i||\hat{Q}_{i+1}) + \mathrm{D}_{\mathrm{KL}}(\hat{Q}_{i+1}||Q^*). \tag{16}$$

Recursively, applying (16) $k+1$ times leads to

$$\mathrm{D}_{\mathrm{KL}}(\hat{Q}_i||Q^*) = \sum_{j=i}^{i+k} \mathrm{D}_{\mathrm{KL}}(\hat{Q}_j||\hat{Q}_{j+1}) + \mathrm{D}_{\mathrm{KL}}(\hat{Q}_{k+1}||Q^*)$$

from which we immediately obtain

$$\sum_{j=i}^{i+k} \mathrm{D}_{\mathrm{KL}}(\hat{Q}_j||\hat{Q}_{j+1}) \leq \mathrm{D}_{\mathrm{KL}}(\hat{Q}_i||Q^*).$$

Since we assume that $\mathrm{D_{KL}}(\hat{Q}_i||Q^*) < \infty$, then, necessarily $\lim_{k\to\infty} \mathrm{D_{KL}}(\hat{Q}_k||\hat{Q}_{k+1}) = 0$, implying the convergence of the sequence $\{\hat{Q}_i\}_{i=0,1,\dots}$ in KL-divergence.

To prove that the limit of the sequence is Q^*, we transform Problem 1 into a specific instance of Problem 2 under an infinite number of marginals moments constraints. As a consequence of the uniqueness of solutions of the Hausdorff moments problem [24], any RV on $[0,1]$ that respects $\mathbb{E}[u^n] = \alpha_n$ for all $n \in \mathbb{N}$, is necessarily uniformly distributed. This allows us to transform the uniform marginal constraints in Problem 1 into a set of countably infinite marginal constraints.

In such form, the only difference between Problems 1 and 2 resides in the finite number i of the moment constraints of the latter. Hence, since $Q^* = \hat{Q}_\infty$, the limit $\lim_{i\to\infty} \mathrm{D_{KL}}(\hat{Q}_i||Q^*) = 0$ holds. This completes the proof.

E Proof Theorem 5

Let $\mathcal{M}^+([0,1]^{2d})$ denote the set of measurable functions on $[0,1]^{2d}$. We notice that the constraint set $\mathcal{A}$ of Problem 2 is defined by linear constraints in the copula d.f. $C_{X,Y}$. The results of [5, Theorem 3.1, (Case A)] ensure that a unique projection Q of R on $\mathcal{A}$ exists with $\mathrm{D_{KL}}(Q||R) < \infty$, hence $\frac{dQ}{dR} \in \mathcal{M}^+([0,1]^{2d})$. More generally, any function $g \in \mathcal{M}^+([0,1]^{2d})$ defines a probability measure $dG = gdR$ on $[0,1]^{2d}$ under the condition that $\int_{[0,1]^{2d}} g(\mathbf{u})dR(\mathbf{u}) = 1$. This enables the definition of an optimization problem over the set $\mathcal{M}^+([0,1]^{2d})$ equivalent to Problem 2 as follows

$$\min_{g\in\mathcal{M}^+([0,1]^{2d})} \int_{[0,1]^{2d}} \log\left(g(\mathbf{u})\right) g(\mathbf{u})dR(\mathbf{u}). \tag{17}$$

$$\text{s.t.} \quad \int_{[0,1]^{2d}} g(\mathbf{u})dR(\mathbf{u}) = 1 \tag{18}$$

$$\int_{[0,1]^{2d}} \Delta(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y))g(\mathbf{u})dR(\mathbf{u}) = D \tag{19}$$

$$\int_{[0,1]^{2d}} u_i^n g(\mathbf{u})dR(\mathbf{u}) = \alpha_n \qquad (i,n) \in I \tag{20}$$

Indicating with $\mu', \theta, \{\nu_{i,n}\}$ the Lagrangian multipliers associated with constraints (18)-(20), we define the Lagrangian functional of the problem as

$$L(g,\mu',\theta,\{\nu_{i,n}\}) = \int S(g(\mathbf{u}),\mu',\theta,\{\nu_{i,n}\})dR(\mathbf{u}) + V(\mu',\theta,\{\nu_{i,n}\}) \tag{21}$$

$$S(z,\mu',\theta,\{\nu_{i,n}\}) = z\Big[\log(z) - \mu' - \sum_{(i,n)\in I} \nu_{i,n}u_i^n - \theta\Delta(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))\Big]$$

$$V(\mu',\theta,\{\nu_{i,n}\}) = \mu' + \theta D + \sum_{(i,n)\in I} \nu_{i,n}\alpha_n.$$

By applying the Euler-Lagrange equation [8], we characterize necessary conditions for the function g to be an extreme point for (21). If g^* is an extreme point of (21) , then the necessary *stationarity condition* holds, i.e.,

$$\frac{dS}{dz}(g^*) = 0$$

from which we obtain

$$g^*(\mathbf{u}) = \frac{dQ}{dR}(\mathbf{u}) = \exp\Big[(\mu'-1) + \sum_{(i,n)\in I} \nu_{i,n}u_i^n + \theta\Delta(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))\Big], \tag{22}$$

which is equivalent to (9) by considering $\mu = \mu' - 1$.

To determine the optimal values of the Lagrangian multipliers $\mu', \theta, \{\nu_{i,n}\}$, we leverage Lagrangian duality theorem [18] and define the dual problem (10) as

$$(10) = \min_{(\mu,\theta,\{\nu_{i,n}\}_{(i,n)\in I})} -L(g^*, \mu+1, \theta, \{\nu_{i,n}\}_{(i,n)\in I}).$$

This concludes the proof.

F Proof Lemma 2

The proof makes use of the Hessian matrix of the optimization problem in (10). Specifically, define the vector $\mathbf{l} \triangleq ((\mu, \theta, \{\nu_{i,n}\}_{(i,n)\in I})) \in \mathbb{R}^{2+2d\cdot N}$ and the mapping $\omega(\mathbf{u}) \triangleq (1, \Delta(\Psi_X(\mathbf{u}_x), \Psi_X(\mathbf{u}_y)), u_1, \dots, u_1^N, \dots, u_{2d}^N)$. Then, the Hessian of the optimization problem (10) can be expressed as

$$\int_{I^{2d}} \omega(\mathbf{u})\omega(\mathbf{u})^T dQ = \mathbb{E}_Q\left[\omega(\mathbf{u})\omega(\mathbf{u})^T\right],$$

which, due to the linear independence of the components of $\omega(\mathbf{u})$, is a strictly positive definite matrix. This is a sufficient condition for the strict convexity of (10). This completes the proof.

G Shannon Lower Bound (SLB) for PR-RDPF

In this appendix, we derive a generalization of the well-known SLB on the classical RDF with MSE distortion [2] 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*

$$R_{PR}(D) \geq R_{PR}^{SLB}(D) = h(X) - h(X^*) + R_{PR}^{G}(D) \tag{23}$$

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

Proof. We start by considering the scalar version of the proposed problem, i.e., $\mathcal{S} \triangleq \{f_X : \mathbb{E}_{f_X}\left[(X - \mathbb{E}[X])^2\right] \leq \sigma^2\}$. Define the constraint set $\mathcal{H}(f_X, D)$ as follows

$$\mathcal{H}(f_X, D) = \{f_{Y|X} : \mathbb{E}_{f_X f_{Y|X}}\left[||X - Y||^2\right] \leq D, X \sim Y\}.$$

Then, by definition of the PR-RDPF, we obtain

$$\begin{aligned} R_{PR}(D) &= h(X) - \max_{f_{Y|X}\in\mathcal{H}(f_X,D)} h(X|Y) \\ &\overset{(a)}{=} h(X) - \max_{f_{Y|X}\in\mathcal{H}(f_X,D)} h(Y|X) \\ &\overset{(b)}{\geq} h(X) - \max_{f_Z\in\mathcal{S}} \max_{f_{Y|Z}\in\mathcal{H}(f_Z,D)} h(Y|Z) \end{aligned} \tag{24}$$

where (a) follows by observing that $h(X|Y) = h(Y|X)$ under the constraint $X \sim Y$, and (b) follows from the maximization over the set of sources $\mathcal{S}$.

Consider the joint distribution on (Z, Y) with $Z \sim Y$. Then, the conditional variance of the RV Y conditioned on Z can be expressed as $\sigma_{Y|Z}^2 = \sigma_Z^2(1 - \rho^2)$, hence the constraint set $\mathcal{H}(f_Z, D)$ can be simplified to

$$\begin{aligned} &\mathbb{E}_{f_Z f_{Y|Z}}\left[||Z - Y||^2\right] = 2\sigma_Z^2(1 - \rho) \leq D \\ \implies\ &\sigma_{Y|Z}^2 \leq \sigma_Z^2\left(1 - \left(1 - \frac{D}{\sigma_Z^2}\right)^2\right) = D'. \end{aligned} \tag{25}$$

Since (25) constraints only the second moment of the distribution $f_{Y|Z}$, we can infer that the maximum $h(Y|Z)$ is attained by $f_{Y|Z} \sim N(0, D')$. Furthermore, since (25) depends only on the second moment of the source σ_Z^2, the maximization over the set of source distributions has to satisfy only the distribution constraint $Z \sim Y$. Assuming $Y|Z$ to be Gaussian, we can select Z to be also Gaussian distributed with $f_Z \sim N(0, \sigma^2)$, which ensures that Y will have the same distribution. Therefore, the distribution on (Y, Z) maximizing $h(Y|Z)$ is itself Gaussian and coincides with the

PR-RDPF achieving distribution assuming a Gaussian source $X^* \sim N(0, \sigma^2)$. This allows the characterization of the following equality

$$\max_{f_Z \in \mathcal{S}} \max_{f_{Y|Z} \in \mathcal{H}(f_Z, D)} h(Y|Z) = \max_{f_{Y|X^*} \in \mathcal{H}(f_{X^*}, D)} h(Y|Z) \tag{26}$$

$$= R_{PR}^{G}(D) - h(X^*) \tag{27}$$

which, together with (24) yields (23).

For the general vector case, we consider that for every source $Z \sim f_Z \in \mathcal{S}$, we have marginals $\{Z_i\}_{i=1,\ldots,N}$ with variance $\sigma_{Z_i}^2 = \lambda_i$, where $\{\lambda_i\}_{i=1,\ldots,N}$ is the set of eigenvalues of Σ. Then, from (24) we obtain

$$\begin{aligned}
\max_{f_Z \in \mathcal{S}} \max_{f_{Y|Z} \in \mathcal{H}(f_Z, D)} h(Y|Z) &\overset{(a)}{\leq} \max_{f_Z \in \mathcal{S}} \max_{f_{Y|Z} \in \mathcal{H}(f_Z, D)} \sum_{i=1}^{N} h(Y_i|Z_i) \\
&\overset{(b)}{\leq} \max_{D_i : \sum_i^N D_i = D} \max_{\substack{f_Z \in \mathcal{S}, f_{Y|Z} \\ \mathbb{E}\left[||Z_i - Y_i||^2\right] \leq D_i \ \forall i = 1,\ldots,N \\ Z \sim Y}} \sum_{i=1}^{N} h(Y_i|Z_i) \\
&\overset{(c)}{\leq} \max_{D_i : \sum_i^N D_i = D} \sum_{i=1}^{N} \max_{\substack{f_{Z_i} \in \mathcal{S}, f_{Y_i|Z_i} \\ \mathbb{E}\left[||Z_i - Y_i||^2\right] \leq D_i \ \forall i = 1,\ldots,N \\ Z_i \sim Y_i}} h(Y_i|Z_i) \\
&\overset{(d)}{\leq} \max_{D_i : \sum_i^N D_i = D} \sum_{i=1}^{N} R_{PR}^{G,i}(D_i) - h(X_i^*) \\
&= -h(X^*) + \max_{D_i : \sum_i^N D_i = D} \sum_{i=1}^{N} R_{PR}^{G,i}(D_i)
\end{aligned} \tag{28}$$

where (a) follows from the property of the differential entropy $h(Y_i|Z_k, Z_j) \leq h(Y_i, Z_k)$; (b) follows from the tensorization properties of the MSE distortion; (c) follows from breaking the maximization of the sum of functions into the sum of the maximum of each function; (d) follows from (27), by considering $X^* \sim N(0, \Sigma)$ with marginals $X_i^* \sim N(0, \lambda_i)$ and $R_{PR}^{G,i}(D_i)$ being the PR-RDPF for source X_i and distortion level D_i. Using [22, Corollary 3], we see that the second term of (28) can be shown to be equivalent to the PR-RDPF for the source X^*, therefore showing that (24), and consequently (23), also hold in the general vector case. This completes the proof. □

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,* (23) *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 [22, Corollary 3].*