Computational Methods for a Class of Constrained Rate-Distortion Functions

Giuseppe Serra
Joint work with: P. A. Stavrou and M. Kountouris

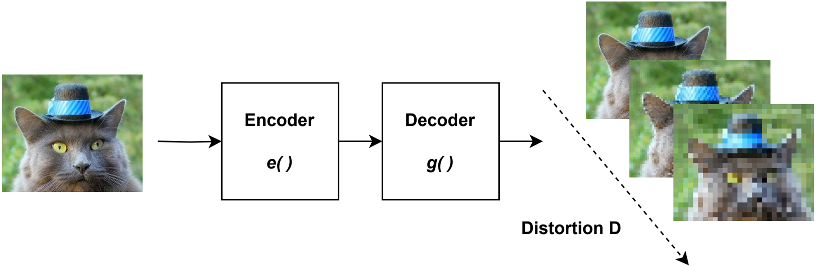
EURECOM

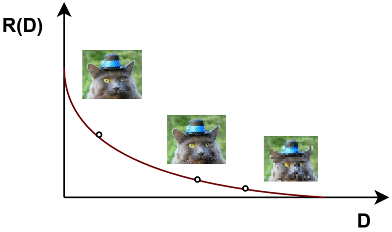
Communication Systems Department

June 7, 2024

Outline

The Rate-Distortion Problem





Rate-Distortion Function

$$R(D) = \min_{P_{X\,|\,\hat{X}}} I(X, \hat{X})$$

$$\text{s.t.} \quad \mathbb{E}\left[d(x, \hat{x})\right] \leq D$$

Human Perception, Semantic and Statistical Divergences

Objective*:*
"Show the picture of a ***cat*** *with a* ***hat****."*

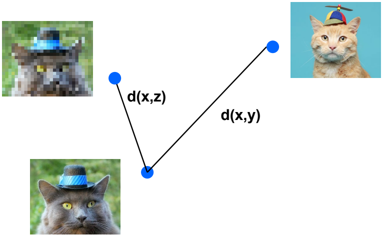
- **High fidelity to the original sample.**
- ***Hardly perceivable semantic characteristics.***

- **Low fidelity to the original sample.**
- ***Maintains recognizable semantic characteristics.***

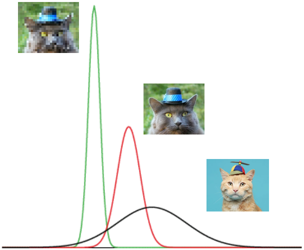
In various domains, perceptual quality has been associated with the deviation of the distribution of output signals from the distribution of the information source[1].

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

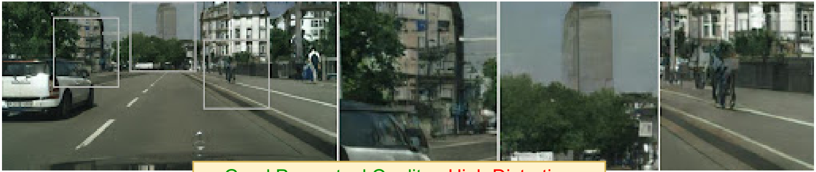
Distortion : How different are they?



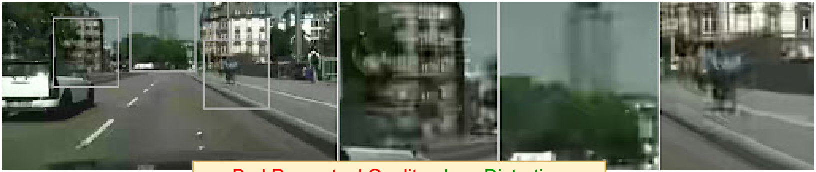
Perception : How different are the distribution they belong to?



Semantic Quality [2]



Good Perceptual Quality - High Distortion

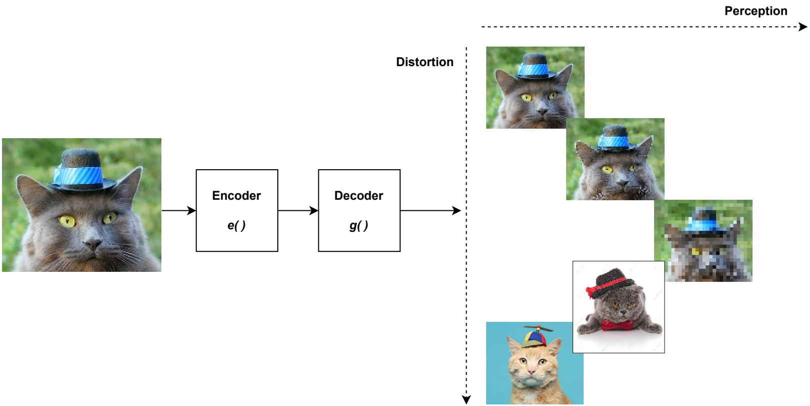


Bad Perceptual Quality - Low Distortion

Good Perceptual Quality ≠ Low Distortion

[2] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, L. Van Gool, Generative adversarial networks for extreme learned image compression, IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 221-231

Rate-Distortion-Perception Problem



Perceptual quality indicator adds to the classical rate distortion problem some human-centric aspects (e.g., human perception of some visualization)

Problem Formulation

- Given an alphabet set $\mathcal{X}$, we consider an I.I.D. source sequence of random variables $X_t : t = 1 \ldots, \infty$ with distribution p_X.

$$\text{(Encoder)} \quad f^E : \mathcal{X}^n \to \mathcal{W} \quad \text{(Decoder)} \quad g^D : \mathcal{W} \to \widehat{\mathcal{X}}^n \tag{1}$$

where $\mathcal{W} = \{1, 2, \ldots, M\}$, $M \in \mathbb{N}_+$.

- For a single-letter distortion metric $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathcal{R}_0^+$ and a divergence measure $D : \mathcal{P} \times \mathcal{Q} \to \mathcal{R}_0^+$, we define the distortion fidelity criteria Δ and perception fidelity criteria Φ as:

$$\Delta \triangleq \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i)\right] \qquad \Phi \triangleq \frac{1}{n}\sum_{i=1}^{n} D(p_{X_i} || q_{\hat{X}_i}).$$

Definition - Operational Rates

Given a distortion level $D > 0$ and a perception constraint $P > 0$, a rate R is said to be (D, P)-achievable if for an arbitrary $\epsilon > 0$, there exists, for large enough n, a lossy source code (n, M, Δ, Φ) with $M \leq 2^{n(R+\epsilon)}$ such that $\Delta \leq D + \epsilon$ and $\Phi \leq P + \epsilon$. Then, we define

$$R_{op}(D, P) \equiv \inf\{R : (R, D, P) \text{ is achievable}\}. \tag{2}$$

Defintion - Rate-Distortion-Perception Function

For a given source distribution p_X, a single-letter distortion measure $d(\cdot,\cdot)$ and a divergence measure $D(\cdot||\cdot)$, the RDPF is characterized as follows:

$$\begin{aligned} R(D,P) = &\min_{P_{\hat{X}|X}} I(X,\hat{X}) \\ &\text{s.t.} \quad \mathbb{E}\left[d(x,\hat{x})\right] \leq D \qquad D(p_X||q_{\hat{X}}) \leq P \end{aligned} \tag{3}$$

where $D \in [D_{\min}, D_{\max}] \subset (0,\infty)$, $P \in [P_{\min}, P_{\max}] \subset (0,\infty)$

When (3) is achievable?

- Theis and Wagner[2] showed that **variable-rate codes**, per-letter average distortion $\mathbf{E}\left[d(\mathbf{x}_i, \hat{\mathbf{x}}_i)\right] \leq D,\ t = 1,\ldots,n$, and **common randomness** at the encoder/decoder achieve (3) for general sources and perception constraints, with $D > 0$, $P \geq 0$
- Chen et al.[3] showed that for $|\mathcal{X}| < \infty$, $D > 0$, $P > 0$, we have that $R_{nr}(D,P) = (3)$

[2] L. Theis and A. B. Wagner, *A coding theorem for the rate-distortion-perception function*, Neural Compression Workshop at ICLR, 2021

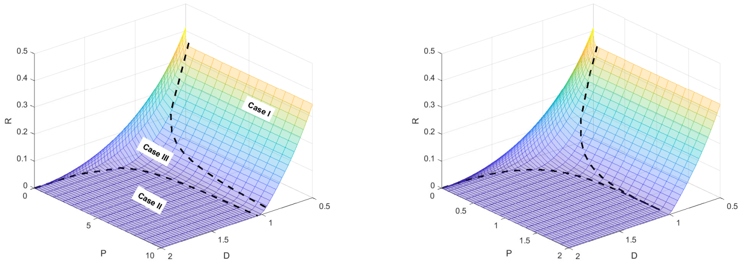
[3] J. Chen, L. Yu, J. Wang, W. Shi, Y. Ge, and W. Tong, *On the Rate-Distortion-Perception function*, IEEE Journal on Selected Areas in Information Theory, 2022.

RDPF for Gaussian Sources[3]

[3] G. Serra, P. A. Stavrou and M. Kountouris, *"On the Computation of the Gaussian Rate–Distortion–Perception Function"*, in IEEE Journal on Selected Areas in Information Theory, vol. 5, pp. 314-330, 2024

Scalar Gaussian RDPF

- We derive analytical bounds of the RDPF under the MSE distortion when the perception constraint belongs to certain well-known and widely-used divergences, e.g. the Kullback–Leibler divergence $D_{KL}(p_X, p_{\hat{X}})$ and the squared Hellinger distance $H^2(p_X, p_{\hat{X}})$.

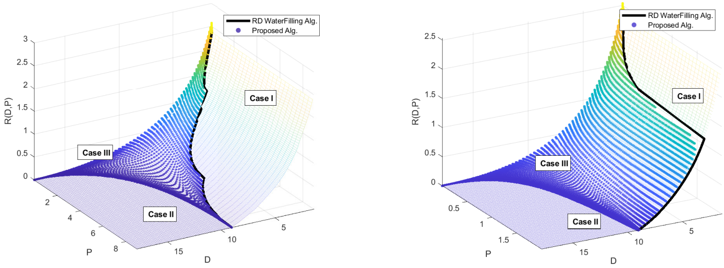


(a) (b)

$R(D, P)$ for a Gaussian source $X \sim \mathcal{N}(0, 1)$ source under (a) $D_{KL}(p_X, p_{\hat{X}})$ divergence and (b) $H^2(p_X, p_{\hat{X}})$ distance.

Vector Gaussian RDPF

- Under tensorizable distortion and perceptual metrics and if the RDPF for the scalar Gaussian source is available, we propose an algorithm for the computation of vector RDPF counterpart.



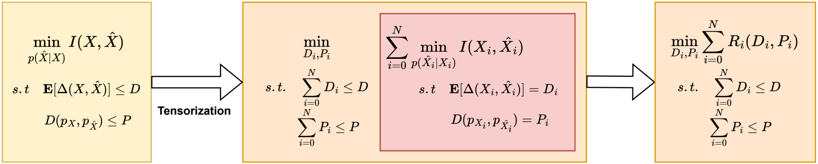
(a) (b)

$R(D, P)$ for a Gaussian source $X \sim \mathcal{N}(0, \Sigma_X)$ source with $\Sigma_X = \text{diag}(1, 3, 5)$ under (a) $W^2(p_X, p_{\hat{X}})$ distance [4] and (b) $H^2(p_X, p_{\hat{X}})$ distance.

[4] The corresponding scalar case is introduced in [*Zhang et. al. - "Universal rate-distortion-perception representations for lossy compression"*]

Algorithm Overview

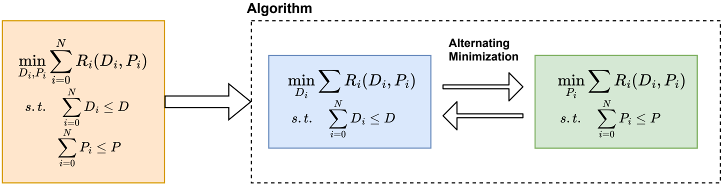
- Leveraging the tensorization properties of the involved metrics, the vector function can be decomposed in a sum scalar functions.



- Introducing the variables $\{D_i, P_i\}_{i=[1:N]}$, the definition of the scalar RDPF function for the i^{th} dimension $R_i(D_i, P_i)$ appears in the optimization problem.

Algorithm Overview

- The resulting optimization problem, despite being simpler than the original problem, is not easily solvable when both constraint are active.



- However, since the two constraints operate on distinct optimization variables, an alternating minimization routine can be implemented to solve optimally the problem.

Perfect Realism RDPF (PR-RDPF)

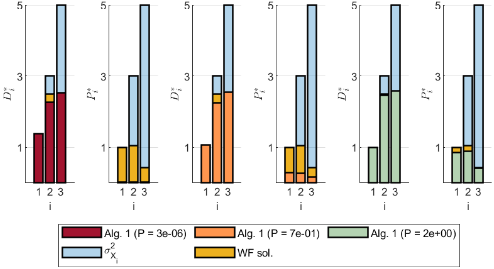
The proposed AM approach allows to derive closed form expression for fixed perception levels P_i. Of particular interest results the case where $P_i = 0$, referred to as *"perfect realism"*, where $\hat{X}$ will have the same distribution as X.

- PR-RDP Waterfilling solution:

$$D^*_{i,RDP} = 2\lambda_i[\Sigma_X] + \frac{1}{2s_D} - \sqrt{4\lambda_i^{\frac{1}{2}}[\Sigma_X] + \frac{1}{4s_D^2}}$$

- RD Waterfilling solution:

$$D^*_{i,RD} = \min\left\{\frac{1}{2s_D}, \lambda_i[\Sigma_X]\right\}$$



Comparison of the per-dimension distortion D_i^* and perception P_i^* levels between the waterfilling solution for D = 6.

Computation of the Output-Constrained RDFs[5]

[5] Serra, G., Stavrou, P. A., Kountouris, M. (2024). *Copula-based Estimation of Continuous Sources for a Class of Constrained Rate-Distortion-Functions*. arXiv preprint arXiv:2401.17089.

Notation

Given a scalar random variable (RV) X on $\mathcal{X} \subseteq \mathbb{R}$, we denote

- the distribution function (d.f.) of X, i.e., its law, F_X;
- the quantile function (q.f.) of X, i.e., $Q_X(u) \triangleq \sup\{x \in \mathcal{X} : F(x) \leq u\}$;
- the probability density function (p.d.f.) of X, i.e., its density, f_X;

Furthermore, given a multivariate random variable $X = (X_1, \ldots, X_d)$ with $X_i \sim F_{X_i}$ for $i = 1, \ldots, d$, we define the following vector functions:

- **Uniform Transformation**:
  $\Phi_X : \mathcal{X} \to [0,1]^d$ defined as $\Phi_X(X) \triangleq (F_{X_1}(X_1), \ldots, F_{X_d}(X_d))$
- **Inverse Uniform Transformation**:
  $\Psi_X : [0,1]^d \to \mathcal{X}$ defined as $\Psi_X(U) \triangleq (Q_{X_1}(U_1), \ldots, Q_{X_d}(U_d))$.

Output-Constrained Rate-Distortion Function (OC-RDF)

OC-RDF [6].

Let $\mathcal{X} \subset \mathbb{R}^d$ and 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

$$R_{OC}(D) = \min_{\substack{f_{Y|X} \in \hat{\Pi}(f_X, f_Y) \\ \mathbb{E}[\Delta(X,Y)] \leq D}} I(X,Y) \tag{4}$$

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

We denote the following differences:

- in the PR-RDPF case, we specifically constrain the reconstruction distribution and source distribution to be identical;
- in the OC-RDF case, we have an additional degree of freedom allowing for the distribution of the reconstruction to be chosen freely.

[6] 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.

Entropic Optimal Transport (EOT)

EOT [7]

Let $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Y} \subset \mathbb{R}^m$ and 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

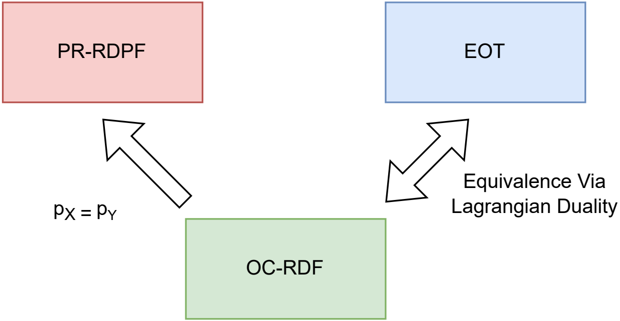
$$D_{EOT}(\epsilon) = \min_{f_{X,Y} \in \bar{\Pi}(f_X, f_Y)} \mathbb{E}\left[\Delta(X, Y)\right] + \epsilon I(X, Y) \tag{5}$$

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

[7] 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.

Links between PR-RDPF, OC-RDF, and EOT

We show that the RDPF, OC-RDF and EOT share a common underlining structure, which allows to map one problem instance to the other.



Therefore, although our analysis will focus on the OC-RDF problem, all the results can be easily adapted to the remaining problems.

What is a Copula?

Copula Distribution

A d-dimensional Copula distribution is a distribution on the volume $[0,1]^d$ with uniform marginals.

Why is it so interesting then? The answer is **Sklar's Theorem**

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 $(x_1, x_2, \ldots, x_d) \in \bar{\mathbb{R}}^d$,

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right). \tag{6}$$

Such a C is uniquely determined on $[0,1]^d$ and, hence, it is unique when $F_1, F_2, \ldots, F_d$ are continuous.

Sklar's Theorem (Corollary)

Let $\mathcal{X} \subset \mathbb{R}^d$ and let $f \in \mathcal{P}(\mathcal{X})$. Then, f can be uniquely factorized as

$$f(x_1, \ldots, x_d) = c\left(F_1(x_1), \ldots, F_d(x_d)\right) \prod_{j=1}^{d} f_j(x_j) \tag{7}$$

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

In other words, any multivariate distribution can be decomposed into:

Multivariate density = Copula density + Marginal densities

- a copula density, expressing the relations between the marginal RVs.
- a product of marginal density, representing the information of each single marginal RV.

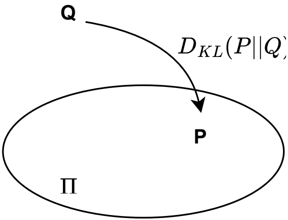
Information Geometry

Information Geometry (IG) is an interdisciplinary field that applies the techniques of differential geometry to the study of statistical manifold.

For the purpose of this work, we will use the tools of IG to characterize the solution of projection problems of the type:

$$\min_{P \in \Pi} D_{KL}(P||Q)$$

where Π is a convex set of probability measures.



Projection Theorems

For specific typologies of convex set Π, the functional characterization of the optimal projection has been characterized in [Csizar,1975]. The characterization is expressed in terms of density, i.e. Radon–Nikodym derivative, of the projection.

The following cases are of interest for our problems:

- **(Case A)** - Π is defined as a linear set of equality constraints, i.e. $\Pi = \{P : E_P[h_i] = \alpha_i, i = 1, \ldots, n\}$, then the projection P is unique, with density:

$$\frac{dP^*}{dQ}(\mathbf{u}) = e^{\mu + \sum_{i=1}^{n} \theta_i h_i(\mathbf{u})}$$

  for some values of μ and θ_i, $i = 1, \ldots, n$.

- **(Case B)** - Π is defined as a set of constraint on the marginals of P, i.e. $\Pi = \{P : P_i \sim F_i, i = 1, \ldots, d\}$

$$\frac{dP^*}{dQ}(\mathbf{u}) = \prod_{i=1}^{d} g_i(u_i).$$

  for some scalar functions g_i, $i = 1, \ldots, d$, such that $\log(g_i) \in l_1$

Successinve Projections Theorem

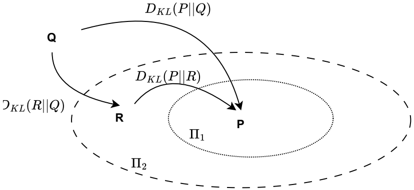
The last result in IG we need characterizes the solution for a sequence of projections on a given sequence of subsets.

Let the sets Π_1 and Π_2 be convex sets such that $\Pi_1 \subseteq \Pi_2$, and let P and R be the respective projections of Q thereon. Then,

$$D_{KL}(P||Q) = D_{KL}(R||Q) + D_{KL}(P||R)$$

and,

$$\frac{dP}{dQ}(\mathbf{u}) = \frac{dR}{dQ}(\mathbf{u}) \cdot \frac{dP}{dR}(\mathbf{u}).$$



OC-RDF is a IG Projection Problem

Lemma - Copula Reparametrization

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

$$I(X, Y) = \mathrm{D_{KL}}(C_{X,Y} || C_X \otimes C_Y) \tag{8}$$

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

$$\mathbb{E}_{F_{X,Y}} \left[\Delta(X, Y)\right] = \mathbb{E}_{C_{X,Y}} \left[\Delta\left(\Psi_X(U_X), \Psi_Y(U_Y)\right)\right] \tag{9}$$

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

Proof: Direct result from Sklar's Theorem

Applying the result of the previous Lemma to the OC-RDF problem:

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

we derive the following copula optimization problem:

Copula-based OC-RDF

The mathematical expression of the OC-RDF Problem in (4) can be reformulated as

$$R_{OC}(D) = \min_{C \in \mathcal{C}_{2d}} \mathrm{D_{KL}}(C || C_X \otimes C_Y) \tag{10}$$
$$\text{s.t. } \mathbb{E}_C\left[\Delta(\Psi_X(U_X), \Psi_Y(U_Y))\right] = D$$

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

Analyzing the structure of the optimization problem, we can understand that the **Copula-based OC-RDF problem is an IG projection problem on a convex domain**.

The set of constraints of the OC-RDF can be decoupled as two sets Π_1 and Π_2 such that:

- Π_2 is the set of all distributions on $[0,1]^{2d}$ respecting the distortion constraint, i.e. Π_2 **is a linear set of constraint**.
- $\Pi_1 \subset \Pi_2$ is the set of all copula distributions respecting the distortion constraint, i.e. Π_1 **imposes conditions on the projection marginals**.

Under this description, using the sequential projection results we can derive:

Solution of the Copula OC-RDF solution

Let $R = C_X \otimes C_Y$. Then, the Copula OC-RDF problem admits a minimizing copula Q with Radon–Nikodym derivative with respect to the measure R of the form

$$\frac{dC}{dR}(\mathbf{u}) = e^{\mu+\theta[\Delta(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))]} \prod_{i=1}^{2d} g_i(u_i) \tag{11}$$

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

Proof: Application of sequential projection using the solutions of Case A (linear constraint) and Case B (marginal constraint).

So... job finished, everybody home?
Sadly no.

Do you think we can numerically compute the previous projection?

Approximate IG Projection

The problem with the previous characterization is that we **don't know the functions** g_i, introduced by the uniform marginal constraint.
$\Rightarrow$ *What if we can enforce the uniform marginal constraints in a different way?*

Hausdorff moments problem: Any RV on $[0,1]$ that respects $\mathbb{E}\left[u^n\right] = \frac{1}{n+1}$ for all $n = 1, \ldots,$ is necessarily uniformly distributed.

Lower bound to Copula OC-RDF

For any integer N, the Copula OC-RDF can be lower bounded as follows

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

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 $[0,1]$.

Solution of the Lower bound ot the Copula OC-RDF solution

Let $R = C_X \otimes C_Y$ and assume there exists a d.f. P on $[0,1]^{2d}$ such that $\mathrm{D_{KL}}(P||R) < \infty$. Then, $R_{OC}^{(N)}(D)$ admits minimizing copula Q with Radon–Nikodym derivative with respect to the measure R of the form

$$\frac{dQ}{dR}(\mathbf{u}) = e^{\mu + \theta\Delta(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y))} \prod_{i=1}^{2d} e^{\sum_{n=0}^{N} \nu_{i,n} u_i^n} \tag{12}$$

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

$$\min_{(\mu,\theta,\{\nu_{i,n}\}_{(i,n)\in I})} -\mu - \theta D - \sum_{(i,n)\in I} \nu_{i,n}\alpha_n + \left(\int_{[0,1]^{2d}} \frac{dQ}{dR}(\mathbf{u})dR(\mathbf{u}) - 1\right). \tag{13}$$

Proof: Existence of the solution derives from the results on IG Case A. The optimization problem is derived imposing stationarity of the first variation.

Algorithm Design

The optimization problem defined in (13) is stictly convex, hence we can rely on gradient methods for its optimal solution.

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

Require: marginal distributions $\{F_{X_i}, F_{Y_i}\}_{i=1,\ldots,d}$; 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, ..., T

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

  $f(\mathbf{l}) \approx (12) + \left(\frac{1}{M}\sum_{i=1}^{M} \frac{dQ}{dR}(\mathbf{l}, \mathbf{u}_i) dR(\mathbf{u}_i)\right)$

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

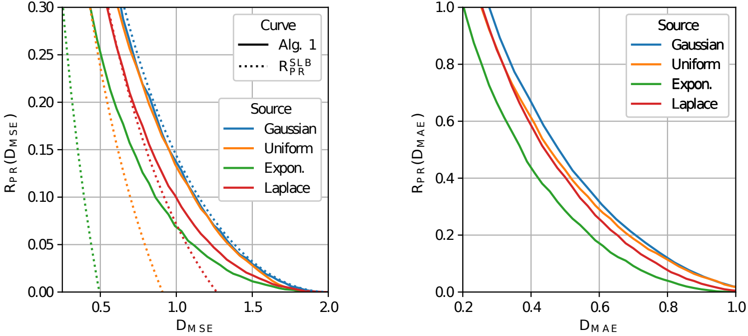
end for

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

MonteCarlo Estimation of

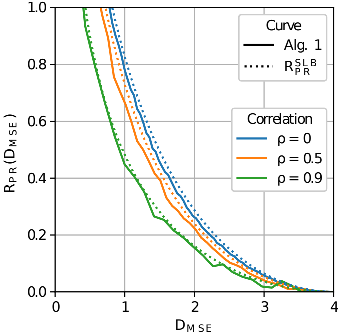
$$\int_{[0,1]^{2d}} \frac{dQ}{dR}(u) dR(u) - 1)$$

Numerical Examples - PR-RDPF



PR-RDPF for various source distributions with $\sigma_X = 1$ under (a) MSE distortion metric and (b) MAE distortion metric.

Numerical Examples - PR-RDPF



PR-RDPF under MSE distortion metric for a (a) Gaussian, and (b) exponential bivariate source with varying degree of correlation coefficient ρ.

Thanks for the Attention!