DOCTORAL THESIS
SORBONNE UNIVERSITY

Scientific Field:

Information and Communication Sciences and Technologies

Doctoral School:

Computer Science, Telecommunications, and Electronics of Paris

Presented by

Giuseppe Serra

To obtain the degree of

Doctor of Philosophy

Thesis Title:

Goal-oriented Semantic Communication: A Rate Distortion Perception Approach

Defended on May 4th, 2026, before the jury composed of

Reviewers:

Prof. Jun Chen, McMaster University, Canada

Prof. Michèle Wigger, Telecom Paris, France

Examiners:

Prof. Elia Petros, EURECOM, France *(Jury President)*

Prof. Albert Guillén i Fàbregas, University of Cambridge, United Kingdom

Prof. Charalambos D. Charalambous, University of Cyprus, Cyprus

Thesis Director:

Prof. Marios Kountouris, EURECOM, France

Thesis Co-Supervisor:

Prof. Photios A. Stavrou, EURECOM, France

THÈSE DE DOCTORAT
SORBONNE UNIVERSITÉ

Spécialité :

Sciences et Technologies de l'Information et de la Communication

École doctorale :

Informatique, Télécommunications et Électronique de Paris

Présentée par

Giuseppe Serra

Pour obtenir le grade de

Doctorat

Sujet de la thèse :

Communication Sémantique Orientée sur les Objectifs: Une Approche par Débit-Distorsion-Perception

Soutenue le 4 Mai 2026 devant le jury composé de

Rapporteurs :

Prof. Jun Chen, McMaster University, Canada

Prof. Michèle Wigger, Telecom Paris, France

Examinateurs :

Prof. Elia Petros, EURECOM, France *(Président du Jury)*

Prof. Albert Guillén i Fàbregas, Université de Cambridge, Royaume-Uni

Prof. Charalambos Charalambous, Université de Chypre, Chypre

Directeur de Thèse:

Prof. Marios Kountouris, EURECOM, France

Co-Encadrant de Thèse:

Prof. Photios A. Stavrou, EURECOM, France

To my family,
harbor in the restless tide,
home after long waves.

"Ma che colpa abbiamo, io e voi, se le parole, per sé, sono vuote? Vuote, caro mio. E voi le riempite del senso vostro, nel dirmele; e io nell'accoglierle, inevitabilmente, le riempio del senso mio. Abbiamo creduto d'intenderci, non ci siamo intesi affatto."

— *Luigi Pirandello,* Uno, Nessuno e Centomila

Abstract

In his seminal communication model, C. Shannon deliberately excluded semantic aspects and message effectiveness from the scope of the communication problem. While this abstraction has been fundamental for decades of progress in information science, the growing need for goal-oriented data usage, including learning, inference, and decision-making, has revived interest in communication frameworks that account for semantics and utility. In this thesis, we discuss a general theoretical framework that extends rate-distortion theory to incorporate perceptual and semantic metrics. By imposing divergence-based constraints, we characterize how information can be effectively transmitted and reconstructed to align with specific goals, such as preserving human-perceived quality or semantic meaning. We analyze this problem for both discrete and continuous information sources, deriving structural properties that characterize the optimal tradeoff between perceptual quality and traditional fidelity metrics. We also propose novel algorithmic solutions and robust coding strategies that provide guarantees on both fidelity and perceptual relevance, even under model uncertainty. The theoretical and algorithmic contributions of this work support the development of scalable, resource-efficient communication systems, enabling more effective integration of compression, perception, and semantics in intelligent networked systems.

Résumé

Dans son modèle de communication fondateur, C. Shannon a délibérément exclu les aspects sémantiques et l'efficacité des messages du cadre du problème de communication. Bien que cette abstraction ait été fondamentale pour des décennies de progrès en sciences de l'information, le besoin croissant d'une utilisation finalisée des données — incluant l'apprentissage, l'inférence et la prise de décision — a ravivé l'intérêt pour des cadres de communication prenant en compte la sémantique et l'utilité. Dans cette thèse, nous présentons un cadre théorique général qui étend la théorie du débit-distorsion afin d'y intégrer des métriques perceptuelles et sémantiques. En imposant des contraintes fondées sur des divergences, nous caractérisons comment l'information peut être efficacement transmise et reconstruite de manière à répondre à des objectifs spécifiques, tels que la préservation de la qualité perçue par l'humain ou du sens sémantique. Nous analysons ce problème pour des sources d'information discrètes et continues, en dérivant des propriétés structurelles qui caractérisent le compromis optimal entre qualité perceptuelle et fidélité traditionnelle. Nous proposons également de nouvelles solutions algorithmiques et des stratégies de codage robustes qui offrent des garanties à la fois sur la fidélité et la pertinence perceptuelle, même en présence d'incertitude sur le modèle de la source. Les contributions théoriques et algorithmiques de ce travail soutiennent le développement de systèmes de communication évolutifs et économes en ressources, permettant une intégration plus efficace de la compression, de la perception et de la sémantique dans des systèmes intelligents interconnectés.

Acknowledgements

Now that this journey is about to end, I look back and realize that this milestone belongs as much to me as to all the people that supported me throughout these years.

First and foremost, I would like to express my deepest gratitude to Fotios Stavrou and Marios Kountouris for their guidance and support throughout this journey. Their mentorship was invaluable to help me grow both as a researcher and as a human being. I hope that the time spent mentoring me was as rewarding for them as it was for me. I wish also to express my sincere appreciation to the European Research Council (ERC) for funding my research and making this work possible.

Equally important have been the interactions with all my colleagues and friends at EURECOM, whose drive and curiosity have inspired me in my research. In particular, I would like to thank Nour, Reza, and Zixuan. Research without the laughs in our office would not have been the same.

To my mother Antonella, my father Giulio, my sister Angelica, and my future nephew Leonardo goes my utmost heartfelt love and gratitude. I am able to go forward only thanks to your love and support. I need to also thank my lifelong friends Antonella, Mario, Dario, Sebastiano, Matteo, Mattia, Alessandro, and Sandro, that, no matter the geographic distance, were always present for all my milestones. Last but not least, I need to thank Andrea, whose presence and support has been fundamental for this success.

There are many other people that would merit a mention, but space would not allow it. After all, this is a thesis about compression. In any case, thank you for all.

Giuseppe Serra
May 4^{th}, 2025

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past two decades, internet traffic has undergone a significant transformation, driven by content creation and delivery. The widespread adoption of smartphones, high-resolution cameras, streaming services, and social media platforms has driven a massive surge in the transmission and storage of human-centric data—primarily images, video, and audio. Recent reports indicate that multimedia content constitutes the vast majority of global internet traffic, with video alone accounting for over 65% of all data [2].

A key enabling factor for the efficient handling of such data volumes is source coding, i.e., the process of compressing data from an information source into a compact representation for transmission or storage. In its lossless form, source coding ensures exact reconstruction of the original data, albeit with stringent requirements in terms of the amount of information to encode. However, for many practical sources such as images, video, and audio, perfect reconstruction is largely unnecessary. In these cases, lossy source coding provides a far more efficient alternative by allowing controlled degradation of the reconstructed signal in exchange for reduced bitrates. Lossy compression techniques are deeply embedded in modern digital infrastructure. For example, streaming and conferencing platforms like YouTube, Netflix, and Zoom rely on aggressive video compression algorithms (e.g., H.264, AV1) to maintain smooth real-time playback under bandwidth constraints. Social media platforms also apply lossy compression to uploaded media to reduce storage demands and accelerate content delivery. These systems must carefully balance compression rates with perceptual quality, aiming to minimize artifacts and preserve subjective fidelity. Classical rate-distortion (RD) theory, introduced by Shannon [3], has traditionally served as a theoretical framework for navigating this tradeoff, quantifying the minimum rate needed to keep the expected distortion between the original and reconstructed signal below a given threshold.

Goal-Oriented Compression and Human Perception Rate-distortion theory relies on the assumption that a single distortion metric can adequately capture all relevant aspects of signal fidelity, thereby implicitly inducing decisions about which signal features matter most for reconstruction. Traditionally, metrics like mean squared error (MSE) or

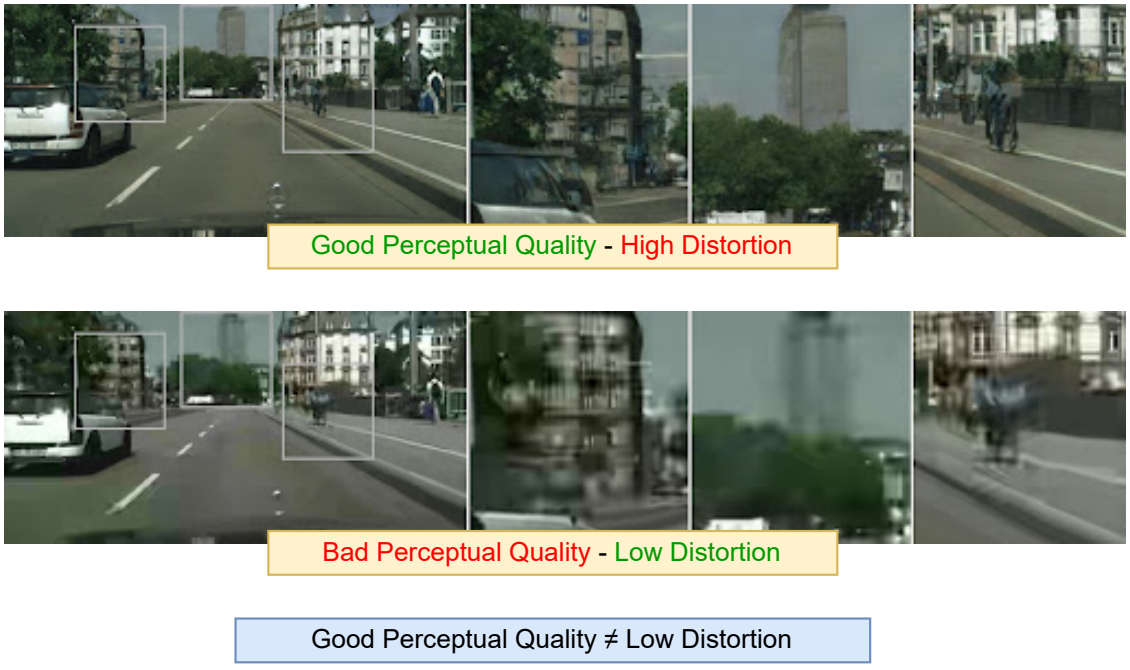


Figure 1.1: Comparison between video frames compressed under MSE distortion with (top) and without (bottom) perceptual regularization [1].

peak-signal-to-noise ratio (PSNR) were favored due to their mathematical interpretability and ease of implementation. However, such metrics often fail to reflect the actual goals of the sender and receiver, especially in modern applications. The recent field of goal-oriented compression seeks to address this by designing distortion metrics that are tailored to specific tasks, while attempting to incorporate varying degrees of mutual intent and contextual relevance for both sender and receiver [4–6].

Nevertheless, when the goal is to produce perceptually pleasing outputs, distortion metrics alone may still be fundamentally inadequate. A growing body of empirical evidence [1, 7, 8] shows that commonly used measures like MSE or PSNR correlate poorly with human perception, especially in heavily constrained compression scenarios. In image and video compression, for instance, minimizing MSE often results in blurry or unnatural reconstructions, while visually appealing outputs can score worse under traditional metrics. This gap has widened with the rise of deep generative models, which are capable of producing realistic samples that diverge substantially from the source in terms of classical distortion, as shown in Figure 1.1.

Rate-Distortion-Perception Tradeoff The rate-distortion-perception (RDP) tradeoff studies the problem of lossy compression under perceptual quality constraints, thus generalizing the classical RD analysis. It was concurrently proposed by Blau and Michaeli in [9] and Matsumoto in [10, 11], motivated by the growing need for a theoretical framework that can capture observations raised by a wide body of research spanning machine learning and multimedia applications (see [12–15]), which highlights an inherent tension

between perceptual quality and fidelity of the compressed samples, where perceptual quality refers to the property of a sample appearing visually pleasing to a human observer.

The mathematical representation of the RDP tradeoff is embodied by the rate-distortion-perception function (RDPF), which complements the classical rate-distortion function (RDF) with a divergence constraint between the source and reconstruction distributions. The additional constraint acts as a proxy for human perception, measuring the deviation from the real source distribution, also referred to as the "natural scene statistic", following principles similar to those used in a class of no-reference image quality metrics [16, 17]. It is worth noting, however, that the selection of a specific divergence metrics may be application-dependent and is still an active area of research. An alternative interpretation of the divergence constraint is as a semantic quality metric, i.e., a quantification of the importance of the reconstructed source from the observer's perspective [18]. For example, in [19], a comparison between the segmentation capabilities of models trained on traditionally compressed samples and compressed samples with enhanced perceptual quality through Generative Adversarial Network (GAN)-based restoration shows a remarkable improvement in segmentation performance, especially for smaller scene objects.

1.1 Open Questions

Due to the novelty of the RDP framework, many of its theoretical properties and optimization aspects remain open problems. For example:

1. As with the classical RDF, are there general algorithms for estimating the RDPF for both discrete and continuous sources?

2. What are the structural properties of the distributions that achieve the RDPF? Are there conceptual links between the RDPF and other distributionally constrained problems, such as optimal transport?

3. How does uncertainty on the statistics of the source distribution affect the RDPF? Can robust coding schemes be developed to address general lossy compression tasks under such uncertainty?

This thesis addresses these open problems, offering both algorithmic and theoretical contributions. The main results are outlined in the following section.

1.2 Contributions and Thesis Outline

This thesis is structured around three main areas of contribution. The chapters are organized as follows, alongside the publications whose results they report:

RDPF for Discrete Sources

- In Chapter 2, we study the computation of the RDPF for discrete memoryless sources under single-letter average distortion and f-divergence perception constraints. We develop an Optimal Alternating Minimization (OAM) scheme with convergence guarantees, though it cannot directly implement a generalized Blahut-Arimoto algorithm due to implicit equations in the iteration structure. To address this limitation, we propose two alternative approaches: a Newton-based Alternating Minimization (NAM) scheme for smooth perception metrics and a Relaxed Alternating Minimization (RAM) scheme based on a relation of the OAM iterates. We prove that both schemes converge to globally optimal solutions through necessary and sufficient conditions, and establish sufficient conditions for exponential convergence rates.

  ▷ [20] *"Computation of rate-distortion-perception function under f-divergence perception constraints."* 2023 IEEE International Symposium on Information Theory (ISIT). IEEE, 2023.

  ▷ [21] *"Alternating minimization schemes for computing rate-distortion-perception functions with f-divergence perception constraints."* IEEE Transactions on Information Theory, vol. 71, no. 11, pp. 9100-9115, 2025.

RDPF for Continuous Sources

- In Chapter 3, we study the computation of the RDPF for multivariate Gaussian sources under jointly Gaussian reconstruction, MSE distortion, and various perception metrics. We first characterize analytical bounds for the scalar Gaussian RDPF and provide the corresponding RDPF-achieving forward test-channel realizations for these divergence functions. For the multivariate case with tensorizable distortion and perception metrics, we establish that the optimization problem can be expressed in terms of scalar Gaussian RDPFs of source marginals subject to global constraints. We design an alternating minimization scheme that optimally solves the problem while identifying the RDPF-achieving realization, and we provide algorithmic implementation details and a convergence characterization.

  ▷ [22] *"Computation of the Multivariate Gaussian Rate-Distortion-Perception Function."* 2024 IEEE International Symposium on Information Theory (ISIT). IEEE, 2024.

  ▷ [23] *"On the computation of the Gaussian rate–distortion–perception function."* IEEE Journal on Selected Areas in Information Theory, vol. 5, pp. 314-330, 2024

- In Chapter 4, we focus on the RDPF of a source modeled as a Gaussian Process (GP) over a measure space under mean squared error distortion and squared Wasserstein-2 perception metric. First, we show that the optimal reconstruction process is itself

a GP, whose covariance operator shares the same set of eigenvectors as the source's covariance operator. This structural property allows us to reformulate the RDPF problem in terms of the Karhunen–Loève (KL) transform coefficients of the involved GPs, which in turn enables us to derive a tight analytical upper bound on the RDPF for GPs. Finally, for stationary GPs over the interval $[0, T]$ with Lebesgue measure, we derive an upper bound on the rate and distortion for a fixed perceptual level as $T \to \infty$, as a function of the spectral density of the source process.

▷ [24] *"On the Rate-Distortion-Perception Function for Gaussian Processes,"* IEEE International Symposium on Information Theory (ISIT), 2025.

- In Chapter 5, we present a novel method to estimate the RDPF in the perfect realism regime (PR-RDPF) for general multivariate continuous sources. The proposed approach is not only able to solve this specific problem but also two related problems: 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 for the optimal projection. Additionally, 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 for the PR-RDPF against the obtained SLB for various sources.

  ▷ [25] *"Copula-based estimation of continuous sources for a class of constrained rate-distortion functions."* 2024 IEEE International Symposium on Information Theory (ISIT). IEEE, 2024.

Distributionally Robust Source Compression

- In Chapter 6, we investigate the problem of distributionally robust source coding, i.e., source coding under uncertainty about the source distribution. We propose two extensions of the so-called *Strong Functional Representation Lemma* (SFRL), considering the cases where, for a fixed conditional distribution, the marginal inducing the joint coupling belongs to an uncertainty set. Using these extensions, we derive distributionally robust coding schemes for both the one-shot and asymptotic regimes, generalizing previous results in the literature. Focusing on the case where the source distribution belongs to a given KL-sphere, we derive an implicit characterization of the points attaining the robust rate-distortion function (R-RDF), later exploited to implement a novel algorithm for its computation.

  ▷ [26] *"On Distributionally Robust Lossy Source Coding."* 2025 IEEE Information Theory Workshop (ITW). IEEE, 2025.

1.3 Notation

Number Sets Let $\mathbb{N}$ denote the set of natural numbers, and $[a : b] \subset \mathbb{N}$ denote the integer interval including both endpoints. Let $\mathbb{R}$ denote the set of real numbers, with $\bar{\mathbb{R}}$ the extended set $\mathbb{R} \cup \{-\infty, +\infty\}$, and $\mathbb{R}_0^+$ the set of non-negative real numbers. For any set $\mathcal{S} \subseteq \mathbb{R}^n$, we define its complement set by $\mathcal{S}^c \triangleq \{x \in \mathbb{R}^n : x \notin \mathcal{S}\}$.

Real Analysis We denote by C^n the set of n-times differentiable functions on $\mathbb{R}$. Given a function $f \in C^0$, we denote by ∂f its sub-gradient [27, Definition 8.3], while, if $f \in C^2$, we denote by $f''(\cdot)$ the second derivative with respect to its argument. Given a multivariate function $(x, y) \to f(x, y)$ with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, we denote by ∇f the gradient of the function $f(x, y)$ and with $\nabla_x f$ (or $\nabla_y f$) the gradient with respect to the specific variable.

Vectors and Matrices The identity operator is denoted as $\mathrm{id}(\cdot)$, whereas the identity matrix is denoted by I. Given a vector $v \in \mathbb{R}^n$, we denote by $\operatorname{diag} v \in \mathbb{R}^{n \times n}$ the matrix with the elements of v on its diagonal and zeros otherwise. Given a matrix $A \in \mathbb{R}^{n \times n}$, we denote by $s_A = [s_{A,i}]_{i=1:N}$ the vector of its singular values and, if A is diagonalizable, by $\lambda_A = [\lambda_{A,i}]_{i \in 1:n}$ the vector of its eigenvalues. However, in the cases where the notation s_A or λ_A is inadequate, we indicate the eigenvalues and singular values using the operators $\lambda[A]$ and $s[A]$, respectively. We introduce the notations $\lambda^{\downarrow}[\cdot]$ or $s^{\downarrow}[\cdot]$ and $\lambda^{\uparrow}[\cdot]$ or $s^{\uparrow}[\cdot]$, respectively, for the vectors whose coordinates are the eigenvalues or singular values arranged in decreasing or increasing order. We denote positive definiteness (resp. positive semi-definiteness) with the notation $A \succ 0$ (resp. $A \succeq 0$). The trace operator is denoted by $\mathrm{Tr}[\cdot]$, while $\|\cdot\|_1$, $\|\cdot\|_F$, and $\langle \cdot, \cdot \rangle_F$ denote the, respectively, Schatten-1 norm [28, Section 4.2], Frobenius norm, and Frobenius inner product [29, Equation 5.2.7].

Measures and Divergences For a Polish space $\mathcal{X}$, 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 probability measures defined thereon. Given $\nu, \mu \in \mathcal{P}(\mathcal{X})$, we denote by $\nu \ll \mu$ the absolute continuity of ν with respect to μ, meaning that for any set $\mathcal{A} \in \mathbb{B}(\mathcal{X}), \mu(\mathcal{A}) = 0 \implies \nu(\mathcal{A}) = 0$. For $\nu \in \mathcal{P}(\mathcal{X})$ and $\mu \in \mathcal{P}(\mathcal{Y})$, we denote their independent product measure as $\mu \times \nu$. Furthermore, given any joint measure $\mu_{X,Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, we denote by $m_X(\mu_{X,Y})$ and $m_Y(\mu_{X,Y})$ the marginal measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively. We denote by $\mathcal{L}(\mathcal{X} \times \mathcal{Y})$ the set of Markov kernels Q such that $Q \cdot p \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ for all $p \in \mathcal{P}(\mathcal{X})$. Depending on the context, we explicitly indicate the functional dependency between mathematical objects with square brackets, for example, $p[h]$ and $Q[h]$ express the functional dependence of a distribution $p \in \mathcal{P}(\mathcal{X})$ or a transition matrix $Q \in \mathcal{L}(\mathcal{X} \times \mathcal{Y})$ on another distribution $h \in \mathcal{P}(\mathcal{X})$.

Random Variables For a random variable (RV) X on $\mathcal{X} \subseteq \mathbb{R}^n$, we specialize our notation by denoting by $P_X \in \mathcal{P}(\mathcal{X})$ as its distribution function (shortly, d.f.) and with p_X its probability density function (shortly, pdf). For a discrete RV X, p_X will equivalently indicate its probability mass function (shortly, pmf). We denote by $\mathbb{E}[\cdot]$ the expectation operator, and by $\mathbb{E}_\mu[\cdot]$ we specify the probability distribution μ on which the expectation operator is applied. For a RV X, we denote by μ_X and Σ_X its mean and covariance matrix, respectively. Given RVs X and Y, $X \perp Y$ indicates their statistical independence, while we denote by $h(X)$, $h(X|Y)$, and $I(X,Y)$, respectively, the (differential) entropy of X, the conditional (differential) entropy of X given Y, and the mutual information between X and Y. When necessary, the mutual information $I(X,Y)$ for $(X,Y) \sim p_X Q_{Y|X}$ will be alternatively indicated with $I(P_X, Q_{Y|X})$ to highlight the dependency on $(P_X, Q_{Y|X}) \in \mathcal{P}(\mathcal{X}) \times \mathcal{L}(\mathcal{X} \times \mathcal{Y})$.

Others The Lambert W function [30] is denoted by $\mathcal{W}(\cdot)$, whereas its primary and secondary branches are denoted by $\mathcal{W}_0(\cdot)$ and $\mathcal{W}_{-1}(\cdot)$, respectively.

1.4 Preliminaries

1.4.1 The Rate-Distortion-Perception Tradeoff

Since its introduction, the RDP tradeoff has received substantial interest from the information theory community, deriving operational characterizations in a variety of operational scenarios. Focusing on the case where infinite common randomness is available both at the encoder and the decoder, Theis and Wagner [31] provide variable-length codes for both the one-shot and asymptotic regimes exploiting the properties of the strong functional representation lemma [32]. In the context of the output-constrained RDF, but also valid for the "perfect realism" RDPF case, Saldi *et al.* [33] provide coding theorems for the case where only finite common randomness between encoder and decoder is available. In [34], Chen *et al.* derives coding theorems for the asymptotic case, focusing on the differences between three operational cases defined by the availability of randomness between encoder and decoder, i.e. infinite common randomness, only private randomness, and the deterministic case.

Similar to its classical counterpart, the RDPF does not enjoy any general analytical solution. However, for specific source distributions and particular choices of the distortion and perception measures, there exist closed-form expressions, such as for binary sources subject to Hamming distortion and total variation (TV) distance [9] and Gaussian sources under mean squared-error distortion and various perception measures [23, 35, 36]. The available closed forms, while providing theoretical insights on the RDP tradeoff, have limited applicability when considering arbitrary distortion and perception metrics. Therefore, numerical solutions for the estimation of the RDPF have been explored to great extent. Data-driven solutions have been proposed, usually employing a

generative adversarial scheme minimizing a linear combination of distortion and perception metrics, see e.g., [9, 35, 37, 38]. Despite the advantage of directly optimizing the image/video codec using only source samples, these methodologies still require particular effort as they are generally highly computational and data-intensive and may suffer from a lack of generalization capabilities. Algorithmic solutions for the RDPF estimation are also present. In the case of discrete alphabets, Chen *et al.* in [39] cast the RDPF problem as an entropic-regularized Wasserstein barycenter problem and propose a solving method based on the Sinkhorn algorithm applicable to arbitrary distortion measures and perception measure being either a Wasserstein-type distance, the Kullback–Leibler divergence, or the TV distance.

The Rate-Distortion-Perception Function We provide in the following definition of the information-theoretic characterization of the RDPF for general alphabets [9].

Definition 1.4.1. *(RDPF) Let a source X be a random variable defined in an alphabet $\mathcal{X}$ and distributed according to $P_X \in \mathcal{P}(\mathcal{X})$. Then, the RDPF for $X \sim P_X$ under the distortion measure $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_0^+$ and divergence function $D : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}_0^+$ is defined as follows:*

$$R(D, P) = \min_{Q_{Y|X} \in \mathcal{L}(\mathcal{X} \times \mathcal{X})} I(X, Y) \tag{1.4.1}$$

$$s.t. \quad \mathbb{E}\left[d(X, Y)\right] \leq D \tag{1.4.2}$$

$$D(P_X \| Q_Y) \leq P \tag{1.4.3}$$

where $D \in [D_{\min}, D_{\max}] \subseteq [0, \infty)$, $P \in [P_{\min}, P_{\max}] \subseteq [0, \infty)$, and $Q_Y = m_Y(P_X Q_{Y|X})$.

In what follows, we remark on certain functional properties related to the Definition 1.4.1.

Remark 1.4.1. *(On Definition 1.4.1 - Functional properties in (D, P)) Following [9], it can be shown that* (1.4.1) *has some useful properties, under mild regularity conditions. In particular, [9, Theorem 1] showed that, for $D \in [D_{\min}, D_{\max}] \subset [0, \infty)$ and $P \in [P_{\min}, P_{\max}] \subset [0, \infty)$, $R(D, P)$ is (i) monotonically non-increasing function in both D and P; (ii) convex in both D and P if the divergence $D(\cdot\|\cdot)$ is convex in its second argument.*

Remark 1.4.2. *(On Definition 1.4.1 - Functional properties in $Q_{Y|X}$) The program defined by* (1.4.1)-(1.4.3) *is convex in the transition kernel $Q_{Y|X}$ for a given P_X if the divergence $D(\cdot\|\cdot)$ is convex in its second argument, since* (1.4.1) *and* (1.4.2) *are respectively convex and linear functionals in $Q_{Y|X}$ [40].*

Operational Aspects of the RDPF In the following section, we provide further details on the operational significance of the RDPF. Multiple works have shown the con-

nection between the RDPF various operational setups, considering both deterministic/stochastic codes with discrete/continuous sources and various formulation of perceptual qualities [31, 33, 34]. In the following, we report the setup proposed in [31], focusing on the asymptotic case. We remark, however, that similar results apply also for the case of one-shot codes, as presented in the same work.

We assume that we are given an i. i. d. sequence of n-length random variables $X^n \in \mathcal{X}^n$ with probability distribution $P_X \in \mathcal{P}(\mathcal{X})$. Formally, a stochastic encoder f_E^n is any function belonging to the set $\mathcal{F}_E^n = \{f : \mathcal{X}^n \times \mathbb{R} \to \mathbb{N}\}$, whereas a stochastic decoder g_D^n is any function in the set $\mathcal{G}_D^n = \{g : \mathbb{N} \times \mathbb{R} \to \mathcal{X}^n\}$. A stochastic code is an element of $\mathcal{F}_E^n \times \mathcal{G}_D^n$. Without loss of generality, the randomness at the encoder and decoder is represented as a single real number, i.e., an infinite number of bits, and is assumed shared by the pair, i.e., common randomness.

We let $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_0^+$ denote a single-letter distortion function and $D : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}_0^+$ denote a divergence measure. Moreover, we define the sets of fidelity criteria $\{\Delta_i\}_{i\in[1:n]}$ and $\{\Phi_i\}_{i\in[1:n]}$ as follows

$$\Delta_i \triangleq \mathbb{E}_{P_{X_i,Y_i}}\left[d(X_i, Y_i)\right], \qquad \Phi_i \triangleq D(P_{X_i}\|Q_{Y_i})$$

where Δ_i is the expected i^{th} symbol distortion and Φ_i is the i^{th} symbol divergence with respect to the reconstructed symbol Y_i. We are now ready to introduce the definition of achievability and that of the infimum of all achievable rates.

Definition 1.4.2. *(Achievability) Given a distortion level $D \geq 0$ and a perception constraint $P \geq 0$, a rate R is said to be (D, P)-achievable if there exists a random variable U and a sequence of codes $(f_E^n, g_D^n) \in \mathcal{F}_E^n \times \mathcal{G}_D^n$ with*

$$K_n = f_E^n(X^n, U), \quad Y^n = g_D^n(K_n, U)$$

such that, for $i = 1, \ldots, n$, the joint distribution p_{X_i,Y_i} satisfies $\Delta_i \leq D$ and $\Phi_i \leq P$ and

$$\lim_{n\to\infty} \frac{H(K_n|U)}{n} \leq R.$$

Then, we define

$$R_{cr}(D, P) \triangleq \inf\{R : R \text{ is } (D, P)\text{-achievable}\}.$$

The following theorem connects the operationally defined $R_{cr}(D, P)$ with $R(D, P)$ for general alphabets.

Theorem 1.4.1. *For $D \geq 0$, $P \geq 0$, we obtain $R_{cr}(D, P) = R(D, P)$.*

Proof. See [31, Theorem 3]. □

1.4.2 Statistical Divergences

Statistical divergences are fundamental measures ubiquitously used in information theory and statistics to quantify the dissimilarity between probability distributions, thus occupying a central role in the RDP framework. In their general definition, a divergence on $\mathcal{P}(\mathcal{X})$ is a function $D : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}_0^+$ such that $D(\mu\|\nu) \geq 0$ for all $\mu, \nu \in \mathcal{P}(\mathcal{X})$, holding with equality if and only if $\mu = \nu$.

The family of f-Divergences A particularly rich and widely studied class of divergences is the family of f-divergences, first introduced in [41] (see also [42]). This family is defined as follows.

Definition 1.4.3. *(f-divergence) Let $f : (0, \infty) \to \mathbb{R}$ be a convex function with $f(1) = 0$. Then the f-divergence $D_f(\cdot\|\cdot)$ associated with f is defined as*

$$D_f(\mu\|\nu) \triangleq \int_{\mathcal{X}} \nu(dx)\, f\left(\frac{d\mu}{d\nu}(x)\right), \qquad \mu, \nu \in \mathcal{P}(\mathcal{X})$$

under the conventions

$$\textit{(i)}\ f(0) = \lim_{x \to 0^+} f(x), \qquad \textit{(ii)}\ 0\, f\left(\frac{0}{0}\right) = 0, \qquad \textit{(iii)}\ \forall\, a \geq 0,\ 0\, f\left(\frac{a}{0}\right) = a\, f'(\infty).$$

Many commonly used divergences belong to this class. Notable examples include:

- *KL divergence* $D_{KL}(\cdot\|\cdot)$, obtained by setting $f(x) = x\log(x)$
- *Jensen-Shannon divergence* $D_{JS}(\cdot\|\cdot)$, where $f(x) = x\log\left(\frac{2x}{x+1}\right) + \log\left(\frac{2}{x+1}\right)$
- *Total Variation distance* $TV(\cdot\|\cdot)$, where $f(x) = \frac{1}{2}|x-1|$
- *α-divergence* $D_\alpha(\cdot\|\cdot)$, parameterized by $\alpha \in \mathbb{R}$:

$$D_\alpha(\mu\|\nu) = \int_{\mathcal{X}} \nu(dx)\, f_\alpha\left(\frac{d\mu}{d\nu}(x)\right), \qquad f_\alpha(x) = \begin{cases} \dfrac{x^\alpha - \alpha x - (1-\alpha)}{\alpha(\alpha-1)} & \alpha \neq 0,\, 1 \\ x\log(x) - x + 1 & \alpha = 1 \\ -\ln(x) + x - 1 & \alpha = 0 \end{cases}$$

The family of f-divergences also enjoys several useful structural properties. For any f-divergence $D_f(\cdot\|\cdot)$, the following hold:

- *(Linearity)* $D_{f_1+f_2}(\cdot\|\cdot) = D_{f_1}(\cdot\|\cdot) + D_{f_2}(\cdot\|\cdot)$
- *(Joint Convexity)* For any $t \in [0,1]$ and $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X})$,

$$D_f(t\mu_1 + (1-t)\mu_2 \,\|\, t\nu_1 + (1-t)\nu_2) \leq t\, D_f(\mu_1\|\nu_1) + (1-t)\, D_f(\mu_2\|\nu_2)$$

- *(Invariance)* Let $\hat{f}(x) = f(x) + c(x-1)$ for some $c \in \mathbb{R}$, then $D_f(\cdot\|\cdot) = D_{\hat{f}}(\cdot\|\cdot)$.

The characterization of the family of f-divergences here provided summarizes the properties useful for the scope of this thesis. For a more in-depth mathematical analysis, we refer the reader to [43].

Selected Divergences and Their Properties While the family of f-divergences provides a general and unified framework, some of the divergences employed in this thesis require a more detailed treatment. In what follows, we provide precise definitions and discuss the key properties of the KL divergence, the Geometric Jensen-Shannon divergence, the squared Hellinger distance, and the squared Wasserstein-2 distance.

Kullback–Leibler (KL) divergence ($\mathrm{D_{KL}}$)**.** The KL divergence is defined as:

$$\mathrm{D_{KL}}(\mu\|\nu) \triangleq \int_{\mathcal{X}} \mu(du) \log\left(\frac{d\mu}{d\nu}(u)\right), \qquad p_X, p_Y \in \mathcal{P}(\mathcal{X})$$

where $\mathrm{D_{KL}}(\mu\|\nu)$ is finite if and only if $\mu \ll \nu$, i.e., the Radon–Nikodym derivative $\frac{d\mu}{d\nu}$ is well defined. It is worth noting that, in general, $\mathrm{D_{KL}}(\mu\|\nu) \neq \mathrm{D_{KL}}(\nu\|\mu)$ [43, Remark 20]. Therefore, throughout the thesis, we will distinguish between the *direct* form $\mathrm{D_{KL}}(\mu\|\cdot)$ and the *reverse* form $\mathrm{D_{KL}}(\cdot\|\mu)$, depending on the placement of the source distribution μ.

Geometric Jensen-Shannon (GJS) divergence ($\mathrm{D_{GJS}}$)**.** The GJS divergence is a close relative of the classical Jensen-Shannon divergence, deviating from the latter in that the symmetrization of the KL divergence is induced by a geometric mean rather than an arithmetic mean. This distinction allows the GJS divergence to admit a closed-form expression for Gaussian distributions, which the classical JS divergence does not. Following [44, Theorem 2], it is defined as follows.

Definition 1.4.4 (GJS Divergence)**.** *Given $\eta, \nu \in \mathcal{P}(\mathcal{X})$, the GJS divergence is defined as:*

$$\mathrm{D_{GJS}}(\eta\|\nu) \triangleq \frac{1}{2}\,\mathrm{D_{KL}}(\eta\|\rho_g) + \frac{1}{2}\,\mathrm{D_{KL}}(\nu\|\rho_g) \tag{1.4.4}$$

where ρ_g denotes the geometric mean measure $\rho_g = \frac{1}{Z}\eta^{\frac{1}{2}}\nu^{\frac{1}{2}}$, with normalizing constant $Z = \int_{\mathcal{X}} \left(\frac{d\eta}{d\lambda}(u)\right)^{\frac{1}{2}} \left(\frac{d\nu}{d\lambda}(u)\right)^{\frac{1}{2}} \lambda(du)$ for some common dominating measure λ. In the Gaussian case, where $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \Sigma_Y)$, the geometric mean ρ_g is itself Gaussian with mean μ_g and covariance Σ_g given by:

$$\Sigma_g = \left(\frac{1}{2}\Sigma_X^{-1} + \frac{1}{2}\Sigma_Y^{-1}\right)^{-1} \tag{1.4.5}$$

$$\mu_g = \Sigma_g\left(\frac{1}{2}\Sigma_X^{-1}\mu_X + \frac{1}{2}\Sigma_Y^{-1}\mu_Y\right). \tag{1.4.6}$$

Squared Hellinger distance ($\mathrm{H^2}$)**.** Following [43], the squared Hellinger distance

is defined as:

$$\mathrm{H}^2(\mu,\nu) \triangleq \frac{1}{2}\int_{\mathcal{X}}\left(\sqrt{\frac{d\mu}{d\lambda}}-\sqrt{\frac{d\nu}{d\lambda}}\right)^2 d\lambda \tag{1.4.7}$$

where λ is any dominating measure for both μ and ν. Note that this definition is independent of the choice of λ. The squared Hellinger distance can be interpreted as the squared ℓ^2 norm between $\sqrt{\mu}$ and $\sqrt{\nu}$, yielding the upper bound

$$\mathrm{H}^2(\mu,\nu) \leq \left\|\sqrt{\frac{d\mu}{d\lambda}}\right\|_2^2 + \left\|\sqrt{\frac{d\nu}{d\lambda}}\right\|_2^2 = 2.$$

Squared Wasserstein-2 distance (W_2^2)**.** Unlike the divergences above, the squared Wasserstein-2 distance does not belong to the family of f-divergences but arises naturally from the theory of optimal transport [45, Chapter 7]. Originally introduced in [46], it is defined as:

$$\mathrm{W}_2^2(\mu,\nu) \triangleq \min_{\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} \|x-y\|^2 \,\pi(dx,dy) \tag{1.4.8}$$

where $\Pi(\mu,\nu)$ denotes the set of all joint measures π on $\mathcal{X}\times\mathcal{X}$ with marginals μ and ν.

Part I:

RDP Trade-off for Discrete Information Sources

Chapter 2

Blahut-Arimoto Type Algorithms for the RDPF

In this chapter, we develop a generic algorithmic approach for the computation of the RDPF, focusing on the case of discrete memoryless sources subject to a single-letter average distortion constraint and a perception constraint that belongs to the class of f-divergences. Our results leverage the fact that the RDPF forms a convex program under mild regularity conditions on the perception constraint (i.e., convexity on the second argument), which are satisfied by the considered class of divergences. This enables us to derive a parametric characterization of the optimal solution of the RDPF (Lemma 2.2.1), which is subsequently utilized to construct an alternating minimization procedure, thereon referred to as the Optimal Alternating Minimization (OAM) scheme, for which we also establish convergence guarantees (Theorem 2.2.1). However, the resulting structure of the OAM scheme relies on a set of implicit equations in the variables of interest, thus prohibiting the direct implementation of a generic BA algorithm similar to what is already known for the classical rate-distortion theory for i. i. d. sources and single-letter distortions [47]. Motivated by this technical difficulty, we propose two alternative minimization approaches that solve the implementability issue and whose applicability depends on the smoothness of the considered perception function.

- In Section 2.2.1, we design a *Newton-based Alternating Minimization* (NAM) scheme where, noticing that the solution of the OAM iterate is equivalent to a root-finding problem (Lemma 2.2.3), we use Newton's root-finding method [48] for the computation of the optimal iteration step (Theorem 2.2.2).
- In Section 2.2.2, we introduce a *Relaxed Alternating Minimization* (RAM) scheme, where we leverage a new relaxed formulation of the structure of the iterations in the OAM scheme and, subsequently, we derive necessary and sufficient conditions to ensure convergence to a globally optimal solution (Theorem 2.2.3).

In Section 2.3 we design the algorithmic embodiment for the NAM and RAM schemes (see Algorithm 1 and Algorithm 2, respectively) and develop suitable stopping criteria

for both algorithms (Theorem 2.3.1). Moreover, we also provide sufficient conditions on the structure of the distortion and the perception constraints based on which our algorithms converge exponentially fast in the number of iterations (Theorems 2.3.3 and 2.3.4). We corroborate our theoretical findings with numerical simulations (Section 2.4), with emphasis on the case of the TV perception metric, for which we develop a smooth approximation (Lemma 2.4.2).

2.1 Preliminaries

Given the relevance for this work, in this section we provide an overview of the alternating minimization methodology and its applications in information theory.

2.1.1 Alternating Minimization and BA-type algorithms

The alternating minimization method is a framework for the minimization of functions of two constrained variables. Consider the following optimization problem

$$\min_{x\in\mathcal{X},y\in\mathcal{Y}} f(x,y)$$

where $\mathcal{X}$ and $\mathcal{Y}$ are two arbitrary non-empty sets and the function $f(x,y)$ satisfies $-\infty < f(x,y) \leq +\infty$ for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Furthermore, we assume that, for each $x \in \mathcal{X}$, there exists $y \in \mathcal{Y}$ with $f(x,y) < +\infty$, implying that $s \coloneqq \inf_{x\in\mathcal{X},y\in\mathcal{Y}} f(x,y) < +\infty$. Depending on the case, the existence or uniqueness of the minimizer (x^*, y^*) such that $f(x^*, y^*) = s$ may also be assumed.

The goal of the alternating minimization method is to create a sequence $\{(x^{(n)}, y^{(n)})\}$ such that $\lim_{n\to\infty} f(x^{(n)}, y^{(n)}) = s$. Under specific conditions, such sequence can be defined using the solutions of two sub-problems; for $x_i \in \mathcal{X}$, $h(x_i) = \arg\min_{y\in\mathcal{Y}} f(x_i, y)$ and, for $y_i \in \mathcal{Y}$, $g(y_i) = \arg\min_{x\in\mathcal{X}} f(x, y_i)$. For some initial point $y^{(0)}$, we can define the n-th element of the sequence as:

$$x^{(n)} = g(y^{(n-1)}) \qquad y^{(n)} = h(x^{(n)}) \qquad \text{for } n = 1, 2 \ldots.$$

Depending on the problem, various sufficient conditions for the existence and optimality of the sequence limit have been studied. For instance,

- In [49], Csiszár and Tusnády prove that, if the sequence $\{(x^{(n)}, y^{(n)})\}$ guarantees

  $$f(x,y) + f(x, y^{(n-1)}) \geq f(x, y^{(n)}) + f(x^{(n)}, y^{(n-1)}) \quad \forall x \in \mathcal{X}, \forall y \in \mathcal{Y}$$

  referred to as *Five-Point property*, then the optimality of the limit is ensured.

- In [50], Grippo and Sciandrone prove convergence of the sequence to a stationary point of f, under the assumption of convexity of the feasible sets $\mathcal{X}$ and $\mathcal{Y}$ and

existence of the sequence limit.

BA-type algorithms, introduced for the numerical computation of channel capacity [51] and RD [47] tradeoffs, represent specific instances of alternating minimization algorithms [52, Chapter 9]. In fact, in their classic formulation, both problems can be expressed as constrained minimization of a convex function on the sets of marginal distributions and transition matrices, where the properties of the sets (e.g., convexity) depend on the constraints for which the problem is formulated.

2.2 Main Results

In this section, we present the derivation of our main theoretical results. We start by providing in the following lemma the parametric characterization of the solution of the RDPF problem, obtained by casting (1.4.1) as a double minimization problem.

Lemma 2.2.1. *(Double minimization) Let $D \geq 0$, $P \geq 0$ and let $D(\cdot\|\cdot) = D_f(\cdot\|\cdot)$. Moreover, let $s = (s_D, s_P)$ with $s_D \geq 0$, $s_P \geq 0$ being the Lagrangian multipliers associated with constraints* (1.4.2) *and* (1.4.3)*. Then* (1.4.1) *can be expressed as a double minimum as*

$$R(D,P) = \min_{\substack{Q_{Y|X}\in\mathcal{L}(\mathcal{X}^2)\\ r_Y\in\mathcal{P}(\mathcal{X})}} D_{KL}(p_X Q_{Y|X}\|p_X r_Y) + s_D\left(\mathbb{E}\left[d(x,y)\right] - D\right) + s_P(D_f(p_X\|q_Y) - P) \tag{2.2.1}$$

where $D = \mathbb{E}_{Q^*_{Y|X}}\left[d(x,y)\right]$ and $P = D_f(p_X\|r^*_Y)$, with $(Q^*_{Y|X}, r^*_Y)$ achieving the minimum. Furthermore, for fixed $Q_{Y|X}$, the right side of (2.2.1) *is minimized by*

$$r_Y[Q_{Y|X}](y) = \sum_{x\in\mathcal{X}} p_X(x) Q_{Y|X}(y|x) \tag{2.2.2}$$

while, for fixed r_Y, the right side of (2.2.1) *is minimized by*

$$Q_{Y|X}[r_Y](y|x) = \frac{r_Y(y)\cdot A[q_Y[r_Y]](x,y,s)}{\sum_{i\in\mathcal{X}} r_Y(i)\cdot A[q_Y[r_Y]](x,i,s)} \tag{2.2.3}$$

where

$$A[u](x,y,s) = \exp\left\{-s_D d(x,y) - s_P g(p_X(y), u(y))\right\} \tag{2.2.4}$$

$$q_Y[u](y) = \sum_{x\in\mathcal{X}} Q_{Y|X}[u](y|x) p_X(x) \tag{2.2.5}$$

$$g(x,y) = f\left(\frac{x}{y}\right) - \frac{x}{y}\partial f\left(\frac{x}{y}\right).$$

Proof. See Appendix 2.5.1. □

We remark that, although showing a close resemblance to the classical BA solution

[53, Theorem 6.3.3], Lemma 2.2.1 differs from it in (2.2.3). In particular, the perception constraint (1.4.3) induces the presence of an additional exponential term, i.e., $s_P g(\cdot,\cdot)$. Clearly, the classical BA implicit solution can be obtained as a special case of (2.2.3) by considering $s_P = 0$, effectively removing the perceptual constraint. The next corollary follows as a direct consequence of Lemma 2.2.1.

Corollary 2.2.1. *Let $s = (s_D, s_P)$ with $s_D \geq 0$, $s_P \geq 0$. Then $R(D,P)$in (2.2.1) can be reformulated as follows*

$$R(D_s,P_s) = -s_D D_s - s_P P_s + \min_{r_Y \in \mathcal{P}(\mathcal{X})} \quad s_P \sum_{y\in\mathcal{X}} p_X(y)\partial f\left(\frac{p_X(y)}{q_Y[r_Y](y)}\right) - \sum_{x\in\mathcal{X}} p_X(x)\log\left(\sum_{y\in\mathcal{X}} r_Y(y)A[q_Y[r_Y]](x,y,s)\right) \tag{2.2.6}$$

where

$$D_s = \sum_{(x,y)\in\mathcal{X}^2} \frac{p_X(x)r_Y^*(y)A[r_Y^*](x,y,s)}{\sum_{i\in\mathcal{X}} r_Y^*(i)A[r_Y^*](x,i,s)} d(x,y) \qquad P_s = D_f(p_X\|r_Y^*)$$

with $r_Y^* \in \mathcal{P}(\mathcal{X})$ achieving the minimum of (2.2.6).

Proof. The proof follows by substitution of (2.2.3) in (2.2.1). □

We note that in Corollary 2.2.1 and in subsequent analysis, the subscript notation (D_s, P_s) is introduced to explicitly indicate the dependence of the constraint levels (D,P) from the fixed Lagrangian multipliers $s = (s_D, s_P)$. The following lemma characterizes a necessary and sufficient condition to ensure that, for given Lagrangian multipliers $s = (s_D, s_P)$, a distribution $r_Y^* \in \mathcal{P}(\mathcal{X})$ is the optimal solution of (2.2.6), i.e., $(r_Y^*, Q_{Y|X}[r_Y^*])$ defines a point achieving the RDPF.

Lemma 2.2.2. *Let $D_f(\cdot\|\cdot)$ be such that $f \in C^1(0,\infty)$ continuous and differentiable on $(0,\infty)$ and let the vector function $c[\cdot,\cdot] : \mathbb{P}(\mathcal{X})^2 \to \mathbb{R}^{|\mathcal{X}|}$ be such that*

$$c[u,r](y) = \sum_{x\in\mathcal{X}} \frac{p_X(x)A[r](x,y,s)}{\sum_{i\in\mathcal{X}} u(i)A[r](x,i,s)}. \tag{2.2.7}$$

Then, a probability vector r_Y yields a point on the $R(D,P)$ curve via the transition matrix $Q_{Y|X}$ defined in (2.2.3) *if and only if $c[r_Y, q_Y[r_Y]](y) \leq 1$ for all $y \in \mathcal{X}$, holding with equality for any y for which $r_Y(y)$ is nonzero.*

Proof. See Appendix 2.5.2. □

Remark 2.2.1. *It can be shown that the function $c[\cdot,\cdot]$ characterizes also the relation between a distribution r_Y and the result of the functional $q_Y[r_Y]$. In fact, we can verify*

that for all $i \in \mathcal{X}$,

$$\frac{q_Y[r_Y](i)}{r_Y(i)} = \sum_{x\in\mathcal{X}} \frac{Q_{Y|X}[r_Y](i|x)}{r_Y(i)} p_X(x) = c[r_Y, q_Y[r_Y]](i).$$

Using the results of Lemma 2.2.1, we now proceed to construct an alternating minimization procedure, thereon referred to as Optimal Alternating Minimization (OAM), proving its convergence to a point of $R(D,P)$.

Theorem 2.2.1. *(OAM) Let the Lagrangian multipliers* $s = (s_D, s_P)$ *with* $s_D \geq 0$, $s_P \geq 0$ *be given. Let* $r_Y^{(0)}$ *denote any probability vector with nonzero components and let* $Q_{Y|X}^{(n+1)} \equiv Q_{Y|X}[r_Y^{(n)}]$ *and* $r_Y^{(n+1)} \equiv q_Y[r_Y^{(n)}]$ *be functions of the current iteration* $r_Y^{(n)}$ *as defined in* (2.2.3) *and* (2.2.5)*, respectively. Then, as* $n \to \infty$*, we obtain*

$$D(Q_{Y|X}^{(n)}) \to D_s,\ P(Q_{Y|X}^{(n)}) \to P_s,\ I(p_X, Q_{Y|X}^{(n)}) \to R(D_s, P_s).$$

Proof. See Appendix 2.5.3. □

Despite being optimal, the OAM scheme does not allow the implementation of a BA-type algorithmic embodiment. The reason stands in the parametric dependencies underlying (2.2.3) and (2.2.5), as highlighted in the following remark.

Remark 2.2.2. *(Implicit Iterate) Due to the structure of the iterations in* (2.2.3) *and* (2.2.5)*, an implicit dependency of* $r_Y^{(n+1)}$ *on itself appears, i.e.,*

$$\frac{r_Y^{(n+1)}(y)}{r_Y^{(n)}(y)} = \sum_{x\in\mathcal{X}} \frac{p_X(x)\exp\left\{-s_D d(x,y) - s_P g\left(p_X, r_Y^{(n+1)}, y\right)\right\}}{\sum_{i\in\mathcal{X}} r_Y^{(n)}(i)\exp\left\{-s_D d(x,i) - s_P g\left(p_X, r_Y^{(n+1)}, i\right)\right\}} = c[r_Y^{(n)}, r_Y^{(n+1)}](y) \tag{2.2.8}$$

showing that the updated term $r_Y^{(n+1)}$ *exists in both the left- and the right-hand side of the equation. Consequently, the structure of* (2.2.8) *impedes the characterization of a closed form expression of the "updated term"* $r_Y^{(n+1)}$ *as a function of only the current iteration term* $r_Y^{(n)}$.

The implementation problem of the OAM scheme prompts us to find alternative ways to compute the alternating minimization iterates. We detail in the following section two different approaches to solve the OAM issue, leveraging either the numerical solution of the implicit equation or through the relaxation of the structure of the iterations.

2.2.1 NAM scheme

The implicit definition of $r_Y^{(n+1)}$ in (2.2.8) suggests the application of numerical methods for its approximation. To this end, we introduce a variation to the OAM scheme, referred to as NAM scheme, where $r_Y^{(n+1)}$ is approximated at each minimization step using Newton's root-finding method [48].

We first demonstrate that the iteration step for $r_Y^{(n+1)}$, i.e., (2.2.8), can be cast as a root finding problem.

Lemma 2.2.3. *Let $r_Y^{(n+1)}$ be defined as in Theorem 2.2.1 and let $T : \mathbb{R}^{|\mathcal{X}|} \to \mathbb{R}^{|\mathcal{X}|}$ be the vector function defined as*

$$T[r_Y^{(n)}, u](i) \triangleq u(i) - r_Y^{(n)}(i) \cdot c[r_Y^{(n)}, u](i), \ \forall i \in \mathcal{X} \tag{2.2.9}$$

where $c[\cdot,\cdot]$ is defined in (2.2.7)*. Then, $r_Y^{(n+1)}$ is a root of $T[r_Y^{(n)}, \cdot]$, i.e., $T[r_Y^{(n)}, r_Y^{(n+1)}] = 0$.*

Proof. The proof follows from the evaluation of (2.2.9) in $r_Y^{(n+1)}$ and the substitution of (2.2.8) therein. □

The application of Newton's method requires the existence and the invertibility of the Jacobian matrix J_T of the functional T [48, Section 10.2]. In our case, ensuring the existence of J_T requires a more restrictive continuity assumption on the divergence measure, i.e., $D_f(\cdot\|\cdot)$ needs to be twice differentiable in its second argument. Although this limitation reduces the generality of the NAM scheme, we note that the majority of the commonly used divergences, i.e., see Section 1.4.2, respect this assumption. Under this restriction, the invertibility of J_T is shown in the following lemma.

Lemma 2.2.4. *Let $T[r_Y^{(n)}, \cdot]$ be the function defined in* (2.2.9) *and let the divergence measure $D_f(\cdot\|\cdot)$ be twice differentiable in its second argument. Then, the Jacobian $J_T : \mathbb{R}^{|\mathcal{X}|} \to \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ of the functional $T[r_Y^{(n)}, \cdot]$, defined as $J_T[r_Y^{(n)}, u] \triangleq \left[\frac{\partial T[r_Y^{(n)}, v](i)}{\partial v(j)}\Big|_u\right]_{(i,j)\in\mathcal{X}^2}$, is positive definite and has the form*

$$J_T[r_Y^{(n)}, u] = I + \Big(C[r_Y^{(n)}, u] - M[r_Y^{(n)}, u]\Big) \cdot \Gamma[r_Y^{(n)}, u] \tag{2.2.10}$$

where

$$M[r_Y^{(n)}, u] = \left[r_Y^{(n)}(i) \sum_{x\in\mathcal{X}} p_X(x) \frac{A[u](x,i,s) \cdot A[u](x,j,s)}{\left(\sum_{k\in\mathcal{X}} r_Y^{(n)}(k) A[u](x,k,s)\right)^2}\right]_{(i,j)\in\mathcal{X}^2} \tag{2.2.11}$$

$$\Gamma[r_Y^{(n)}, u] = s_P \operatorname{diag}\left[r_Y^{(n)}(i) \cdot \frac{\partial^2 D_f(p_X\|v)}{\partial v(i)^2}\Big|_u\right]_{i\in\mathcal{X}} \tag{2.2.12}$$

$$C[r_Y^{(n)}, u] = \operatorname{diag}\Big[c[r_Y^{(n)}, u](i)\Big]_{i\in\mathcal{X}}. \tag{2.2.13}$$

Proof. See Appendix 2.5.4. □

We are now ready to define the structure of the iteration of Newton's root-finding method applied to the functional $T(\cdot)$, which, as shown in Lemma 2.2.3, provides an approximation of $r_Y^{(n+1)}$.

Theorem 2.2.2. *(Newton's method) Assume the divergence measure $D_f(\cdot\|\cdot)$ to be twice differentiable in its second argument and let $r_Y^{(n+1)}$ and $r_Y^{(n)}$ be defined as in Theorem 2.2.1.*

Let $T[r_Y^{(n)}, \cdot]$ *and* $J_T[r_Y^{(n)}, \cdot]$ *be as defined in* (2.2.9) *and* (2.2.10)*, respectively. Furthermore, let the sequence* $\{u^{(k)}\}_{k=1,2,\dots}$ *for some initial point* $u^{(0)} \in \mathbb{R}^{|\mathcal{X}|}$ *be defined as*

$$u^{(k+1)} \triangleq u^{(k)} - \left(J_T[r_Y^{(n)}, u^{(k)}]\right)^{-1} \cdot T[r_Y^{(n)}, u^{(k)}].$$

Then, $\lim_{k\to\infty} u^{(k)} = r_Y^{(n+1)}$.

Proof. The proof follows by direct application of Newton's root-finding method [48, Section 10.2], since Lemma 2.2.3 proves that the set of solutions $r_Y^{(n+1)}$ and the set of the roots of $T(\cdot)$ coincide, while Lemma 2.2.4 proves that $T(\cdot)$ satisfies the assumption for the convergence of the method. □

The implementation of the NAM algorithm illustrated in Algorithm 1 is obtained by introducing the results of Theorem 2.2.2 in the OAM scheme defined in Theorem 2.2.1. However, despite solving the main technical issues of the OAM scheme, the NAM scheme imposes limitations on the choice of the perception metric. In the next section, we provide an alternative minimization scheme that circumvents these issues.

2.2.2 RAM scheme

An alternative approach to solve the implementation problems of the OAM scheme stems from a relaxed formulation of the OAM iterations. Through the introduction of an auxiliary design variable v in (2.2.3), we define an approximation to the original OAM scheme, referred to as the RAM scheme. The main advantage of the RAM scheme stands in the fact that, for v properly selected, the iterative scheme is directly implementable and does not require additional assumptions on the continuity of the perception constraints, while still being able to achieve a globally optimal solution. The following theorem provides the formal formulation of the RAM iterative scheme.

Theorem 2.2.3. *(RAM) Let the Lagrangian multipliers* $s = (s_D, s_P)$ *with* $s_D \geq 0$, $s_P \in [0, s_{P,\max}]$ *be given and define*

$$\hat{Q}_{Y|X}[u](y|x) \triangleq \frac{u(y)A[v[u]](x,y,s)}{\sum_{i\in\mathcal{X}} u(i)A[v[u]](x,i,s)} \tag{2.2.14}$$

$$\hat{q}_Y[u](y) \triangleq \sum_{x\in\mathcal{X}} \hat{Q}_{Y|X}[u](y|x)p_X(x) \tag{2.2.15}$$

where $A[\cdot]$ *is defined in* (2.2.4) *and* $v[\cdot] : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ *is any functional defining a probability distribution. Let* $\hat{r}_Y^{(0)}$ *be any probability vector with nonzero components and let* $\hat{Q}_{Y|X}^{(n+1)} \equiv \hat{Q}_{Y|X}[\hat{r}_Y^{(n)}]$, $\hat{r}_Y^{(n+1)} \equiv \hat{q}_Y[\hat{r}_Y^{(n)}]$, *and* $v^{(n)} = v[\hat{r}_Y^{(n)}]$. *Then, as* $n \to \infty$, *we obtain*

$$D(\hat{Q}_{Y|X}^{(n)}) \to D_s,\ P(\hat{Q}_{Y|X}^{(n)}) \to P_s,\ I(p_X, \hat{Q}_{Y|X}^{(n)}) \to R(D_s, P_s)$$

if $\lim_{n\to\infty} \|r_Y^{(n+1)} - v^{(n)}\| = 0$ *with at least linear rate of convergence.*

Proof. See Appendix 2.5.5. □

Theorem 2.2.3 enables the implementation of the alternating minimization scheme by introducing an auxiliary variable $v[r_Y^{(n)}]$, which approximates the correct iteration $r_Y^{(n+1)}$ while still being a function of only the current iteration of $r_Y^{(n)}$. Nevertheless, depending on v, this approximation may incur restrictions on the domain of the Lagrangian multiplier s_P that affect convergence guarantees, as we will later discuss in Section 2.3.

We conclude this section with the following technical remark, highlighting the differences between the NAM and RAM schemes.

Remark 2.2.3. *(NAM vs. RAM) The main advantage of NAM is the convergence for any value of the Lagrangian multipliers (s_D, s_P), without needing any additional condition. However, the introduction of Newton's method requires the differentiability of the perceptual metric and the introduction of additional complexity at each iteration. RAM, on the other hand, lifts the differentiability requirement and avoids the additional computational cost at each iteration, but at the expense of a potentially smaller set of (s_D, s_P) for which the algorithm achieves the optimal solution, potentially prohibiting the computation of the complete RDP curve.*

2.3 Algorithmic Implementation and Convergence Analysis

This section is about the algorithmic implementation of the alternating minimization schemes derived in Section 2.2 and the characterization of their convergence rate.

We start by presenting the implementation of the NAM and RAM schemes, respectively, in Algorithm 1 and 2. Subsequently, we discuss the derivation of stopping conditions suitable for both algorithms.

Algorithm 1 Newton-based Alternating Minimization (NAM)

Require: source distribution p_X; Lagrangian parameters $s = (s_D, s_P)$ with $s_D \geq 0$ and $s_P \geq 0$; error tolerances $\epsilon > 0$, $\delta > 0$; distortion measure $d(x, y)$; initial assignment $r_Y^{(0)}$.

1: $\omega \leftarrow +\infty$; $n \leftarrow 0$;
2: **while** $\omega \geq \epsilon$ **do**
3: $\quad r_Y^{(n+1)} \leftarrow$ Newton Approximation(p_X, $r_Y^{(n)}$, s, δ)
4: $\quad c^{(n)} \leftarrow c\left[r_Y^{(n)}, r_Y^{(n+1)}\right]$
5: $\quad \omega \leftarrow \log c_{\max}^{(n)}(y) - \sum_{y\in\mathcal{X}} r_Y^{(n)} c^{(n)}(y) \log(c^{(n)}(y))$
6: $\quad n \leftarrow n + 1$
7: **end while**

Ensure: $D_s = \mathbf{E}_{p_X Q_{Y|X}^{(n)}}[d(x, y)]$, $P_s = D_f(p_X, r_Y^{(n)})$, $R(D_s, P_s) = \hat{W}[r_Y^{(n)}] - s_D D_s - s_P P_s - \sum_{y\in\mathcal{X}} r_Y^{(n)} c^{(n)} \log(c^{(n)})$, $\hat{W}[\cdot] =$ (2.3.3).

Algorithm 2 Relaxed Alternating Minimization (RAM)

Require: source distribution p_X; Lagrangian multipliers $s = (s_D, s_P)$ with $s_D \geq 0$ and $s_P \in [0, s_{P,\max}]$; error tolerance $\epsilon > 0$; divergence measure $D_f(\cdot\|\cdot)$; distortion measure $d(\cdot,\cdot)$; initial assignment $\hat{r}_Y^{(0)}$.

1: $\omega \leftarrow +\infty$; $n \leftarrow 0$;
2: **while** $\omega \geq \epsilon$ **do**
3: $\quad c^{(n)} \leftarrow c[\hat{r}_Y^{(n)}, v^{(n)}]$
4: $\quad \hat{r}_Y^{(n+1)} \leftarrow \hat{r}_Y^{(n)} \cdot c^{(n)}$
5: $\quad \omega \leftarrow \log c_{\max}^{(n)}(y) - \sum_{y\in\mathcal{X}} \hat{r}_Y^{(n)} c^{(n)}(y) \log(c^{(n)}(y))$
6: $\quad n \leftarrow n + 1$
7: **end while**

Ensure: $D_s = \mathbf{E}_{p_X \hat{Q}_{Y|X}^{(n)}}[d(x,y)]$, $P_s = D_f(p_X, \hat{r}_Y^{(n)})$, $R(D_s, P_s) = \hat{W}[r_Y^{(n)}] - s_D D_s - s_P P_s - \sum_{y\in\mathcal{X}} r_Y^{(n)} c^{(n)} \log(c^{(n)})$, $\hat{W}[\cdot] = (2.3.3)$.

Stopping Criterion We first focus on the derivation of a stopping criterion for the RAM case, as the NAM case can be obtained by fixing the auxiliary variable $v[r_Y^{(n)}] = q_Y[r_Y^{(n)}]$ in Theorem 2.2.3, i.e., recovering the original OAM iterates. For this purpose, we need the following theorem in which we derive bounds on the RDPF.

Theorem 2.3.1. *(Bounds on RDPF) Let $\hat{Q}_{Y|X}$ and $\hat{q}_Y$ be defined as in Theorem 2.2.3, $c(y)$ be as defined as in Lemma 2.2.2, and $c_{\max} = \max_{y\in\mathcal{X}} c[\hat{r}_Y, \hat{q}_Y[\hat{r}_Y]](y)$. Then, at the point $D = \mathbb{E}_{p_X \hat{Q}_{Y|X}}[d(x,y)]$, and $P = D_f(p_X \| \hat{q}_Y[\hat{r}_Y])$, the following bounds hold*

$$R(D,P) \geq R_L[\hat{r}_Y](D,P) = -s_D D - s_P P + \hat{W}[\hat{r}_Y] - \log(c_{\max}) \tag{2.3.1}$$

$$\begin{aligned} R(D,P) \leq R_U[\hat{r}_Y](D,P) = &-s_D D - s_P P + \hat{W}[\hat{r}_Y] \\ &- \sum_{y\in\mathcal{X}} \hat{r}_Y(y) c[\hat{r}_Y, \hat{q}_Y[\hat{r}_Y]](y) \log(c[\hat{r}_Y, \hat{q}_Y[\hat{r}_Y]](y)) \end{aligned} \tag{2.3.2}$$

where $\hat{W}[\cdot]$ is given by

$$\begin{aligned} \hat{W}[u] = &- \sum_{x\in\mathcal{X}} p_X(x) \log\left(\sum_{y\in\mathcal{X}} u(y) A[v[u]](x,y,s)\right) + s_P \sum_{y\in\mathcal{X}} \hat{q}_Y[u](y) \frac{p_X(y)}{v[u](y)} \partial f\left(\frac{p_X(y)}{v[u](y)}\right) \\ &+ s_P \left[\sum_{y\in\mathcal{X}} \hat{q}_Y[u] \left(f\left(\frac{p_X(y)}{\hat{q}_Y[u](y)}\right) - f\left(\frac{p_X(y)}{v[u](y)}\right)\right)\right]. \end{aligned} \tag{2.3.3}$$

Proof. See Appendix 2.5.6. □

Leveraging the bounds in (2.3.1) and (2.3.2), we can estimate the precision of the estimation of $R(D,P)$ at the n-th iteration by considering the estimation error $\omega = R_U[\hat{r}_Y^{(n)}](D,P) - R_L[\hat{r}_Y^{(n)}](D,P)$, as implemented in line 5 of both Algorithm 1 and 2.

2.3.1 Asymptotic Convergence Rate Analysis

In this section, we characterize the asymptotic convergence rate of the proposed minimization schemes. We start with the analysis of the convergence rate of the OAM scheme, which, although not directly implementable, serves as a reference for the characterization of the convergence rate of both the NAM and RAM schemes.

OAM Convergence Rate We remark that the iteration structure in Theorem 2.2.1, i.e., $r_Y^{(n+1)} = q_Y[r_Y^{(n)}]$ given the current iteration n, can be described as an implicit vector function $S : \mathbb{R}^{|\mathcal{X}|} \to \mathbb{R}^{|\mathcal{X}|}$ with $S[r_Y](i) = r_Y(i) \cdot c[r_Y, S[r_Y]](i)$, such that $r_Y^{(n+1)} = S[r_Y^{(n)}]$. The results of Lemma 2.2.2 characterize a distribution r_Y^* achieving the RDPF as a fixed point of $S(r_Y)$, i.e. $r_Y^* = S[r_Y^*]$, since $c[r_Y^*, S[r_Y^*]](i) = 1,\ i = 1\ldots, |\mathcal{X}|$. Under these observations, we can analyze the convergence rate of the OAM scheme following similar steps as in [54].

The first order Taylor expansion of $S[r_Y]$ around a fixed point r_Y^* is defined as

$$S[r_Y] = S[r_Y^*] + J[r_Y^*] \cdot (r_Y - r_Y^*) + o(\|r_Y - r_Y^*\|)$$

where $J[r_Y]$ is the Jacobian matrix of $S[r_Y]$ with entries $J[r_Y](i,j) \triangleq \frac{\partial S[r_Y](i)}{\partial r_Y(j)}, (i,j) \in \mathcal{X}^2$. The next theorem provides the functional form of the Jacobian for the case of Theorem 2.2.1.

Theorem 2.3.2. *(Jacobian form) The Jacobian $J(r_Y)$ computed at the fixed point r_Y^* is given as*

$$J[r_Y^*] = (I - M^*)\,(I - \Gamma^* J[r_Y^*]) \tag{2.3.4}$$

where $M^* = M[r_Y^*, r_Y^*]$ and $\Gamma^* = \Gamma[r_Y^*, r_Y^*]$ as defined in (2.2.11) *and* (2.2.12)*, respectively.*

Proof. See Appendix 2.5.7. □

Next, we introduce two lemmas, in which we use the structure of (2.3.4) to identify properties of matrix M^*.

Lemma 2.3.1. *Let $\lambda_{M^*} = \{\lambda_i\}_{i\in[1:|\mathcal{X}|]}$ be the set of eigenvalues of $M^* = M[r_Y^*, r_Y^*]$. Given a distortion function $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$ that induces a full-rank matrix $D = [e^{-s_D d(i,j)}]_{(i,j)\in\mathcal{X}^2}$, then $\lambda_i > 0, \forall i \in [1 : |\mathcal{X}|]$, i.e., M^* has only positive eigenvalues.*

Proof. See Appendix 2.5.8. □

Remark 2.3.1. *(On Lemma 2.3.1) We note that a popular example that satisfies the assumptions imposed on Lemma 2.3.1 is the Hamming distortion denoted hereinafter as d_H [55].*

Lemma 2.3.2. *Let $\lambda_{M^*} = \{\lambda_i\}_{i\in[1:|\mathcal{X}|]}$ be the set of eigenvalues of $M^* = M[r_Y^*, r_Y^*]$. Then, we have that $\lambda_i \leq 1, \forall i \in [1 : |\mathcal{X}|]$.*

Proof. See Appendix 2.5.9. □

Using Lemmas 2.3.1 and 2.3.2, we can characterize the interval that contains the set of eigenvalues of $J[r_Y^*]$ and subsequently the convergence rate of Theorem 2.2.1.

Theorem 2.3.3. *(Convergence rate of Theorem 2.2.1) Let $\{\theta_i\}_{i\in[1:|\mathcal{X}|]}$ be the eigenvalues of $J[r_Y^*]$. Then,*

$$0 \leq \{\theta_i\}_{i\in[1:|\mathcal{X}|]} < 1.$$

Moreover, let $\gamma \in [\theta_{\max}, 1)$. Then, there exists $\delta > 0$ and $K > 0$ such that if $r_Y^{(0)} \in \{r_Y : \|r_Y - r_Y^*\| \leq \delta\}$, we obtain

$$\|r_Y^{(n)} - r_Y^*\| < K \cdot \|r_Y^{(0)} - r_Y^*\| \cdot \gamma^n \tag{2.3.5}$$

i.e., the iterations converge exponentially.

Proof. See Appendix 2.5.10. □

Summarizing, under the structural constraints on the distortion function d reported in Lemma 2.3.1, the exponential convergence of the OAM scheme is guaranteed by Theorem 2.3.3.

NAM Convergence Rate The convergence rate of the NAM scheme is immediate from the OAM scheme analysis, given the close relation between the two schemes. Since the only difference lies in the introduction of Newton's root-finding method for the estimation of the optimal iteration step, the NAM scheme enjoys the same convergence rate in terms of the number of iterations as the OAM scheme, i.e., an exponential convergence $\mathcal{O}(e^{-n})$ under the assumptions of Lemma 2.3.1. However, the added complexity from the application of Newton's method at each iteration increases the overall iteration complexity, due to the at least linear convergence rate $\mathcal{O}(\frac{1}{m})$ of the root approximant. Therefore, the total complexity is approximately $\mathcal{O}(\frac{e^{-n}}{m})$, where n and m depend on the error tolerances ϵ and δ given as input in Algo. 1.

RAM Convergence Rate Following similar steps that led to Theorem 2.3.3, the Jacobian $\hat{J}(r_Y^*)$ associated with the iteration scheme in Theorem 2.2.3 is characterized as

$$\hat{J}[r_Y^*] = (I - M[r_Y^*, r_Y^*])(I - \Gamma[r_Y^*, r_Y^*])$$

where M and Γ are given by (2.2.11) and (2.2.12), respectively. Unlike Theorem 2.2.1, where the structure of (2.3.4) bounds its own eigenvalues, in this case, we need to bound the Lagrangian multiplier s_P, hence matrix Γ, to guarantee exponential convergence of the algorithm. This is proved in the following theorem.

Theorem 2.3.4. *For a given $s_D \geq 0$, let $I_{s_P} = [0, s_{P,\max}]$ be the domain of s_P, $\{\theta_{a,i}\}_{i\in\mathcal{X}}$ the set of eigenvalues of $\hat{J}(r_Y^*)$ and $\theta_{\max}$ the maximum eigenvalue of $J(r_Y^*)$ in (2.3.4). Define the set $I_{s_P}^{\epsilon} = [0, s_{P,\max} - \epsilon]$ for $0 < \epsilon < s_{P,\max}$. Then, there exists an ϵ' such that if $s_P \in I_{s_P}^{\epsilon'}$ then $0 \leq \{\theta_{a,i}\}_{i\in\mathcal{X}} < 1$.*

Proof. See Appendix 2.5.11. □

Theorem 2.3.4 guarantees exponential convergence for Theorem 2.2.3 only for $s_P \in I_{s_P}^{\epsilon}$ which means that we can consider $P \in [P_{\min}(\epsilon), P_{\max}]$, depending on the characteristics of the input in a specific problem.

2.4 Numerical Results

In this section, we validate our theoretical findings with simulations. In particular, we consider the computation of the RDPF under the NAM and RAM schemes, respectively using Algorithms 1 and 2.

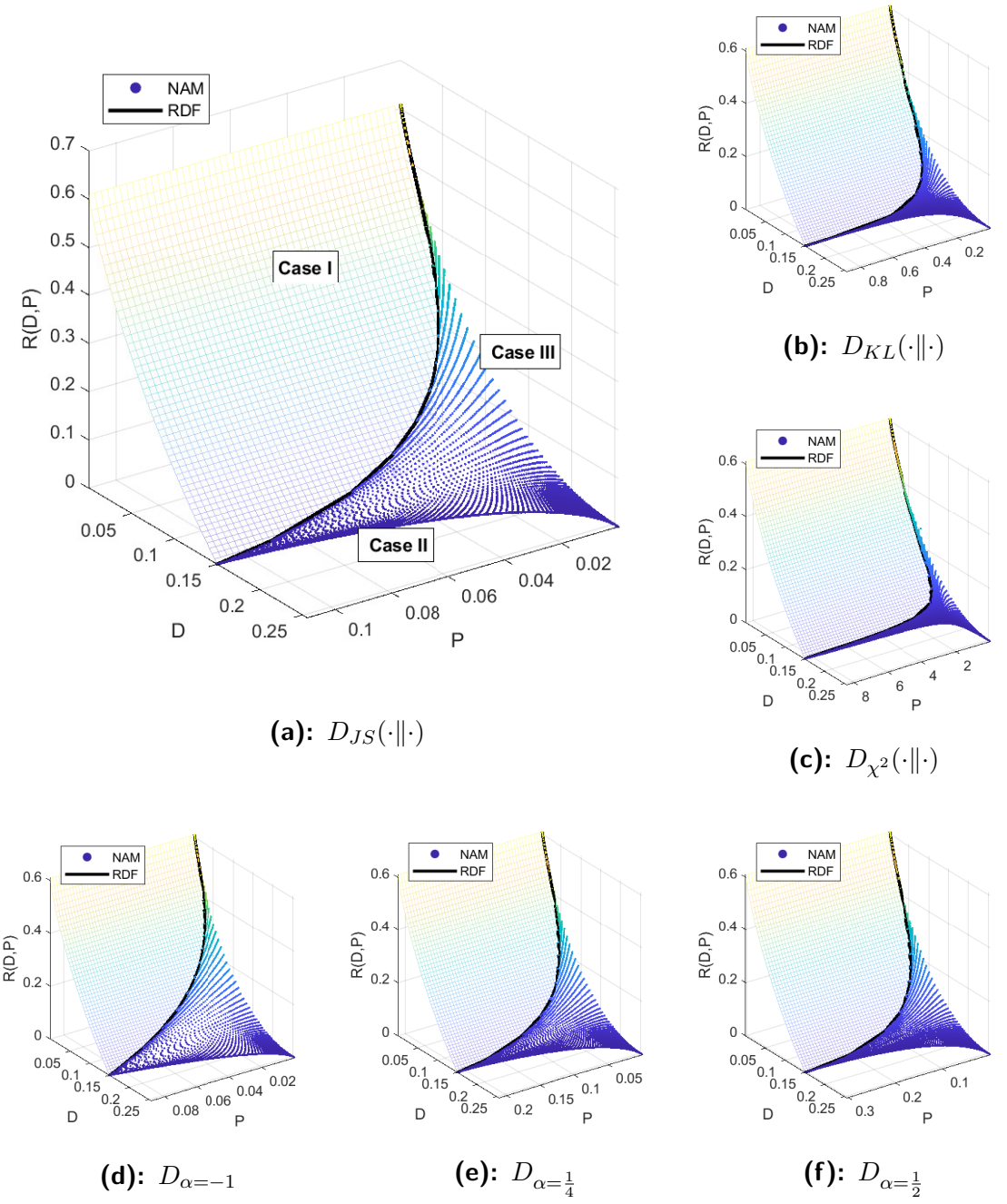
2.4.1 RDPF Computation - NAM scheme

Example 2.4.1. *Suppose that $\mathcal{X} = \{0, 1\}$ with $p_X \sim Ber(0.15)$, and let $d(\cdot,\cdot) = d_H(\cdot,\cdot)$ with perception constraint chosen to be either (a) $D_f(\cdot\|\cdot) = D_{JS}(\cdot\|\cdot)$, (b) $D_f(\cdot\|\cdot) = D_{KL}(\cdot\|\cdot)$, (c) $D_f(\cdot\|\cdot) = D_{\chi^2}(\cdot\|\cdot)$, (d) $D_f(\cdot\|\cdot) = D_{\alpha=-1}(\cdot\|\cdot)$, (e) $D_f(\cdot\|\cdot) = D_{\alpha=\frac{1}{4}}(\cdot\|\cdot)$, (f) $D_f(\cdot\|\cdot) = D_{\alpha=\frac{1}{2}}(\cdot\|\cdot)$. In Fig. 2.1, we display the $R(D,P)$ estimates resulting from Algorithm 1 for each divergence metric.*

We observe that all the computed RDPFs share similarities in the structure of the operating regions on the (D, P) plane. Taking as a reference the plot of Fig. 2.0(a), we distinguish three cases:

- *Case I*, where the perception level P is not met with equality. In this region, the RDPF is equal to the RDF.
- *Case II*, where, due to the distortion level D not met with equality, the RDPF function is identically zero.
- *Case III*, where both the distortion level D and perception level P are met with equality.

The results show that Algorithm 1 can completely cover the *Case III* region. The boundary between Cases I-III (black line) representing the RDF is obtained by setting the Lagrangian multiplier $s_P = 0$. Furthermore, the *Case I* region can be obtained by extension of the relative boundaries, since all the (D, P) points share the solution given by the RD problem.



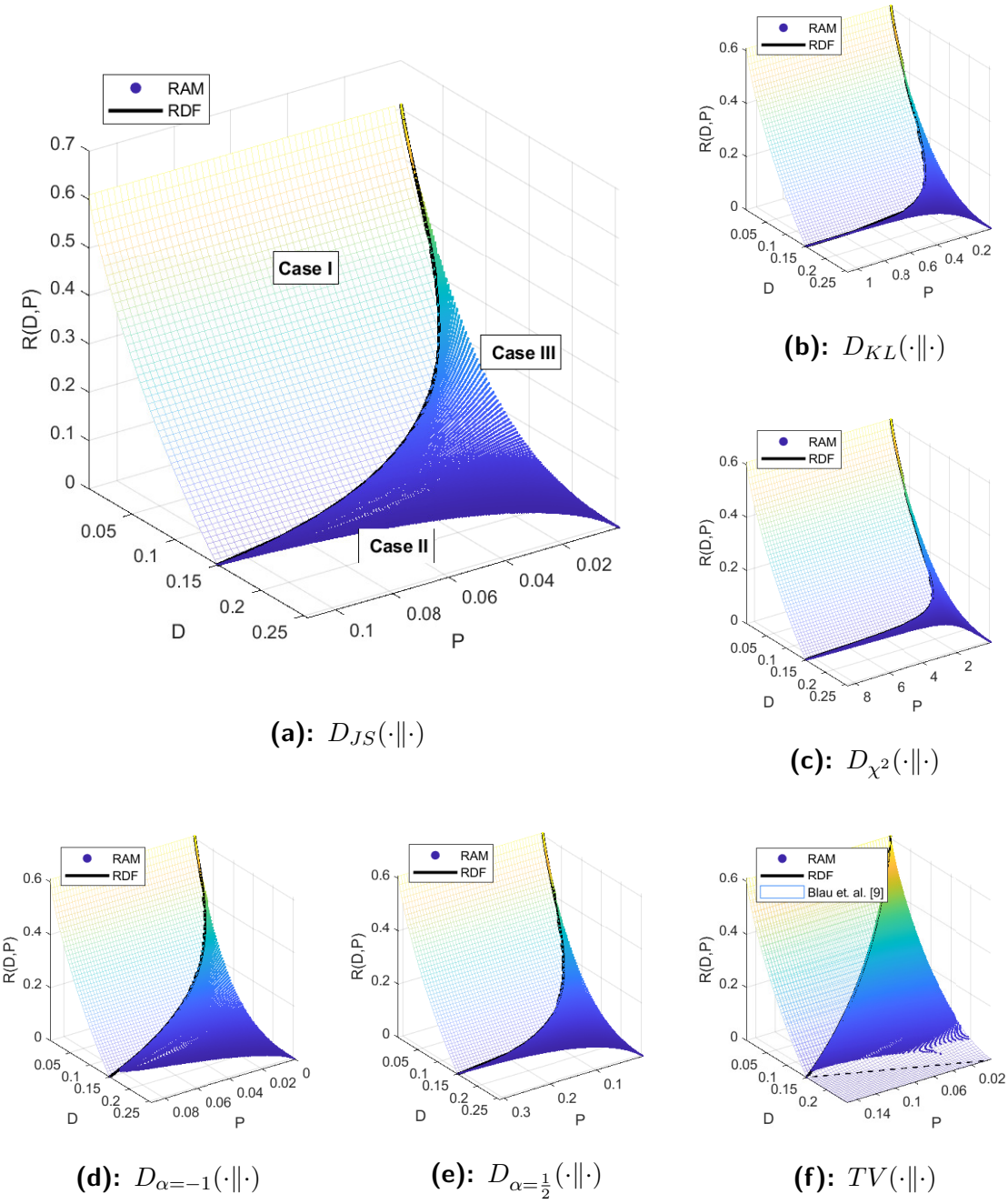
(a): $D_{JS}(\cdot\|\cdot)$ **(b):** $D_{KL}(\cdot\|\cdot)$ **(c):** $D_{\chi^2}(\cdot\|\cdot)$

(d): $D_{\alpha=-1}$ **(e):** $D_{\alpha=\frac{1}{4}}$ **(f):** $D_{\alpha=\frac{1}{2}}$

Figure 2.1: $R(D,P)$ **for a Bernoulli source under a Hamming distortion and various perception constraints.**

2.4.2 RDPF Computation - RAM scheme

Example 2.4.2. *Suppose that* $\mathcal{X} = \hat{\mathcal{X}} = \{0,1\}$ *with* $p_X \sim Ber(0.15)$*, and let* $d(\cdot,\cdot) = d_H(\cdot,\cdot)$ *with perception constraint chosen to be either (a)* $D_f(p_X\|r_Y) = D_{JS}(\cdot\|\cdot)$*, (b)* $D_f(\cdot\|\cdot) = D_{KL}(\cdot\|\cdot)$*, (c)* $D_f(\cdot\|\cdot) = D_{\chi^2}(\cdot\|\cdot)$*, (d)* $D_f(\cdot\|\cdot) = D_{\alpha=-1}(\cdot\|\cdot)$*, (e)* $D_f(\cdot\|\cdot) = D_{\alpha=\frac{1}{2}}(\cdot\|\cdot)$*, (f)* $D_f(\cdot\|\cdot) = TV(\cdot\|\cdot)$*. In Fig. 2.2, we display the* $R(D,P)$ *estimates resulting from Algorithm 2 for each divergence metric.*



(a): $D_{JS}(\cdot\|\cdot)$

(b): $D_{KL}(\cdot\|\cdot)$

(c): $D_{\chi^2}(\cdot\|\cdot)$

(d): $D_{\alpha=-1}(\cdot\|\cdot)$

(e): $D_{\alpha=\frac{1}{2}}(\cdot\|\cdot)$

(f): $TV(\cdot\|\cdot)$

Figure 2.2: $R(D,P)$ **for a Bernoulli source under a Hamming distortion and various perception constraints.**

Remarks similar to those in Example 2.4.1 can be made regarding the operating regions of the computed RDPFs. In addition, in Fig. 2.1(f) we compare the theoretical results of [9, Equation 6] with the numerical results obtained using Algorithm 2. This case is of particular interest due to the non-differentiability of the TV distance. We observe that Algorithm 2 achieves exactly the theoretical solution of [9, Equation 6] as long as $D \leq D_{\max} = 0.15$. We attribute this behavior to the limitations in the domain of the Lagrangian multiplier s_P, as discussed in Theorem 2.3.4. We address this issue in detail

in the following section.

2.4.3 On the convergence under TV perception constraint

As observed in Example 2.4.2, in the case of TV distance the RAM algorithm is not able to converge to proper solutions in the region $\Omega = \{(D, P) \in \mathbb{R}_+^2 : D \geq p = 0.15\}$. We argue that the problem stems from the values of the Lagrangian multipliers $s = (s_D, s_P)$ associated with Ω requires values of s_P that do not guarantee the convergence of the algorithm, as reported in Theorem 2.3.4.

To solve the issue, we propose an approximation of the $TV(\cdot\|\cdot)$ distance through a sequence of f-divergences $\{D_{f_n}\}_{n\in\mathbb{N}}$ such that $D_{f_n} \to TV$ for $n \to \infty$. We start with the following general property.

Lemma 2.4.1. *For a divergence metric $D_f(\cdot\|\cdot)$, let the set $\mathcal{L}_{D_f}(D, P)$ be defined as*

$$\mathcal{L}_{D_f}(D, P) \triangleq \{Q_{Y|X} \in \mathcal{L}(\mathcal{X}^2) : \mathbb{E}_{Q_{Y|X}}[d(x, y)] \leq D \wedge D_f(p_X\|q_Y) \leq P\}.$$

Given D_f, D_g divergence metrics with $D_f(p\|q) \leq D_g(p\|q), \forall p, q \in \mathcal{P}$, then $\mathcal{L}_{D_g}(D, P) \subseteq \mathcal{L}_{D_f}(D, P)$. Moreover, for the associated RDPF problems

$$R_{D_f}(D, P) = \min_{Q_{Y|X}\in\mathcal{L}_{D_f}(D,P)} I(p_X, Q_{Y|X}) \quad \textit{and} \quad R_{D_g}(D, P) = \min_{Q_{Y|X}\in\mathcal{L}_{D_g}(D,P)} I(p_X, Q_{Y|X})$$

the inequality $R_{D_g}(D, P) \geq R_{D_f}(D, P)$ holds.

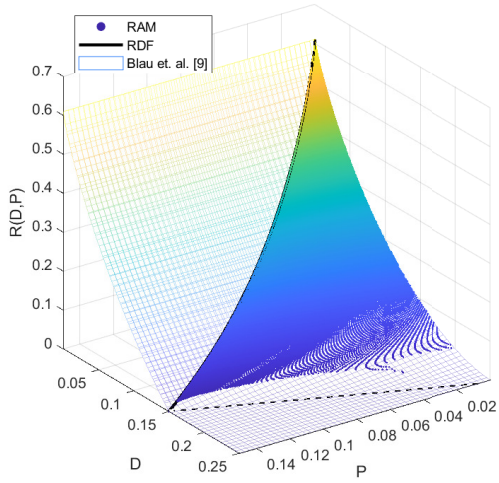
Proof. The inequality $R_{D_g}(D, P) \geq R_{D_f}(D, P)$ holds if $\mathcal{L}_{D_g} \subseteq \mathcal{L}_{D_f}$, which is a trivial implication of $D_f(p\|q) \leq D_g(p\|q), \forall p, q \in \mathcal{P}$. □

We can now characterize a sequence $\{D_{f_n}\}$ such that $D_{f_n} \to TV$ for $n \to \infty$ and $D_{f_n} \leq TV$.

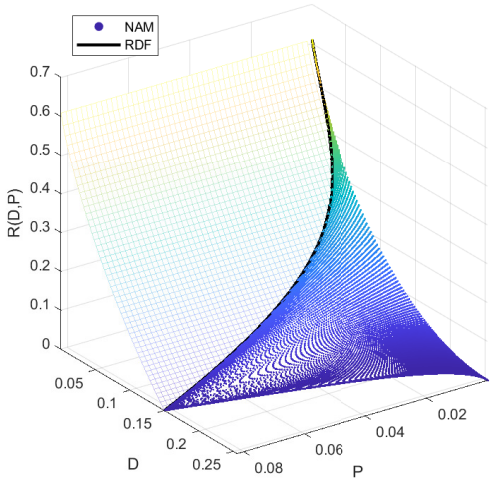
Lemma 2.4.2. *Let f_n be the sequence of convex functions defined as $f_n(x) = \frac{2}{\pi}(x - 1)\arctan(n(x - 1))$ and let $\{D_{f_n}(\cdot\|\cdot)\}_{n=1,2,\dots}$ be the sequence of associated f-divergences. Then, for $n \to \infty$, $D_{f_n} \to TV$ uniformly. Furthermore, for $n = 1, 2, \dots$ and for all $p, q \in \mathcal{P}(\mathcal{X})$, $D_{f_n}(p\|q) \leq TV(p\|q)$.*

Proof. See Appendix 2.5.12. □

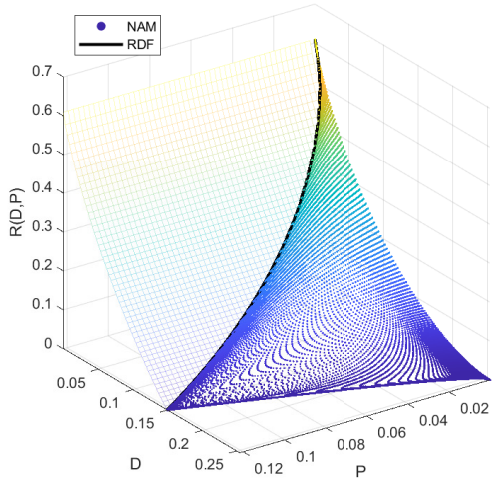
The results from Lemmas 2.4.1 and 2.4.2 guarantee that, $\forall n \in \mathbb{N}$, the RDP problem defined for the perception metric $D_{f_n}(\cdot\|\cdot)$ act as lower-bound for the RDP problem defined with $TV(\cdot\|\cdot)$ perception metric. Furthermore, since $D_{f_n}(\cdot\|\cdot)$ is a smooth function, we can apply NAM scheme to compute the associate RDPF. Fig. 2.3 shows the RDPFs for a source $X \sim Ber(0.15)$ under Hamming distortion and perception measure D_{f_n}, for $n = \{1, 10, 100\}$. As expected, for increasing n, the D_{f_n} perception metric becomes a better approximation of the $TV(\cdot\|\cdot)$ perception.



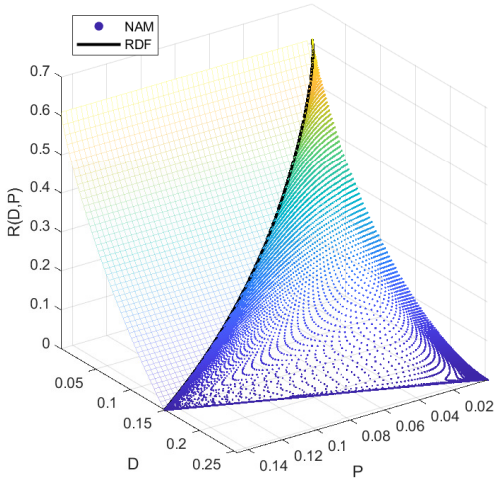
(a): $TV(\cdot\|\cdot)$



(b): $n = 1$



(c): $n = 10$



(d): $n = 100$

Figure 2.3: Comparison between the RDPF under TV perception, computed with the RAM scheme, and the RDPF under D_{f_n}, computed with the NAM scheme, for $n = \{1, 10, 100\}$.

2.5 Appendix

2.5.1 Proof of Lemma 2.2.1

The double minimization follows immediately from properties of the KL divergence, reported in [53, Theorem 5.2.6], which allows rewriting (1.4.1) as (2.2.1). Moreover, the characterization (2.2.2) of the minimizer r_Y follows from the same properties.

On the other hand, we remark that (1.4.1) is a convex program in the variable $Q_{Y|X}$ for

a given p_X, and respects Slater's condition, since it is easy to show that $Q_{Y|X}(y|x) = \delta_{y=x}$ is an interior point of the constraint set, as highlighted in Remark 1.4.2. Therefore, the minimizer $Q^*_{Y|X}$ can be characterized by applying the Karush-Kuhn-Tucker (KKT) conditions [56] on the dual formulation of (1.4.1).

The Lagrangian associated with the primal problem has the form:

$$L(Q_{Y|X}, s, \gamma, \zeta) = D_{KL}\left(p_X Q_{Y|X} \| p_X r_Y\right) + s_D\left(\mathbb{E}\left[d(x,y)\right] - D\right) + s_P\left(D_f(p_X\|q_Y) - P\right) + \sum_{x\in\mathcal{X}} \gamma_x \left(1 - \sum_{y\in\mathcal{X}} Q_{Y|X}(y|x)\right) - \sum_{(x,y)\in\mathcal{X}^2} \zeta_{x,y} Q_{Y|X}(y|x)$$

where the last two sets of constraints refer to the positivity and normalization constraints on $Q_{Y|X}$. Differentiating $L(Q_{Y|X}, s, \gamma, \zeta)$ with respect to the primal variables $Q_{Y|X}(y|x), \forall (x,y) \in \mathcal{X}^2$, we obtain:

$$\frac{\partial L(Q_{Y|X}, s, \gamma, \zeta)}{\partial Q_{Y|X}(y|x)} = \frac{\partial D_{KL}(p_X Q_{Y|X}\|p_X r_Y)}{\partial Q_{Y|X}(y|x)} + s_D \frac{\partial E[d(X,Y)]}{\partial Q_{Y|X}(y|x)} + s_P \frac{\partial D_f(p_X\|q_Y)}{\partial Q_{Y|X}(y|x)} + \gamma_x - \zeta_{x,y}$$

where:

$$\frac{\partial D_{KL}(p_X Q_{Y|X}\|p_X r_Y)}{\partial Q_{Y|X}(y|x)} = p_X(x)\left(\log(\frac{Q_{Y|X}(y|x)}{r_Y(y)}) + 1\right)$$
$$\frac{\partial E[d(X,Y)]}{\partial Q_{Y|X}(y|x)} = p_X(x) d(x,y)$$
$$\frac{\partial D_f(p_X\|q_Y)}{\partial Q_{Y|X}(y|x)} = p_X(x) g(p_X, q_Y, y).$$

Enforcing *stationarity* and *complementary slackness*

$$\begin{cases} \dfrac{\partial L(Q_{Y|X}, s, \gamma, \zeta)}{\partial Q_{Y|X}(y|x)} = 0 \\ \zeta(x,y) Q_{Y|X}(y|x) = 0 \end{cases} \tag{2.5.1}$$

we solve for $Q_{Y|X}(y|x)$ while choosing $\gamma(x)$ such that $\sum_{y\in\mathcal{X}} Q_{Y|X}(y|x) = 1$, $\forall x \in \mathcal{X}$, resulting in (2.2.3) and (2.2.4). This completes the proof.

2.5.2 Proof of Lemma 2.2.2

Using the results of Corollary 2.2.1, we can apply KKT conditions to (2.2.6). Thus, a minimum for r_Y must satisfy:

$$\frac{\partial}{r_Y(y)}\left[- \sum_{x\in\mathcal{X}} p_X(x) \log\left(\sum_{y\in\mathcal{X}} r_Y(y) A[r_Y](x,y,s)\right) + s_P \sum_{y\in\mathcal{X}} p_X(y) \partial f\left(\frac{p_X(y)}{q_Y[r_Y](y)}\right) + \gamma \sum_{y\in\mathcal{X}} r_Y(y)\right] \geq 0$$

which reduces to

$$\gamma - c[r_Y, q_Y[r_Y]](y) \geq 0.$$

The Lagrangian multiplier γ is evaluated by multiplying by r_Y and summing over $y \in \mathcal{X}$, giving $\gamma = 1$. This concludes the proof.

2.5.3 Proof of Theorem 2.2.1

Let $\mathbf{h} = (r_Y^{(0)}, r_Y^{(1)}, \ldots)$, and $\mathbf{Q} = (Q_{Y|X}^{(1)}, Q_{Y|X}^{(2)}, \ldots)$ be the sequences of probability vectors and transition matrices obtained by the chain of alternating minimization $r_Y^{(n)} \to Q_{Y|X}^{(n+1)} \to r_Y^{(n+1)}$. Let $A^{(n)} = A[r_Y^{(n+1)}]$ and define the functionals $V[r_Y^{(n)}]$ and $W[r_Y^{(n)}]$ as

$$V[r_Y^{(n)}] = D_{KL}\left(p_X \cdot Q_{Y|X}^{(n+1)} \| p_X \cdot r_Y^{(n)}\right) + s_D \mathbf{E}_{Q_{Y|X}^{(n+1)} p_X}[d(x,y)] + s_P D_f(p_X \| r_Y^{(n+1)}) \tag{2.5.2}$$

$$W[r_Y^{(n)}] = s_P \sum_{y \in \mathcal{X}} p_X(y) \partial f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - \sum_{x \in \mathcal{X}} p_X(x) \log\left(\sum_{y \in \mathcal{X}} r_Y^{(n)}(y) A^{(n)}(x,y,s)\right). \tag{2.5.3}$$

Using Theorem 2.2.1 and Corollary 2.2.1, we can show that $V[\mathbf{q}]$ is non-increasing on the sequence $\mathbf{q}$ by first fixing $r_Y^{(n)}$ and minimizing over $\mathbf{Q}$ (resulting in $Q_{Y|X}^{(n+1)}$) and then fixing $Q_{Y|X}^{(n+1)}$ and minimizing over $\mathbf{q}$ (resulting in $r_Y^{(n+1)}$). $W[r_Y^{(n+1)}]$ is the result of the minimization over $\mathbf{Q}$ between $V[r_Y^{(n)}]$ and $V[r_Y^{(n+1)}]$, resulting in a non-increasing sequence $V[r_Y^{(n)}] \geq W[r_Y^{(n)}] \geq V[r_Y^{(n+1)}]$. Given $V[\mathbf{q}]$ non-increasing and bounded from below, $V[\mathbf{q}]$ converges to some number V^∞.

Let now r_Y^* be any probability vector and $Q_{Y|X}^* = Q_{Y|X}[r_Y^*]$ such that $R(D,P) = V[r_Y^*] - s_D D - s_P P$ and let

$$\begin{aligned}\sum_{(x,y) \in \mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(n)}}{Q_{Y|X}^{(n+1)}}\right) = &\sum_{(x,y) \in \mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(n)}}{r_Y^{(n)}}\right) \\ &- \sum_{(x,y) \in \mathcal{X}^2} p_X Q_{Y|X}^* \log\left(A^{(n)}(x,y,s)\right) \\ &+ \sum_{(x,y) \in \mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\sum_{i \in \mathcal{X}} r_Y^{(n)}(i) A^{(n)}(x,i,s)\right)\end{aligned} \tag{2.5.4}$$

where

$$\sum_{(x,y) \in \mathcal{X}^2} p_X Q_{Y|X}^* \log\left(A^{(n)}(x,y,s)\right) = -s_D \mathbb{E}_{Q_{Y|X}^*}[d(x,y)] - s_P \sum_{y \in \mathcal{X}} r_Y^*(y) f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right)$$

$$+ s_P \sum_{y\in\mathcal{X}} r_Y^*(y)\frac{p_X(y)}{r_Y^{(n+1)}(y)}\partial f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right).$$

We can introduce an upper bound to (2.5.4) by noticing that

$$\begin{aligned}\sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(n)}}{r_Y^{(n)}}\right) - \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^*}{r_Y^*}\right)\\ = \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(n)}\cdot r_Y^*}{r_Y^{(n)}\cdot Q_{Y|X}^*}\right) \le \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^{(n)} \log\left(\frac{r_Y^*}{r_Y^{(n)}}\right) - 1 = 0.\end{aligned} \tag{2.5.5}$$

As a result, we obtain

$$\begin{aligned}&\sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(n)}}{Q_{Y|X}^{(n+1)}}\right) \le\\
&\le \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^*}{r_Y^*}\right) + s_D\mathbb{E}_{Q_{Y|X}^*}\left[d(x,y)\right] + s_P D_f(p_X\|r_Y^*)\\
&\quad + s_P\left[\sum_{y\in\mathcal{X}} r_Y^*(y) f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - D_f\left(p_X\|r_Y^*\right)\right]\\
&\quad - \left[-\sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\sum_{i\in\mathcal{X}} r_Y^{(n)}(i)A^{(n)}(x,i,s)\right) + s_P\sum_{y\in\mathcal{X}} r_Y^{(n+1)}(y)\frac{p_X(y)}{r_Y^{(n+1)}(y)}\partial f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right)\right]\\
&\quad + s_P\left[\sum_{y\in\mathcal{X}} r_Y^{(n+1)}(y)\frac{p_X(y)}{r_Y^{(n+1)}(y)}\partial f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - \sum_{y\in\mathcal{X}} r_Y^*(y)\frac{p_X(y)}{r_Y^{(n+1)}(y)}\partial f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right)\right]\\
&\le V[r_Y^*] - W[r_Y^{(n)}] + s_P\left[\sum_{y\in\mathcal{X}} r_Y^*(y) f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - D_f(p_X\|r_Y^*)\right]\\
&\quad + s_P\left[\sum_{y\in\mathcal{X}} r_Y^{(n+1)}(y)\frac{p_X(y)}{r_Y^{(n+1)}(y)}\partial f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - \sum_{y\in\mathcal{X}} r_Y^*(y)\frac{p_X(y)}{r_Y^{(n+1)}(y)}\partial f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right)\right].\end{aligned}$$

Notice that for any iteration $n \ge 0$, we have $W[r_Y^{(n)}] \ge V[r_Y^{(n+1)}] \ge V[r_Y^*]$ and, subsequently, the following inequalities

$$0 \le W[r_Y^{(n)}] - V[r_Y^*] \le G[r_Y^{(n)}] \tag{2.5.6}$$

where

$$\begin{aligned}G[r_Y^{(n)}] = &\sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(n+1)}}{Q_{Y|X}^{(n)}}\right) + s_P\left\{\mathbb{E}_{r_Y^*}\left[f\left(\frac{p_X}{r_Y^{(n+1)}}\right)\right] - D_f\left(p_X\|r_Y^*\right)\right\}\\
&+ s_P\left\{\mathbb{E}_{r_Y^{(n+1)}}\left[\frac{p_X}{r_Y^{(n+1)}}\partial f\left(\frac{p_X}{r_Y^{(n+1)}}\right)\right] - \mathbb{E}_{r_Y^*}\left[\frac{p_X}{r_Y^{(n+1)}}\partial f\left(\frac{p_X}{r_Y^{(n+1)}}\right)\right]\right\}.\end{aligned}$$

Since $s_P \geq 0$ and due to the fact that

$$\begin{aligned}
&\mathbb{E}_{r_Y^*}\left[f\left(\frac{p_X}{r_Y^{(n+1)}}\right)\right] - D_f\left(p_X \| r_Y^*\right) + \mathbb{E}_{r_Y^{(n+1)}}\left[\frac{p_X}{r_Y^{(n+1)}}\partial f\left(\frac{p_X}{r_Y^{(n+1)}}\right)\right] - \mathbb{E}_{r_Y^*}\left[\frac{p_X}{r_Y^{(n+1)}}\partial f\left(\frac{p_X}{r_Y^{(n+1)}}\right)\right] \\
&= D_f(p_X \| r_Y^{(n+1)}) - D_f(p_X \| r_Y^{(*)}) - \sum_{i\in\mathcal{X}}\left(r_Y^*(i) - r_Y^{(n+1)}(i)\right) f\left(\frac{p_X(i)}{r_Y^{(n+1)}}\right) \\
&\quad + \sum_{i\in\mathcal{X}}\left(r_Y^*(i) - r_Y^{(n+1)}(i)\right)\frac{p_X(i)}{r_Y^{(n+1)}}\partial f\left(\frac{p_X(i)}{r_Y^{(n+1)}}\right) \\
&= D_f\left(p_X \| r_Y^{(n+1)}\right) - D_f\left(p_X \| r_Y^{(*)}\right) - \sum_{i\in\mathcal{X}}\left(r_Y^{(n+1}(i) - r_Y^*(i)\right)\left.\frac{\partial D_f(p_X\|u)}{\partial u(i)}\right|_{r_Y^{(n+1)}} \\
&= -\left(D_f\left(p_X \| r_Y^*\right) - T_{D_f(p\|\cdot),r_Y^{(n+1)}}(r_Y^*)\right) \overset{(a)}{\leq} 0
\end{aligned}$$

where $T_{D_f(p\|\cdot),r_Y^{(n+1)}}(r_Y^*))$ is the first order Taylor expansion of $D_f(p_X\|\cdot)$, centered in $r_Y^{(n+1)}$ and evaluated in $r_Y^{(*)}$, and (a) is verified since $D_f(p_X\|\cdot)$ is a convex function in its second argument. Therefore $\forall n \in \mathbf{N}$, we obtain

$$W[r_Y^{(n)}] - V[r_Y^*] \leq \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(n+1)}}{Q_{Y|X}^{(n)}}\right).$$

Summing over N terms, we obtain

$$\begin{aligned}
\sum_{n=1}^{N}(W[r_Y^{(n)}] - V[r_Y^*]) &\leq \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \sum_{n=1}^{N}\log\left(\frac{Q_{Y|X}^{(n+1)}}{Q_{Y|X}^{(n)}}\right) \\
&= \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{(N+1)}}{Q_{Y|X}^{(1)}}\right) \\
&\overset{(b)}{\leq} \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{Q_{Y|X}^{*}}{Q_{Y|X}^{(1)}}\right)
\end{aligned}$$

where (b) follows using the logarithm inequality.
Since at any iteration n, $W[r_Y^{(n)}] - V[r_Y^*] \geq 0$ and for all integers $N > 0$ the partial sum is upper-bounded by a constant $L(r_Y^*, r_Y^{(0)})$ dependent only on the initial probability assignment $r_Y^{(0)}$, we have that $\lim_{N\to\infty}\sum_{n=1}^{N}(W[r_Y^{(n)}] - V[r_Y^*])$ exists and it is finite hence $\lim_{n\to\infty} W[r_Y^{(n)}] - V[r_Y^*] = 0$. This completes the proof.

2.5.4 Proof of Lemma 2.2.4

We first derive the functional form of J_T:

$$\frac{\partial T[r_Y^{(n)}, u](i)}{\partial u(j)} = \delta_{i,j} - r_Y^{(n)}(i)\sum_{x\in\mathcal{X}} p_X(x)\frac{\frac{\partial A[u](x,i,s)}{\partial u(j)}}{\sum_{k\in\mathcal{X}} r_Y^{(n)}(k)A[u](x,k,s)}$$

$$+ r_Y^{(n)}(i) \sum_{x\in\mathcal{X}} p_X(x) \frac{A[u](x,i,s)\left(\sum_{k\in\mathcal{X}} r_Y(k)\frac{\partial A[u](x,k,s)}{\partial u(j)}\right)}{\left(\sum_{k\in\mathcal{X}} r_Y^{(n)}(k)A[u](x,k,s)\right)^2}$$

$$\frac{\partial A[u](x,i,s)}{\partial u(j)} = A[u](x,i,s)\left(-s_P \frac{\partial^2 D(p_X\|v)}{\partial v(i)^2}\bigg|_u\right)\delta_{i,j}.$$

By defining matrices M, Γ and C, respectively, as in (2.2.11), (2.2.12), (2.2.13), the matrix $J_T(\cdot)$ can be rewritten as in (2.2.10). To prove its invertibility, we need to ensure that 0 is not part of the set of eigenvalues of $J_T(y)$, i.e., $0 \notin \lambda_{J_T(y)},\ \forall y \in \mathbb{R}^{|\mathcal{X}|}$. Noticing that $M[r_Y^{(n)},u](i,j) \geq 0, \forall (i,j) \in \mathcal{X}^2$ and $\sum_{i\in\mathcal{X}} M[r_Y^{(n)},u](i,j) = C[r_Y^{(n)},u](i)$, we can define the sets D_i as:

$$D_i = \{\lambda \in \mathbb{R} : \left|\lambda - \left(C[r_Y^{(n)},u](i) - M[r_Y^{(n)},u](i,i)\right)\right| \leq C[r_Y^{(n)},u](i) - M[r_Y^{(n)},u](i,i)\} \tag{2.5.7}$$

Since $\bigcup_{i\in\mathcal{X}} D_i \subseteq \mathbb{R}_0^+$, we can apply Gershgorin Circle Theorem [29] to prove that $C[r_Y^{(n)},u] - M[r_Y^{(n)},u]$ has only non-negative eigenvalues. Moreover, since $D(p_X\|\cdot)$ is a convex function in its second argument, $\Gamma[r_Y^{(n)},u]$ is a positive semi-definite matrix. Therefore, we obtain

$$J_T[r_Y^{(n)},u] = I + \left(C[r_Y^{(n)},u] - M[r_Y^{(n)},u]\right)\cdot\Gamma[r_Y^{(n)},u] \geq I > 0$$

This concludes the proof.

2.5.5 Proof of Theorem 2.2.3

Let $V[\cdot]$, $W[\cdot]$ be the functionals defined in (2.5.2) and (2.5.3), respectively. Moreover, let $\hat{W}[\cdot]$ be a functional obtained by the alternating sequence $\hat{r}_Y^{(n)} \to \hat{Q}_{Y|X}^{(n+1)} \to \hat{r}_Y^{(n+1)}$ substituting $\hat{Q}_{Y|X}$ with fixed $\hat{r}_Y$ as follows

$$\begin{aligned}\hat{W}[\hat{r}_Y^{(n)}] = &- \sum_{x\in\mathcal{X}} p_X(x)\log\left(\sum_{y\in\mathcal{X}} \hat{r}_Y^{(n)}(y)A^{(n)}(x,y,s)\right) + s_P \sum_{y\in\mathcal{X}} \hat{r}_Y^{(n+1)}(y)\frac{p_X(y)}{v^{(n)}(y)}\partial f\left(\frac{p_X(y)}{v^{(n)}(y)}\right)\\ &+ s_P\left[\sum_{y\in\mathcal{X}} \hat{r}_Y^{(n+1)} f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - \sum_{y\in\mathcal{X}} \hat{r}_Y^{(n+1)} f\left(\frac{p_X(y)}{v^{(n)}(y)}\right)\right]\end{aligned} \tag{2.5.8}$$

where $A^{(n)} = A[v^{(n)}]$. Similarly to Theorem 2.2.1, we let r_Y^* be any probability vector and $Q_{Y|X}^* = Q_{Y|X}[r_Y^*]$ such that $R(D,P) = V[r_Y^*] - s_D D - s_P P$ and consider that

$$\sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{\hat{Q}_{Y|X}^{(n)}}{\hat{Q}_{Y|X}^{(n+1)}}\right) = \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{\hat{Q}_{Y|X}^{(n)}}{\hat{r}_Y^{(n)}}\right) - \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(A^{(n)}(x,y,s)\right) + \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\sum_{i\in\mathcal{X}} \hat{r}_Y^{(n)}(i) A^{(n)}(x,i,s)\right).$$

Substituting the definition of $A(x,y,s)$ and using (2.5.5), we obtain

$$\begin{aligned}\sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}^* \log\left(\frac{\hat{Q}_{Y|X}^{(n)}}{\hat{Q}_{Y|X}^{(n+1)}}\right) \leq \quad & V[r_Y^*] - \hat{W}[\hat{r}_Y^{(n)}] \\ + s_P \Bigg[& \sum_{y\in\mathcal{X}} r_Y^*(y) f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) - D_f(p_X\|r_Y^*) + \sum_{y\in\mathcal{X}} r_Y^{(n+1)}(y)\frac{p_X(y)}{v^{(n)}(y)}\partial f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) \\ & - \sum_{y\in\mathcal{X}} r_Y^*(y)\frac{p_X(y)}{v^{(n)}(y)}\partial f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) + \sum_{y\in\mathcal{X}} \hat{r}_Y^{(n+1)} f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - \sum_{y\in\mathcal{X}} \hat{r}_Y^{(n+1)} f\left(\frac{p_X(y)}{v^{(n)}(y)}\right)\Bigg].\end{aligned} \tag{2.5.9}$$

Note that the right-hand side of (2.5.9) can be bounded by

$$\begin{aligned}&\sum_{y\in\mathcal{X}} r_Y^*(y) f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) - D_f\left(p_X\|r_Y^*\right) + \sum_{y\in\mathcal{X}} r_Y^{(n+1)}(y)\frac{p_X(y)}{v^{(n)}(y)}\partial f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) \\ &\quad - \sum_{y\in\mathcal{X}} r_Y^*(y)\frac{p_X(y)}{v^{(n)}(y)}\partial f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) + \sum_{y\in\mathcal{X}} \hat{r}_Y^{(n+1)} f\left(\frac{p_X(y)}{r_Y^{(n+1)}(y)}\right) - \sum_{y\in\mathcal{X}} \hat{r}_Y^{(n+1)} f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) \\ &= D_f\left(p_X\|v^{(n)}\right) - D_f\left(p_X\|r_Y^*\right) + \sum_{i\in\mathcal{X}} (r_Y^* - v^{(n)})\left[f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) - \frac{p_X(y)}{v^{(n)}(y)}\partial f\left(\frac{p_X(y)}{v^{(n)}(y)}\right)\right] \\ &\quad + D_f\left(p_X\|r_Y^{(n+1)}\right) - D_f\left(p_X\|v^{(n)}\right) + \sum_{i\in\mathcal{X}} \left(v^{(n)} - r_Y^{(n+1)}\right)\left[f\left(\frac{p_X(y)}{v^{(n)}(y)}\right) - \frac{p_X(y)}{v^{(n)}(y)}\partial f\left(\frac{p_X(y)}{v^{(n)}(y)}\right)\right] \\ &= -\left[D_f\left(p_X\|r_Y^*\right) - T_{D_f(p\|\cdot),v^{(n)}}(r_Y^*)\right] + \left[D_f(p_X\|r_Y^{(n+1)}) - T_{D_f(p\|\cdot),v^{(n)}}\left(r_Y^{(n+1)}\right)\right] \\ &\overset{(a)}{\leq} D_f\left(p_X\|r_Y^{(n+1)}\right) - T_{D_f(p\|\cdot),v^{(n)}}\left(r_Y^{(n+1)}\right)\end{aligned}$$

where $T_{D_f(p\|\cdot),v^{(n)}}\left(r_Y^{(n+1)}\right)$ is the first order Taylor expansion of $D_f(p_X\|\cdot)$, centered in $v^{(n)}$ and evaluated in $r_Y^{(n+1)}$, and (a) is verified since $D_f(p_X\|\cdot)$ is a convex function in its second argument.

Since $\hat{W}[\hat{r}_Y^{(n)}] \geq W[\hat{r}_Y^{(n)}] \geq V[r_Y^*]$ we can rewrite

$$0 \leq \hat{W}[\hat{r}_Y^{(n)}] - V[r_Y^*] \leq G[\hat{r}_Y^{(n)}]$$

where

$$G[\hat{r}_Y^{(n)}] = \sum_{(x,y)\in\mathcal{X}^2} p_X Q^*_{Y|X} \log\left(\frac{\hat{Q}^{(n+1)}_{Y|X}}{\hat{Q}^{(n)}_{Y|X}}\right) + s_P\left[D_f\left(p_X\|r_Y^{(n+1)}\right) - T_{D_f(p\|\cdot),v^{(n)}}\left(r_Y^{(n+1)}\right)\right].$$

Summing over N terms we obtain

$$\begin{aligned}\sum_{n=1}^{N}\hat{W}[\hat{r}_Y^{(n)}] - V[r_Y^*] &\leq \sum_{(x,y)\in\mathcal{X}^2} p_X Q^*_{Y|X}\sum_{n=1}^{N}\log\left(\frac{\hat{Q}^{(n+1)}_{Y|X}}{\hat{Q}^{(n)}_{Y|X}}\right)\\ &\quad + s_P\sum_{n=1}^{N} D_f\left(p_X\|r_Y^{(n+1)}\right) - T_{D_f(p\|\cdot),v^{(n)}}\left(r_Y^{(n+1)}\right)\\ &\leq \sum_{(x,y)\in\mathcal{X}^2} p_X Q^*_{Y|X}\log\left(\frac{Q^*_{Y|X}}{\hat{Q}^{(1)}_{Y|X}}\right) + \sum_{n=1}^{N} o\left(\|r_Y^{(n+1)} - v^{(n)}\|\right)\\ &\leq \tilde{L}(r_Y^*, r_Y^{(0)})\end{aligned}$$

where $\tilde{L}(r_Y^*, r_Y^{(0)})$ is finite if the limit $\lim_{n\to\infty}\|r_Y^{(n+1)} - v^{(n)}\| = 0$ converges at least linearly. Thus, we can rewrite

$$0 \leq \lim_{N\to\infty}\sum_{n=1}^{N}\hat{W}[\hat{r}_Y^{(n)}] - V[r_Y^*] \leq \tilde{L}(r_Y^*, r_Y^{(0)})$$

proving the convergence as in Theorem 2.2.1. This completes the proof.

2.5.6 Proof of Theorem 2.3.1

We start by first introducing the following auxiliary lemma.

Lemma 2.5.1. *Let $s_D \geq 0$, $s_P \in [0, s_{P,\max})$ be given with $s = (s_D, s_P)$ and let $Q_{Y|X}$ be a transition matrix included in the set $\mathcal{L}_{(D,P)}$ defined as follows:*

$$\mathcal{L}_{(D,P)} = \{Q_{Y|X} : \mathbb{E}_{Q_{Y|X}}[d(x,y)] \leq D \wedge D_f(p_X\|q_Y) \leq P\}.$$

where $q_Y = \sum_{x\in\mathcal{X}} p_X Q_{Y|X}$. Furthermore, let $\Lambda_{s,v[q_Y]}$ be the set defined as

$$\Lambda_{s,v[q_Y]} \triangleq \{\gamma \in \mathbb{R}^{|\mathcal{X}|} : \forall x \in \mathcal{X}, \gamma(x) \geq 0 \wedge \forall y \in \mathcal{X}, \sum_{x\in\mathcal{X}} p_X(x)\gamma(x)A[v[q_Y]](x,y,s) \leq 1\}.$$

Then, $\forall\gamma \in \Lambda_{s,v[q_Y]}$, we obtain

$$R(D,P) \geq \sum_{x\in\mathcal{X}} p_X(x)\log(\gamma(x)) - s_D D - s_P\sum_{(x,y)\in\mathcal{X}^2} p_X(x)Q_{Y|X}(x,y)g(p_X(y), v[q_Y](y)).$$

Proof. Let $\gamma \in \Lambda_{s,v[q_Y]}$ and $Q_{Y|X} \in \mathcal{L}_{(D,P)}$, then:

$$
\begin{aligned}
&I\left(p_X, Q_{Y|X}\right) + \sum_{x\in\mathcal{X}} p_X(x)\log\left(\frac{1}{\gamma(x)}\right) + s_D D + s_P \sum_{(x,y)\in\mathcal{X}^2} p_X(x)Q_{Y|X}(x,y)g(p_X(y), v[q_Y](y)) \\
&\geq \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X} \log\left(\frac{Q_{Y|X}}{q_Y\gamma(x)A[v[q_Y]](x,y,s)}\right) \\
&\geq 1 - \sum_{(x,y)\in\mathcal{X}^2} p_X Q_{Y|X}\left(\frac{q_Y\gamma(x)A[v[q_Y]](x,y,s)}{Q_{Y|X}}\right) \\
&= 1 - \sum_{y\in\mathcal{X}} q_Y \sum_{x\in\mathcal{X}} p_X\gamma(x)A[v[q_Y]](x,y,s) \\
&\geq 1 - \sum_{y\in\mathcal{X}} q_Y = 0.
\end{aligned}
$$

The equality is ensured by the possibility of choosing $\gamma(x) \in \Lambda_{s,v[q_Y]}$ as

$$\gamma(x) = \frac{1}{\sum_{y\in\mathcal{X}} q_Y(y)A[v[q_Y]](x,y,s)}$$

that once substituted describes the optimization problem found in Corollary 2.2.1. This completes the proof of the lemma. □

Now we use Lemma 2.5.1 to prove Theorem 2.3.1. In particular, (2.3.2) can be derived from the following inequality:

$$
\begin{aligned}
R(D,P) \leq \quad & I\left(p_X, \hat{Q}_{Y|X}[\hat{r}_Y]\right) \\
&= \sum_{(x,y)\in\mathcal{X}^2} p_X(x)\hat{Q}_{Y|X}[\hat{r}_Y](y|x)\log\left(\frac{\hat{Q}_{Y|X}[\hat{r}_Y](y,x)}{\sum_{x\in\mathcal{X}} p_X(x)\hat{Q}_{Y|X}[\hat{r}_Y](y|x)}\right) \\
&= -s_D D - s_P P - \sum_{y\in\mathcal{X}} \hat{r}_Y(y)c(y)\log(c(y)) + s_P \sum_{y\in\mathcal{X}} q_Y[\hat{r}_Y](y)\frac{p_X(y)}{v[\hat{r}_Y](y)}\partial f\left(\frac{p_X(y)}{v[\hat{r}_Y](y)}\right) \\
&\quad - s_P\left\{\sum_{y\in\mathcal{X}} q_Y[\hat{r}_Y](y) f\left(\frac{p_X(y)}{v[\hat{r}_Y](y)}\right) - P\right\} - \sum_{x\in\mathcal{X}} p_X(x)\log\left(\sum_{y\in\mathcal{X}} \hat{r}_Y(y)A[v[\hat{r}_Y]](x,y,s)\right) \\
&= -s_D D - s_P P - \sum_{y\in\mathcal{X}} \hat{r}_Y(y)c(y)\log(c(y)) + \hat{W}[\hat{r}_Y].
\end{aligned}
$$

(2.3.1) is derived as an application of Lemma 2.5.1 by choosing $\gamma(x)$ as

$$\gamma(x) = \left(c_{\max}\sum_{y\in\mathcal{X}} \hat{r}_Y(y)A[v[\hat{r}_Y]](x,y,s)\right)^{-1}$$

which respects the assumption of the theorem. This completes the proof.

2.5.7 Proof of Theorem 2.3.2

The functional form of $J[r_Y](i,j)$ in the case of Theorem 2.2.1 is obtained from

$$\frac{\partial S[r_Y](i)}{\partial r_Y(j)} = r_Y(i)\sum_{x\in\mathcal{X}} p_X(x)\frac{\partial}{\partial r_Y(j)}\left(\frac{A[S[r_Y]](x,i,s)}{\sum_{k\in\mathcal{X}} r_Y(k)A[S[r_Y]](x,k,s)}\right) + c[r_Y, S[r_Y]](i)\delta_{i,j}$$

where

$$\begin{aligned}
\frac{\partial(\sum_{k\in\mathcal{X}} r_Y(k)A[S[r_Y]](x,k,s))}{\partial r_Y(j)} &= \sum_{k\in\mathcal{X}} r_Y(k)\frac{\partial A[S[r_Y]](x,k,s)}{\partial r_Y(j)} + A[S[r_Y]](x,j,s)\\
\frac{\partial A(x,i,s)}{\partial r_Y(j)} &= -s_P A[S[r_Y]](x,i,s)\frac{p_X(i)^2}{(S[r_Y(i)])^3}\partial f\left(\frac{p_X(i)}{S[r_Y(i)]}\right)\frac{\partial S[r_Y](i)}{\partial r_Y(j)}.
\end{aligned}$$

By noticing that

$$\frac{p_X(i)^2}{(S[r_Y](i))^3}\partial f\left(\frac{p_X(k)}{S[r_Y](i)}\right) = \frac{\partial^2}{\partial q(i)^2} D_f(p_X\|q)\Big|_{S[r_Y]}$$

and defining the matrices M and Γ as in (2.2.11) and (2.2.12), we can rewrite the entries of $J[r_Y^*]$ as

$$\begin{aligned}
J[r_Y^*](i,j) = {} & c[r_Y^*, r_Y^*](i)\left(\delta_{i,j} - \Gamma[r_Y^*, r_Y^*](i)J[r_Y^*](i,j)\right)\\
& + \sum_{k\in\mathcal{X}} \Gamma[r_Y^*, r_Y^*](k)M[r_Y^*, r_Y^*](i,k)J[r_Y^*](k,j) - M[r_Y^*, r_Y^*](i,j)
\end{aligned} \tag{2.5.10}$$

where $\delta_{i,j}$ is the Kronecker delta. As a result, we can express (2.5.10) in matrix form as follows

$$\begin{aligned}
J[r_Y^*] &= C[r_Y^*, r_Y^*](I - \Gamma[r_Y^*, r_Y^*]J[r_Y^*]) - M[r_Y^*, r_Y^*] + M[r_Y^*, r_Y^*]\Gamma[r_Y^*, r_Y^*]J[r_Y^*]\\
&= (C[r_Y^*, r_Y^*] - M[r_Y^*, r_Y^*])(I - \Gamma J[r_Y^*])
\end{aligned}$$

where $C[\cdot,\cdot]$ is defined in (2.2.13). Finally, we obtain (2.3.4) noticing that $C[r_Y^*, r_Y^*] = I$ due to the optimality conditions found in Lemma 2.2.2, thus concluding the proof.

2.5.8 Proof of Lemma 2.3.1

Let matrices Φ and Q be defined as

$$\begin{aligned}
\Phi &\triangleq \left[\sqrt{p_X(x)}\frac{A[r_Y^*](x,i,s)}{\sum_{k\in\mathcal{X}} r_Y(k)A[r_Y^*](x,k,s)}\right]_{(i,x)\in\mathcal{X}^2}\\
Q &\triangleq \operatorname{diag}\left[r_Y^*(i)\right]_{i\in\mathcal{X}}.
\end{aligned}$$

Then, the following identity can be verified

$$Q^{\frac{1}{2}}M^*Q^{-\frac{1}{2}} = Q^{\frac{1}{2}}\Phi\Phi^T Q^{\frac{1}{2}} = (Q^{\frac{1}{2}}\Phi)(Q^{\frac{1}{2}}\Phi)^T$$

where $Q^{\frac{1}{2}}M^*Q^{-\frac{1}{2}}$ is necessarily symmetric and at least semi-positive definite. To guarantee positive definiteness of $Q^{\frac{1}{2}}M^*Q^{-\frac{1}{2}}$, and thus the fact that the eigenvalues of M are strictly positive, we need to impose conditions on the full rank of Φ. To address them, we can factorize Φ into the product $\Phi = UDV$, where

$$\begin{aligned} D &= \left[e^{-s_D d(i,j)}\right]_{(i,j)\in\mathcal{X}\times\mathcal{X}} \\ U &= \operatorname{diag}\left[e^{-s_P g(p_X, r_Y^*, i)}\right]_{i\in\mathcal{X}} \\ V &= \operatorname{diag}\left[\frac{\sqrt{p_X(x)}}{\sum_{k\in\mathcal{X}} r_Y(k) A[r_Y^*](x,k,s)}\right]_{x\in\mathcal{X}}. \end{aligned}$$

Since both U and V are positive definite matrices, it is easy to verify that Φ is a full-rank matrix if and only if D is full rank too. This completes the proof.

2.5.9 Proof of Lemma 2.3.2

We can verify that the lemma is the result of the Gershgorin Circle Theorem [29] applied to the columns of M^*. Noticing that all entries $M^*(i,j)$ are strictly positive, the disk radius $R(j)$ for column j is:

$$R(j) + M^*(i,i) = \sum_{i\in\mathcal{X}} M^*(i,j)$$

where

$$\begin{aligned} \sum_{i\in\mathcal{X}} M^*(i,j) &= \sum_{i\in\mathcal{X}} r_Y^*(i) \sum_{x\in\mathcal{X}} p_X(x) \frac{A[r_Y^*](x,i,s) A[r_Y^*](x,j,s)}{(\sum_{k\in\mathcal{X}} r_Y^*(k) A[r_Y^*](x,k,s))^2} \\ &= \sum_{x\in\mathcal{X}} p_X(x) \frac{A[r_Y^*](x,j,s)}{\left(\sum_{k\in\mathcal{X}} r_Y^*(k) A[r_Y^*](x,k,s)\right)^2} \left(\sum_{i\in\mathcal{X}} r_Y^*(i) A[r_Y^*](x,i,s)\right) \\ &= c[r_Y^*, r_Y^*](j) = 1. \end{aligned}$$

Thus the eigenvalues of M^* are each in at least one of the disks $I_i = \{z \in \mathbb{R} : |z - M^*(i,i)| \leq 1 - M^*(i,i)\}$, which are all contained in the disk $I = \{z \in \mathbb{R} : |z| \leq 1\}$. This completes the proof.

2.5.10 Proof of Theorem 2.3.3

Using Lemmas 2.3.1 and 2.3.2, the following inequalities hold

$$\begin{aligned} 0 < \lambda[M^*] \leq 1 &\implies 0 \leq \lambda[I - M^*] < 1 \\ 0 \leq \lambda[I - M^*] < 1 &\overset{(a)}{\implies} 1 \leq \lambda[I + (I - M^*)\Gamma] \\ 1 \leq \lambda[I + (I - M^*)\Gamma] &\implies 0 < \lambda[(I + (I - M^*)\Gamma^*)^{-1}] \leq 1 \end{aligned}$$

where (a) is due to Γ being a positive definite matrix. Using the previous inequalities, we can rewrite (2.3.4) as follows:

$$J(r_Y^*) = (I + (I - M^*)\Gamma^*)^{-1}(I - M^*).$$

Define $\theta_{\text{sup}} \triangleq \lambda_{\max}[I - M^*] \cdot \lambda_{\max}[I + (I - M^*)\Gamma^*)^{-1}]$. Then, we can show that $0 \leq \lambda_{J[r_Y^*]} \leq \theta_{\text{sup}} < 1$ is always verified. The second part of the theorem follows directly from [54, Theorem 5] hence we omit it. This completes the proof.

2.5.11 Proof of Theorem 2.3.4

Since $R(D, P)$ is a non-increasing convex function in D_s and P_s, we can derive:

$$\frac{\partial^2 R(D,P)}{\partial P_s^2} = -\frac{\partial s_P}{\partial P_s} \geq 0$$

meaning that more constrained values of P_s are associated with larger s_P. Thus, for a given s_D, let $s_{P,\max}$ be the value of the Lagrangian s_P associated with the constraint $P = 0$. Then the solution r_Y^* is necessarily unique and must be $r_Y^* = p_X$. Then, due to the properties of the Jacobian $J(r_Y^*)$,

$$(I - \Gamma J(r_Y^*)) \geq 0 \implies s_{P,\max} \leq \frac{1}{\theta_{\max} f''(0)}.$$

In order to guarantee $0 \leq \{\theta_{a,i}\}_{i \in \mathcal{X}} < 1$, it is sufficient to have

$$(I - \Gamma) \geq 0 \implies s_{P,\max} \leq \frac{1}{f''(0)}.$$

Since in the non-degenerate case we have $0 < \theta_{\max} < 1$, we can construct ϵ' considering ;

$$\epsilon' = \min\left\{ s_{P,\max}, \frac{1}{f''(0)} \left(\frac{1}{\theta_{\max}} - 1 \right) \right\} > 0$$

and define the set $I_{s_P}^{\epsilon'}$ where the exponential convergence of the approximate algorithm is guaranteed. This concludes the proof.

2.5.12 Proof of Lemma 2.4.2

We remind that $f = |x-1|$ is the function associated with the TV distance $TV(\cdot\|\cdot)$. The first statement can be proved by first establishing the uniform convergence of $f_n \to f$ as $n \to \infty$

$$\begin{aligned} \sup_{x \in \mathbb{R}} |f_n(x) - f(x)| &= \sup_{x \in \mathbb{R}} \left| \frac{2}{\pi}(x-1)\arctan(n(x-1)) - |x-1| \right| \\ &= \sup_{x \in \mathbb{R}_0^+} \left| \frac{2}{\pi} x \arctan(nx) - x \right| \end{aligned}$$

$$
\begin{aligned}
&= \sup_{x \in \mathbb{R}_0^+} \frac{2}{\pi} x \left(\frac{\pi}{2} - \arctan(nx) \right) \\
&= \sup_{x \in \mathbb{R}_0^+} \frac{2}{\pi} x \arctan\left(\frac{1}{nx} \right) \\
&\leq \sup_{x \in \mathbb{R}_0^+} \frac{2}{n\pi} = \frac{2}{n\pi}
\end{aligned}
$$

meaning that $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |f_n(x) - f(x)| = 0$. A direct consequence of the above is that $D_{f_n} \to TV$ uniformly in the limit of $n \to \infty$ since, for any $p, q \in \mathcal{P}(\mathcal{X})$,

$$
\begin{aligned}
\lim_{n\to\infty} |D_{f_n}(p\|q) - TV(p\|q)| &= \lim_{n\to\infty} \left| \sum_{x\in\mathcal{X}} q(x) \left(f_n\left(\frac{p(x)}{q(x)} \right) - f\left(\frac{p(x)}{q(x)} \right) \right) \right| \\
&\leq \sum_{x\in\mathcal{X}} q(x) \lim_{n\to\infty} \left| f_n\left(\frac{p(x)}{q(x)} \right) - f\left(\frac{p(x)}{q(x)} \right) \right| = 0.
\end{aligned}
$$

Instead, the inequality $f_n(x) \leq f(x), \forall x \in \mathbb{R}$, implies that for all $n \in \mathbb{N}$ and $\forall p, q \in \mathcal{P}(\mathcal{X})$, the inequality $D_{f_n}(p\|q) \leq TV(p\|q)$ holds. This concludes the proof.

Part II:

RDP Trade-off for Continuous Information Sources

Chapter 3

Vector Gaussian RDPF

With this chapter, we shift our attention on the RDP framework from discrete to continuous sources, focusing on Gaussian sources. The objective is twofold. First, we aim to derive analytical expressions of the RDPF under MSE distortion when the perception constraint belongs to certain well-known and widely-used divergence, that is, the KL divergence [43], the geometric Jensen-Shannon divergence [44], the squared Hellinger distance [43], and the squared Wasserstein-2 distance [46] for scalar-valued Gaussian sources with jointly Gaussian reconstruction. Although the assumption of jointly Gaussian reconstruction generally induces an upper bound to the RDPF, it can be shown that depending on the specific perception metric, the bound can become exact, and this observation extends to multivariate Gaussian sources with jointly Gaussian reconstructions. The second and most important goal is the construction of a generic algorithmic approach for the optimal computation of the RDPF of a multivariate Gaussian source with jointly Gaussian reconstruction, under convex and tensorizable distortion and perception metrics, when an analytical solution of the associated scalar Gaussian RDPF is available. To summarize, in this chapter we derive the following new results:

- In Section 3.1, we characterize closed-forms expressions of the scalar Gaussian RDPF for direct or reverse KL divergence, the geometric Jensen-Shannon divergence, the squared Hellinger distance, and the squared Wasserstein-2 distance perception constraints.

- In Section 3.2.1, we prove that, under tensorizable distortion and divergence constraints, the optimal solution of the multivariate Gaussian RDPF subject to convex and tensorizable distortion and perception metrics, can be found on the space of the eigenvectors of the source covariance matrix (to obtain this result we make use of Lemma 3.4.1 in Appendix 3.4). In other words, under the assumption of jointly Gaussian reconstruction, the problem achieves the optimal solution when the involved covariance matrices commute by pairs [29, Section 0.7.7]. The resulting optimization problem can be solved optimally using an alternating minimization approach, by means of the *block nonlinear Gauss–Seidel method* [50], for which we

also develop its algorithmic embodiment (see Alg. 3). For the specific algorithm, we show convergence (Theorem 3.2.1) and provide an upper bound on the worst-case convergence rate (Theorem 3.2.2).

- In Section 3.2.2, we provide, as an application example, the implementation of Algorithm 3 for the MSE distortion and squared Wasserstein-2 divergence constraints. Although only the solution of one of the subproblems of Algorithm 3 is available in closed-form (Theorems 3.2.3), the complementary subproblem can be solved optimally through numerical methods (Theorem 3.2.4). We note that similar implementations can be derived for the other divergence constraints that are used in this work via the results summarized in Table 3.1.

- In Section 3.2.3, we leverage the analytical results obtained from the alternating minimization approach that led to Algorithm 3, to characterize the closed-form solution of the multivariate Gaussian RDPF in the regime of *perfect realism* (Corollary 3.2.2). Specifically, this result provides the optimal stagewise distortion allocation, i.e., the distortion introduced on each dimension of the Gaussian reconstruction, according to which can be interpreted as an *adaptive water-level* (see Figure 3.4). In fact, unlike the reverse water-filling solution in the classical RDF problem, the obtained solution presents dependency on the stagewise second-order moment of the source.

We highlight that for the KL divergence and the squared Wasserstein-2 distance in the scalar case, we solve the problem differently from [35, 57] in a way that allows to also find the optimal linear encoder and decoder that achieve the specific closed-form expression (i.e., the optimal forward test-channel realization). When it comes to the vector Gaussian case, we note that the reverse water-filling parametric solutions in [57] are only applicable to specific problems. In contrast, our alternating minimization approach and the corresponding algorithmic embodiment obtained here can be applied to any Gaussian RDPF problem with MSE distortion, as long as the scalar Gaussian RDPF admits a characterization. In addition, our approach provides, similar to the scalar case, the optimal linear encoder and decoder pair that achieves the Gaussian RDPF as a result of the optimization procedure.

3.1 RDPF for Scalar-valued Gaussian sources

In this section, we derive closed-form expressions of the RDPF for a scalar-valued Gaussian source assuming jointly Gaussian reconstruction, under MSE distortion constraint and various perception constraints. The considered class of perception constraints consists of direct or reverse D_{KL}, $\mathrm{D}_{\mathrm{GJS}}$, H^2, and W_2^2 perception constraints[1], respectively.

[1]For the squared Wasserstein-2 divergence, the same closed-form solution is derived using a different methodology in [35, Theorem 1].

Moreover, on top of the closed-form expressions of Gaussian RDPF for the previous class of perception constraints, our methodology completely specifies the design variables of the *optimal linear realization* (i.e., the optimal selection of encoder and decoder policies) that achieve the obtained Gaussian RDPF bounds for each perception constraint. In what follows, we characterize the RDPF (1.4.1) for jointly Gaussian random variables and MSE distortion constraint.

Problem 3.1.1. *Given a Gaussian source $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $\sigma_X^2 > 0$, assume that the reconstructed message Y is chosen such that the joint tuple (X, Y) is also Gaussian. Then, the reconstructed message admits a linear (forward) realization of the form $Y = aX + W$, where $a \in \mathbb{R}$, $W \sim \mathcal{N}(\mu_W, \sigma_W^2)$, $W \perp X$ and $\sigma_W^2 \geq 0$, such that $\mu_Y = a\mu_X + \mu_W$ and $\sigma_Y^2 = a^2\sigma_X^2 + \sigma_W^2$. By considering the case where $\mu_X = \mu_Y$ (by setting $\mu_W = (1-a)\mu_X$), and $\mathbb{E}\left[d(X,Y)\right] \equiv \mathbb{E}\left[(X-Y)^2\right]$, we can cast* (1.4.1)-(1.4.3) *as follows:*

$$\begin{aligned} R(D,P) \leq R^G(D,P) = \min_{a\in\mathbb{R},\sigma_W^2\geq 0} \frac{1}{2}\log\left(1 + a^2\frac{\sigma_X^2}{\sigma_W^2}\right) \\ s.t. \quad (1-a)^2\sigma_X^2 + \sigma_W^2 \leq D, \qquad D(p_X\|p_Y) \leq P \end{aligned} \tag{3.1.1}$$

where $R^G(D,P)$ indicates the Gaussian RDPF (assuming jointly Gaussian reconstruction) and $D(\cdot\|\cdot)$ depends on the specific choice of perception constraint.

We point out the following technical remarks on Problem 3.1.1.

Remark 3.1.1. *(On Problem 3.1.1) If in Problem 3.1.1 the choice of the perception constraint is restricted to $\mathrm{W}_2^2(\cdot,\cdot)$ or the reverse $\mathrm{D}_{\mathrm{KL}}(\cdot\|\cdot)$, it can be shown that their respective minimizing distribution is itself Gaussian, (see e.g., [58] and [35, Appendix A]). This guarantees that* (3.1.1) *in these two cases provides an exact characterization of the Gaussian RDPF, i.e., $R(D,P) = R^G(D,P)$. Alas, the same property does not hold, in general, for many divergences such as the cases of direct-$\mathrm{D}_{\mathrm{KL}}(\cdot\|\cdot)$, $\mathrm{D}_{\mathrm{GJS}}(\cdot\|\cdot)$ or $\mathrm{H}^2(\cdot,\cdot)$. In such cases, the characterization in* (3.1.1) *serves as an upper bound to the optimal solution.*

Remark 3.1.2. *(Mean Matching) The assumption $\mu_X = \mu_Y$ does not hinder the generality of the derivation, which can be proved to be optimal. In fact, matching the mean of the source X and the reconstruction Y does not affect their mutual information $I(X,Y)$ when minimizing the MSE distortion metric. Moreover, all the considered divergence functions $D(\cdot\|\cdot)$ are minimized by matching the first moment of the involved distributions (for more details, see Lemma 3.5.1 in Appendix 3.5).*

In the following theorem, we compute in closed-form the characterization of the Gaussian RDPF defined in Problem 3.1.1, when $D(\cdot\|\cdot)$ corresponds to the direct $\mathrm{D}_{\mathrm{KL}}(p_X\|p_Y)$.

Theorem 3.1.1 (RDPF under direct $\mathrm{D}_{\mathrm{KL}}(\cdot\|\cdot)$)**.** *Consider the characterization in Problem 3.1.1 with $D(\cdot\|\cdot) \equiv D_{KL}(\cdot\|\cdot)$. Then the solution of the Gaussian RDPF in* (3.1.1) *is*

obtained analytically, as follows:

$$R^G(D,P)=\begin{cases}\max\left\{\frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right),0\right\} & \text{if } (D,P)\in\mathcal{S}^c\\ \frac{1}{2}\log\left(1+\frac{\sigma_X^2\left(1-\frac{D}{\sigma_X^2}-g(P)\right)^2}{4D-\sigma_X^2\left(1+\frac{D}{\sigma_X^2}+g(P)\right)^2}\right) & \text{if } (D,P)\in\mathcal{S}\end{cases} \tag{3.1.2}$$

where

$$g(P)=\frac{1}{\mathcal{W}_{-1}\left(-e^{-(2P+1)}\right)} \tag{3.1.3}$$

$$\mathcal{S}=\left\{(D,P):\frac{|\sigma_X^2-D|}{\sigma_X^2}\leq -g(P)\right\}. \tag{3.1.4}$$

Moreover, the solution in (3.1.2) *is achieved by a linear realization* $Y=aX+W$ *such that the design variables* (a,σ_W^2) *are obtained in closed-form as follows:*

$$a=\begin{cases}\max\left\{1-\frac{D}{\sigma_X^2},0\right\} & \text{if } (D,P)\in\mathcal{S}^c\\ \frac{1}{2}\left(1-\frac{D}{\sigma_X^2}-g(P)\right) & \text{if } (D,P)\in\mathcal{S}\end{cases} \tag{3.1.5}$$

$$\sigma_W^2=\min\{D,(1-g(P))\sigma_X^2\}-(1-a)^2\sigma_X^2. \tag{3.1.6}$$

Proof. See Appendix 3.6.1. □

Next, we use Theorem 3.1.1 to obtain a similar result when the perception constraint of the upper bound characterization in (3.1.1) is the reverse KL divergence, which corresponds to a case where the upper bound is an exact solution of the $R(D,P)$ (see Remark 3.1.1). Our derivation extends [57] providing, on top of a closed-form solution to the problem, the design of the optimal (forward) realization that achieves the optimal solution. This is reported in the following corollary.

Corollary 3.1.1 (RDPF under reverse $D_{KL}(\cdot\|\cdot)$)**.** *Consider the characterization in Problem 3.1.1 with* $D(\cdot\|\cdot)\equiv D_{KL}(p_Y\|p_X)$*. Then, the exact solution of* (3.1.1) *corresponds to* $R(D,P)=R^G(D,P)=$ (3.1.2)*, where in place of* (3.1.3) *we now have* $g(P)=\mathcal{W}_0\left(-e^{-(2P+1)}\right)$*. Moreover, the optimal realization* $Y=aX+W$ *that achieves* $R(D,P)$ *corresponds to the choice of the design variables* (a,σ_W^2) *given by* (3.1.5) *and* (3.1.6)*, with the updated form of* $g(P)$ *given above.*

Proof. The proof follows almost verbatim to the proof of Theorem 3.1.1 hence we omit it. □

In the next lemma, we provide an alternative derivation to a closed-form result first obtained by Zhang *et al.* in [35, Theorem 1]. This result corresponds to the exact computation of $R(D,P)$ in (3.1.1) because we consider $D(\cdot\|\cdot)\equiv \mathrm{W}_2^2(p_X,p_Y)$. As in

Corollary 3.1.1, we note that our methodology reveals on top of the closed-form solution to this problem, the design of the optimal (forward) realization that achieves the optimal solution. These results are reported in the following lemma.

Lemma 3.1.1. *Consider the characterization in Problem 3.1.1 with $D(\cdot\|\cdot) \equiv W_2^2(\cdot,\cdot)$. Then the optimal solution of R(D,P) in* (3.1.1) *is obtained analytically as follows:*

$$R(D,P) = R^G(D,P) = \begin{cases} \max\left\{\frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right), 0\right\} & \text{if } (D,P) \in \mathcal{S}^c \\ \frac{1}{2}\log\left(1 + \frac{\left[\sigma_X^2 + (\sigma_X - \sqrt{P})^2 - D\right]^2}{(D-P)\left[(2\sigma_X - \sqrt{P})^2 - D\right]}\right) & \text{if } (D,P) \in \mathcal{S} \end{cases} \tag{3.1.7}$$

where $\mathcal{S} = \left\{(D,P) : \sqrt{P} \leq \sigma_X - \sqrt{|\sigma_X^2 - D|}\right\}$. *Moreover, the solution in* (3.1.2) *is achieved by a linear realization $Y = aX + W$ such that the design variables (a, σ_W^2) are obtained in closed-form as follows:*

$$a = \begin{cases} \max\left\{1 - \frac{D}{\sigma_X^2}, 0\right\} & \text{if } (D,P) \in \mathcal{S}^c \\ \frac{1}{2}\frac{\sigma_X^2 + \left(\sigma_X - \sqrt{P}\right)^2 - D}{\sigma_X^2} & \text{if } (D,P) \in \mathcal{S} \end{cases} \tag{3.1.8}$$

$$\sigma_W^2 = \min\{D, \sigma_X^2 + (\sigma_X - \sqrt{P})^2\} - (1-a)^2\sigma_X^2.$$

Proof. See Appendix 3.6.2. □

In the next theorem, we derive the closed-form solution of (3.1.1) when the perception constraint is the $\mathrm{D_{GJS}}$ divergence.

Theorem 3.1.2 (RDPF under $\mathrm{D_{GJS}}(\cdot\|\cdot)$)**.** *Consider the characterization in Problem 3.1.1 with $D(\cdot\|\cdot) \equiv D_{GJS}(\cdot\|\cdot)$. Then, the solution of $R^G(D,P)$ in* (3.1.1) *corresponds to the following analytical expression:*

$$R^G(D,P) = \begin{cases} \max\left\{\frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right), 0\right\} & \text{if } (D,P) \in \mathcal{S}^c \\ \frac{1}{2}\log\left(1 + \frac{\sigma_X^2\upsilon^2}{D - \sigma_X^2(1+\upsilon)^2}\right) & \text{if } (D,P) \in \mathcal{S} \end{cases} \tag{3.1.9}$$

where $\upsilon = \frac{D}{2\sigma_X^2} - \frac{g(P)}{4} + \frac{\sqrt{g(P)(g(P)-4)}}{4}$ *with* $g(P) = -2\mathcal{W}_{-1}\left(-2e^{-(4P+2)}\right)$, *and*

$$\mathcal{S}^c = \left\{(D,P) : f_L(P) \leq |\sigma_X^2 - D| \leq f_U(P)\right\}$$
$$f_L(P) = \frac{1}{2}\sigma_X^2\left(-\sqrt{g(P)(g(P)-4)} - (2 - g(P))\right)$$
$$f_U(P) = \frac{1}{2}\sigma_X^2\left(\sqrt{g(P)(g(P)-4)} - (2 - g(P))\right).$$

Moreover, the solution in (3.1.9) *is achieved by a linear realization $Y = aX + W$ such*

that the design variables (a, σ_W^2) *are obtained in closed-form as follows:*

$$a = \begin{cases} \max\left\{1 - \frac{D}{\sigma_X^2}, 0\right\} & \text{if } (D,P) \in \mathcal{S}^c \\ \frac{1}{2}\left(1 - \frac{D - f_L(P)}{\sigma_X^2}\right) & \text{if } (D,P) \in \mathcal{S} \end{cases}, \quad \sigma_W^2 = \min\{D, \sigma_X^2 + f_L(P)\} - (1-a)^2 \sigma_X^2. \tag{3.1.10}$$

Proof. See Appendix 3.6.3. □

Our last result in this section provides the closed-form solution of $R^G(D,P)$ in (3.1.1) when the perception constraint is the H^2 distance.

Theorem 3.1.3 (RDPF under $\mathrm{H}^2(\cdot,\cdot)$)**.** *Consider the characterization in Problem 3.1.1 with* $D(\cdot\|\cdot) \equiv \mathrm{H}^2(\cdot,\cdot)$*. Then, the solution of* $R^G(D,P)$ *in* (3.1.1) *corresponds to the following analytical expression:*

$$R^G(D,P) = \begin{cases} \max\left\{\frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right), 0\right\} & \text{if } (D,P) \in \mathcal{S}^c \\ \frac{1}{2}\log\left(1 + \frac{\sigma_X^2\left(g(P) - \frac{D}{2\sigma_X^2}\right)^2}{D - \sigma_X^2\left(1 - g(P) + \frac{D}{2\sigma_X^2}\right)^2}\right) & \text{if } (D,P) \in \mathcal{S} \end{cases} \tag{3.1.11}$$

where

$$g(P) = \frac{1 - \sqrt{1 - (1 - \frac{P}{2})^4}}{(1 - \frac{P}{2})^4}, \qquad \mathcal{S} = \left\{(D,P) : 1 - g(P) \le \frac{D}{2\sigma_X^2} \le g(P)\right\}.$$

Moreover, the solution in (3.1.11) *is achieved by a linear realization* $Y = aX + W$ *such that the design variables* (a, σ_W^2) *are obtained in closed-form as follows:*

$$a = \begin{cases} \max\left\{1 - \frac{D}{\sigma_X^2}, 0\right\} & \text{if } (D,P) \in \mathcal{S}^c \\ g(P) - \frac{D}{2\sigma_X^2} & \text{if } (D,P) \in \mathcal{S} \end{cases}, \quad \sigma_W^2 = \min\{D, 2\sigma_X^2 g(P)\} - (1-a)^2 \sigma_X^2. \tag{3.1.12}$$

Proof. See Appendix 3.6.4. □

3.2 RDPF for Vector-Valued Gaussian sources

In this section, our goal is twofold. First, we derive (under certain conditions) a generic alternating minimization approach and its algorithmic embodiment, which allows the computation of any RDPF for multidimensional Gaussian sources with jointly Gaussian reconstruction. Second, we apply this algorithm when having an MSE distortion constraint and the class of perception constraints studied in Section 3.1, namely $\mathrm{D_{KL}}$, $\mathrm{D_{GJS}}$, H^2, W_2^2.

3.2.1 A Generic Alternating Minimization Approach

We start with the generalization of Problem 3.1.1 to vector-valued Gaussian sources. In contrast to Problem 3.1.1 we do not specify the type of fidelity constraints that will be utilized. This is because we aim to first provide a general approach along with its algorithmic embodiment, able to tackle *any* vector-valued jointly Gaussian problems for RDPF with jointly Gaussian reconstruction.

Problem 3.2.1. *Given a Gaussian source $X \sim \mathcal{N}(\mu_X, \Sigma_X)$, $\Sigma_X \succ 0$, assume that the reconstructed random vector $Y \in \mathbb{R}^N$ is chosen such that the joint tuple (X, Y) is jointly Gaussian. Then, the reconstructed message admits a linear (forward) realization of the form $Y = AX + W$, where $A \in \mathbb{R}^{N\times N}$, $W \sim \mathcal{N}(\mu_W, \Sigma_W)$, $W \perp X$ and $\Sigma_W \succeq 0$, such that $\mu_Y = A\mu_X + \mu_W$ and $\Sigma_Y = A\Sigma_X A^T + \Sigma_W$. Moreover, we can cast* (5.1.1)-(1.4.3) *as follows:*

$$\begin{aligned} R(D,P) \leq R^G(D,P) = \min_{A\in\mathbb{R}^{N\times N},\Sigma_W\succeq 0} \frac{1}{2}\log\left(\frac{|A\Sigma_X A^T+\Sigma_W|}{|\Sigma_W|}\right) \\ s.t. \quad \mathbb{E}\left[d(X,Y)\right] \leq D, \quad D(p_X\|p_Y) \leq P \end{aligned} \tag{3.2.1}$$

where the specific form of the fidelity constraints $\mathbb{E}\left[d(\cdot,\cdot)\right]$ and $D(\cdot\|\cdot)$ are not yet specified.

Assuming that also the fidelity constraints $\mathbb{E}\left[d(\cdot,\cdot)\right]$ and $D(\cdot\|\cdot)$ are tensorizable, i.e.,

$$\mathbb{E}\left[d(X,Y)\right] \geq \sum_{i=1}^{N} g\left(\mathbb{E}\left[d(X_i,Y_i)\right]\right) \qquad D(p_X\|p_Y) \geq \sum_{i=1}^{N} h\left(D(p_{X_i}\|p_{Y_i})\right)$$

with $g(\cdot)$ and $r(\cdot)$ being convex functions, and X_i (resp. Y_i) being the projection of X (resp. Y) onto the i^{th} eigenvector of Σ_X (resp. Σ_Y), ordered according to $\lambda^{\downarrow}[\Sigma_X]$ (resp. $\lambda^{\downarrow}[\Sigma_Y]$). Then, applying [Appendix 3.4, Lemma 3.4.1] in (3.2.1) leads to the following lower bound:

$$\begin{aligned} R^G(D,P) \overset{(\star)}{\geq} \min_{\lambda_{A,i},\lambda_{\Sigma_W,i}} \sum_{i=1}^{N} \frac{1}{2}\log\left(1+\frac{{\lambda_{A,i}}^2\lambda_{X,i}}{\lambda_{W,i}}\right) \\ \text{s.t.} \quad \sum_{i=1}^{N} g\left(\mathbb{E}\left[d(X_i,Y_i)\right]\right) \leq D, \qquad \sum_{i=1}^{N} h\left(D(p_{X_i}\|p_{Y_i})\right) \leq P \end{aligned} \tag{3.2.2}$$

where $(\star)$ holds with equality if the triplet (A, Σ_W, Σ_X) commute by pairs [29, Section 0.7.7]. In fact, for jointly Gaussian random vectors (X, Y), the sufficient condition that achieves the lower bound in (3.2.2) can easily be realized, since the matrices (A, Σ_W) are design variables and can be chosen such that they have the *same eigenvectors* as Σ_X [2], hence one can replace inequality with equality without loss of generality. We note that the inequality in (3.2.2) does not hold with equality, in general, beyond i. i. d. random vectors.

[2] For details see, e.g., [59, Proposition 1].

Proposed alternating minimization method To solve (3.2.2), we first introduce the (vector) optimization variables $\mathbf{D} = [D_i]_{i\in 1:N}$ and $\mathbf{P} = [P_i]_{i\in 1:N}$ such that

$$D_i = \mathbb{E}\left[d(X_i, Y_i)\right], \quad P_i = D(p_{X_i}\|p_{Y_i}) \quad \forall i \in 1:N.$$

Once the slack variables above are substituted in (3.2.2), we yield:

$$R^G(D,P) = \min_{\mathbf{D},\mathbf{P}} \sum_{i=1}^{N} R_i^G(D_i,P_i) \quad \text{s.t.} \quad \sum_{i=1}^{N} g(D_i) \leq D, \qquad \sum_{i=1}^{N} r(P_i) \leq P \tag{3.2.3}$$

where $R_i^G(D_i,P_i)$ corresponds to the stagewise RDPF given by

$$R_i^G(D_i,P_i) = \min_{\lambda_{A,i},\lambda_{\Sigma_W,i}} \frac{1}{2}\log\left(1+\frac{{\lambda_{A,i}}^2\lambda_{X,i}}{\lambda_{W,i}}\right) \text{ s.t. } D_i = \mathbb{E}\left[d(X_i,Y_i)\right],\ P_i = D(p_{X_i}\|p_{Y_i}). \tag{3.2.4}$$

Note that in (3.2.3) we have three distinct "rate region" cases, which combined, result in the complete computation of the Gaussian RDPF. Particularly, we may have only the distortion constraint to be active (hereinafter referred to as **Case I**), only the perception constraint to be active (hereinafter referred to as **Case II**), or both constraints to be active (hereinafter referred to as **Case III**). We note that only **Case III** is interesting as the computation of the other two cases follows from the computation of that case.

To find the optimal pair $(\mathbf{D}^*,\mathbf{P}^*)$ in (3.2.3) we resort to an application of an alternating minimization technique. Specifically, we define the following two subproblems of (3.2.3):

- For fixed $\mathbf{P}$, (3.2.3) simplifies to

$$\min_{\mathbf{D}} \sum_{i=1}^{N} R_i^G(D_i,P_i) \quad \text{s.t.} \quad \sum_{i=1}^{N} g(D_i) \leq D. \tag{3.2.5}$$

- For fixed $\mathbf{D}$, (3.2.3) simplifies to

$$\min_{\mathbf{P}} \sum_{i=1}^{N} R_i^G(D_i,P_i) \quad \text{s.t.} \quad \sum_{i=1}^{N} r(P_i) \leq P. \tag{3.2.6}$$

It should be noted that solving the optimization problems of (3.2.5) and (3.2.6) is of primary interest because the solution of these two problems forms the basis of an alternating minimization scheme that can optimally solve (3.2.3). To make this point clear, next, we state and prove the convergence to an optimal point of an alternating minimization approach that relies on the solution of (3.2.5) and (3.2.6). This is often encountered in the literature by the name *block nonlinear Gauss–Seidel method* [50].

Theorem 3.2.1. *(Convergence) Let the optimization problem* (3.2.3) *be defined for finite distortion and perception levels* (D,P)*. Let* $(\mathbf{D}^{(0)},\mathbf{P}^{(0)})$ *be an initial point and*

let the sequence $\{(\mathbf{D}^{(n)}, \mathbf{P}^{(n)}) : n = 1, 2, \ldots\}$ *be the sequence obtained by the alternating optimization of problems* (3.2.5) *and* (3.2.6)*. Then the sequence has a limit* $\lim_{n\to\infty}(\mathbf{D}^{(n)}, \mathbf{P}^{(n)}) = (\mathbf{D}^*, \mathbf{P}^*)$ *and the limit is an optimal solution of* (3.2.3)*.*

Proof. See Appendix 3.6.5. □

Subproblems (3.2.5) and (3.2.6), despite being more manageable than the original problem (3.2.3), are still constrained optimization problems whose solutions may not be easily obtained. Therefore, to further simplify the optimization task, we define the unconstrained optimization problem associated with (3.2.3). Let $s = (s_D, s_P)$, with $s_D > 0$ and $s_P > 0$, be the vector of Lagrangian multipliers associated with the distortion (i.e., s_D) and perception (i.e., s_P) constraints, respectively. Then, the Lagrangian functional $L_{R^G}(s)$ associated with (3.2.3) is defined as:

$$\min_{\mathbf{D},\mathbf{P}} L_{R^G}(\mathbf{D}, \mathbf{P}, s) \triangleq \min_{\mathbf{D},\mathbf{P}} \left[\sum_{i=1}^{N} R_i^G(D_i, P_i) + s_D \sum_{i=1}^{N} g(D_i) + s_P \sum_{i=1}^{N} r(P_i) \right]. \tag{3.2.7}$$

Similarly to the constrained case, the optimal pair $(\mathbf{D}^*, \mathbf{P}^*)$ in (3.2.7) can be characterized through an alternate minimization scheme. Hence, the associated subproblems are as follows:

- For fixed $\mathbf{P}$,

$$\min_{\mathbf{D}} \sum_{i=1}^{N} R_i^G(D_i, P_i) + s_D \sum_{i=1}^{N} g(D_i). \tag{3.2.8}$$

- For fixed $\mathbf{D}$,

$$\min_{\mathbf{P}} \sum_{i=1}^{N} R_i^G(D_i, P_i) + s_P \sum_{i=1}^{N} r(P_i). \tag{3.2.9}$$

Assume the Lagrangian multiplier vector s is given and let $\mathbf{D}_s^*$ and $\mathbf{P}_s^*$ be the optimal solutions obtained from the Gauss-Seidel method employing subproblems (3.2.8) and (3.2.9), respectively. Furthermore, let $D_s = \sum_{i=1}^{N} g(D_{s,i}^*)$ and $P_s = \sum_{i=1}^{N} r(P_{s,i}^*)$. Then, leveraging Lagrangian duality [60], we can easily compute $R^G(D_s, P_s)$ as

$$R^G(D_s, P_s) = L_{R^G}(\mathbf{D}_s^*, \mathbf{P}_s^*, s) - s_D D_s - s_P P_s. \tag{3.2.10}$$

Remark 3.2.1. *The assumption of strictly positive Lagrangian multipliers* (s_D, s_P) *implies finite* (D, P) *levels in Theorem 3.2.1, hence guaranteeing the convergence of Algorithm 3. However, under the assumption of bounded perception metric* $D(\cdot\|\cdot)$*, the case* $s_P = 0$ *does not violate the assumptions of Theorem 3.2.1. In this regime, the perception constraint of Problem 3.2.1 is inactive, thus making the problem equivalent to the classical RDF problem.*

Algorithm 3 Algorithm of Theorem 3.2.1

Require: source distribution $p_X = \mathcal{N}(\mu_X, \Sigma_X)$ with $\Sigma_X \succ 0$; Lagrangian parameters $s = (s_D, s_P)$ with $s_D > 0$ and $s_P > 0$; error tolerances ϵ; initial point $(\mathbf{D}^{(0)}, \mathbf{P}^{(0)})$.

1: $n \leftarrow 0$, $\omega^{(0)} \leftarrow +\infty$;
2: **while** $\omega^{(n)} > \epsilon$ **do**
3: $\quad n \leftarrow n + 1$
4: $\quad \mathbf{D}^{(n)} \leftarrow$ Solution Problem (3.2.8) for $(\mathbf{P}^{(n-1)}, s_D)$
5: $\quad \mathbf{P}^{(n)} \leftarrow$ Solution Problem (3.2.9) for $(\mathbf{D}^{(n)}, s_P)$
6: $\quad \omega^{(n)} \leftarrow L_{R^G}(\mathbf{D}^{(n-1)}, \mathbf{P}^{(n-1)}) - L_{R^G}(\mathbf{D}^{(n)}, \mathbf{P}^{(n)})$
7: **end while**

Ensure: $D_s = \sum_{i=1}^{N} g(D_i^{(n)})$, $P_s = \sum_{i=1}^{N} r(P_i^{(n)})$, $R^G(D_s, P_s) = \sum_{i=1}^{N} R_i^G(D_i^{(n)}, P_i^{(n)})$.

In Algorithm 3 we implement the alternating minimization scheme of Theorem 3.2.1 using the unconstrained Lagrangian formulation that was previously discussed, which allows for the computation of any $R^G(D, P)$ of the form characterized in (3.2.1) as long as we can have a characterization of the problem for the univariate case. In other words, we can always cast the general jointly Gaussian problem into the optimization problem in (3.2.3), which further means that the proposed alternating minimization approach can be applied whenever the RDPF can be obtained in closed-form for scalar-valued Gaussian sources assuming jointly Gaussian reconstructions.

The computation of the optimal pair $(\mathbf{D}^*, \mathbf{P}^*)$ via Algorithm 3, can be used to obtain numerically the value of matrices (A, Σ_W). This means that we can numerically compute the linear realization $Y = AX + W$ that achieves $R^G(D, P)$ in (3.2.1). On top of that, by leveraging the fact that (A, Σ_W, Σ_X), commute by pairs, we can expect that the aforementioned design variables will be of the form:

$$A = V \cdot \operatorname{diag}\left([\lambda_{A,i}]_{i=1:N}\right) \cdot V^T, \qquad \Sigma_W = V \cdot \operatorname{diag}\left([\lambda_{\Sigma_W,i}]_{i=1:N}\right) \cdot V^T$$

where $V \in \mathbb{R}^{N \times N}$ is a non-singular orthogonal matrix, and $\{(\lambda_{A,i}, \lambda_{\Sigma_W,i}) : \ i = 1, 2, \ldots, N\}$ are both functions of $\{(\lambda_{X,i}, D_i, P_i) : \ i = 1, 2, \ldots, N\}$ which are associated with the parameters for the associated stagewise solution of $R^G(D, P)$.

Analysis of Algorithm 3 To characterize the worst-case performance of Algorithm 3, we provide an upper bound on its convergence rate.

Theorem 3.2.2. *(Upper bound on the Convergence Rate) Let $\{(\mathbf{D}^{(n)}, \mathbf{P}^{(n)})\}_{n=0,\ldots,T}$ be the sequence of iterations generated by Algorithm 3 in T iterations and let $(\mathbf{D}^*, \mathbf{P}^*)$ be a minimizer of $L_{R^G}(\cdot, \cdot)$. Then, there exists positive and finite constant C such that Alg. 3 guarantees that*

$$L_{R^G}(\mathbf{D}^{(T)}, \mathbf{P}^{(T)}) - L_{R^G}(\mathbf{D}^*, \mathbf{P}^*) \leq \frac{C}{T} \tag{3.2.11}$$

i.e., the asymptotic rate of convergence of Alg. 3 is upper bounded by $\mathcal{O}\left(\frac{1}{T}\right)$ (sublinear

convergence rate).

Proof. See Appendix 3.6.6. □

In what follows, we demonstrate experimentally that indeed at worst Alg. 3 can achieve sublinear convergence, but depending on the value of the Lagrangian multipliers the performance can significantly improve, achieving even a linear convergence rate.

Example 3.2.1. *Let $X \sim \mathcal{N}(0, \Sigma_X)$ with $\Sigma_X = \mathrm{diag}([1, 3, 5, 7, 10])$, let Problem 3.2.1 be defined for MSE distortion metric and W_2^2 perception metric and let $\omega^{(n)}$ be defined as in Alg. 3, i.e., $\omega^{(n)} = L_{R^G}(\mathbf{D}^{(n-1)}, \mathbf{P}^{(n-1)}) - L_{R^G}(\mathbf{D}^{(n)}, \mathbf{P}^{(n)})$. Then, Figure 3.1 shows the convergence rate of Alg. 3 for different values of the Lagrangian multipliers $s_D \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ and $s_P \in \{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. Clearly, Figure 3.1 illustrates that depending on the values of the Lagrangian multipliers, the convergence rate of the algorithm can be improved, achieving even linear convergence rates.*

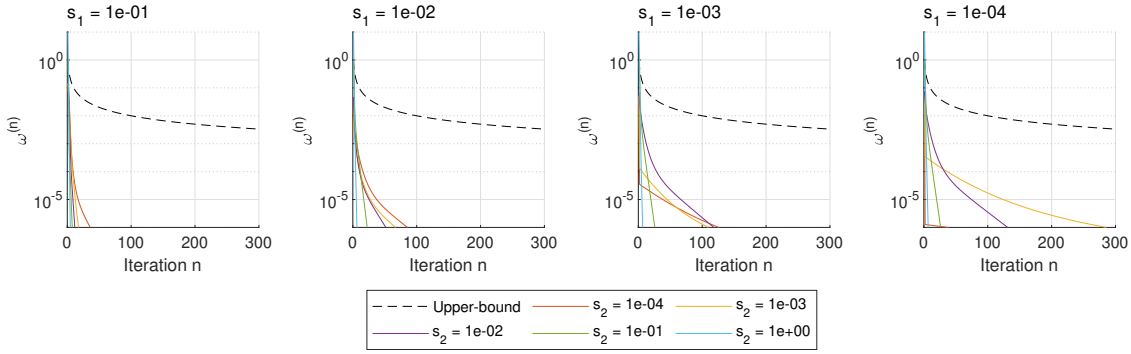


Figure 3.1: Experimental convergence rate of Alg. 3 for a 5-dimensional Gaussian source.

3.2.2 Application of the Alternating Minimization Approach

In this subsection, we specialize the analysis of Subsection 3.2.1 to the MSE distortion constraint whereas as perception constraint we consider the squared Wasserstein-2 distance. We remark that similar specializations can be developed for other divergence measures, under the assumption that a closed-form solution for scalar-valued Gaussian RDPF, $R^G(D, P)$, is available (see e.g., Section 3.1). We start by solving the subproblems (3.2.8) and (3.2.9). To this end, we leverage Lemma 3.1.1 to characterize the function $R_i^G(D_i, P_i)$, which is the stagewise Gaussian RDPF for the i^{th} dimension, under MSE distortion metric and W_2^2 perception metric, for a Gaussian source $X_i \sim \mathcal{N}(0, \lambda_{\Sigma_X,i})$. Furthermore, Lemma 3.1.1 and Proposition 3.4.3 defines the functional form of the auxiliary optimization variables $\mathbf{D} = [D_i]_{i\in 1:N}$ and $\mathbf{P} = [P_i]_{i\in 1:N}$ and of the tensorization functions $g(\cdot)$ and $r(\cdot)$, introduced in (3.2.3), as follows:

$$D_i = \mathbb{E}\left[\|X_i - Y_i\|^2\right] = (1 - \lambda_{A,i})^2 \lambda_{\Sigma_X,i} + \lambda_{\Sigma_W,i}$$

Table 3.1: Closed-Form Solutions of Subproblem (3.2.5)

$d(p_X\|p_Y)$	$\mathbf{D}^*$
$\mathrm{D}_{\mathrm{KL}}(p_X\|p_Y)$	$D_i^* = \frac{1}{2s_D} + \lambda_{\Sigma_X,i}\left(1 - \frac{1}{Z(P_i)}\right) + \frac{1}{Z(P_i)}\sqrt{Z(P_i)\left(\frac{Z(P_i)}{4s_D^2} - 4\lambda_{\Sigma_X,i}^2\right)}$ $Z(P_i) = \mathcal{W}_{-1}\left(-e^{-(2P_i+1)}\right)$
$\mathrm{D}_{\mathrm{KL}}(p_Y\|p_X)$	$D_i^* = \frac{1}{2s_D} + \lambda_{\Sigma_X,i}\left(1 - Z(P_i)\right) - \sqrt{4\lambda_{\Sigma_X,i}^2(-Z(P_i)) + \frac{1}{4s_D^2}}$ $Z(P_i) = \mathcal{W}_0\left(-e^{-(2P_i+1)}\right)$
$\mathrm{H}^2(p_X,p_Y)$	$D_i^* = \frac{1}{2s_D} + 2Z(P_i)\lambda_{\Sigma_X,i} - \sqrt{\frac{1}{4s_D^2} + 4\lambda_{\Sigma_X,i}^2\left(2Z(P_i) - 1\right)}$ $Z(P_i) = \left(1 - \sqrt{1 - (1 - \frac{P_i}{2})^4}\right)\left(1 - \frac{P_i}{2}\right)^{-4}$
$\mathrm{D}_{\mathrm{GJS}}(p_X\|p_Y)$	$D_i^* = \frac{1}{2s_D} - \lambda_{\Sigma_X,i}\left(Z(P_i) + \sqrt{Z(P_i)(Z(P_i)+2)}\right)$ $-\sqrt{\frac{1}{4s_D^2} - 4\lambda_{\Sigma_X,i}^2\left(Z(P_i) + \sqrt{Z(P_i)(Z(P_i)+2)} + 1\right)}$, $Z(P_i) = \mathcal{W}_{-1}\left(-e^{-(4P_i+2)}\right)$

$$
\begin{aligned}
P_i &= \mathrm{W}_2^2(p_{X_i}, p_{Y_i}) = \lambda_{\Sigma_X,i} - \sqrt{{\lambda_{A,i}}^2\lambda_{\Sigma_X,i} + \lambda_{\Sigma_W,i}} \\
g(\cdot) &= r(\cdot) = \mathrm{id}(\cdot)
\end{aligned}
$$

where $Y_i \sim \mathcal{N}(0, \lambda_{\Sigma_{X_i},i})$ is the stagewise linear realization of the form $Y_i = \lambda_{A,i}X_i + W_i$ with $W_i \sim \mathcal{N}(0, \lambda_{\Sigma_W,i})$.

In the following theorem we derive the characterization of the optimal solution of the subproblem (3.2.8).

Theorem 3.2.3. *Let the Lagrangian multiplier $s_D > 0$ be given. Then, for fixed $\mathbf{P}$, the optimal stagewise distortions levels $\mathbf{D}^*(\mathbf{P}) = [D_i^*(P_i)]_{i\in i:N} \in \mathcal{S}$ achieving the minimum of* (3.2.8) *are*

$$
D_i^* = P_i + 2\sqrt{\lambda_{\Sigma_X,i}}\left(\sqrt{\lambda_{\Sigma_X,i}} - \sqrt{P_i}\right) + \left(\frac{1}{2s_D} - \sqrt{4\lambda_{\Sigma_X,i}(\sqrt{\lambda_{\Sigma_X,i}} - \sqrt{P_i})^2 + \frac{1}{4s_D^2}}\right). \tag{3.2.12}
$$

Proof. See Appendix 3.6.7. □

We remark that similar expressions to (3.2.12) can be derived for other perception constraints. Such cases are reported in Table 3.1. We now move to the characterization of the optimal solution of the subproblem (3.2.6).

Theorem 3.2.4. *Let the Lagrangian multiplier $s_P > 0$ be given. Then, for fixed $\mathbf{D}$, the optimal stagewise perception levels $\mathbf{P}^*(\mathbf{D}) = [P_i^*(D_i)]_{i\in i:N} \in \mathcal{S}$ achieving the minimum*

of (3.2.9) *can be characterized as the zeros of the vector function* $T(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$ *where each component is defined as follows:*

$$T_i(x) \triangleq \left.\frac{\partial R_i^G(D_i, P_i)}{\partial P_i}\right|_{x_i} + s_P \tag{3.2.13}$$

$$\frac{\partial R_i^G(D_i, P_i)}{\partial P_i} \triangleq \begin{cases} \frac{1}{2}\frac{(\sqrt{\lambda_{\Sigma_X,i}}-\sqrt{P_i})^4-(\lambda_{\Sigma_X,i}-D_i)^2}{\sqrt{P_i}(D_i-P_i)(\sqrt{P_i}-\sqrt{\lambda_{\Sigma_X,i}})(D_i-(2\sqrt{\lambda_{\Sigma_X,i}}-\sqrt{P_i})^2)} & \text{if } P_i \in \mathcal{S} \\ 0 & \text{if } P_i \in \mathcal{S}^c. \end{cases} \tag{3.2.14}$$

Proof. See Appendix 3.6.8. □

Remark 3.2.2. *Similar expressions to* (3.2.13) *can be derived for the direct* D_{KL}*, reverse* D_{KL} *and* $\mathrm{D}_{\mathrm{GJS}}$ *divergences cases. On the contrary, the case of the* H^2 *distance requires particular care. Using Proposition 3.4.4, we can express the* H^2 *tensorization inequality as*

$$\mathrm{H}^2(p_X, p_Y) \geq r^{-1}\left(\sum_{i=1}^N r\left(\mathrm{H}^2(p_{X_i}, p_{Y_i})\right)\right)$$

where $r : [0, 2) \to \mathbb{R}^+$ *is the convex strictly increasing bijection defined as* $r(x) = -\log(1 - \frac{x}{2})$ *and* r^{-1} *is its inverse. Therefore, the perception constraint in* (3.2.2) *can be expressed as follows:*

$$r^{-1}\left(\sum_{i=1}^N r\left(\mathrm{H}^2(p_{X_i}, p_{Y_i})\right)\right) \leq P \overset{(a)}{\Longrightarrow} \sum_{i=1}^N r\left(\mathrm{H}^2(p_{X_i}, p_{Y_i})\right) \leq r(P).$$

where (a) is due to the monotonicity of h*. Using the modified perception level* $P' = r(P)$*, the new formulation respects the constraint format described in* (3.2.3)*.*

Corollary 3.2.1. *Let* $T_i : \mathbb{R}^N \to \mathbb{R}$ *be the* i^{th} *component of the vector function* $T(\cdot)$ *defined in Theorem 3.2.4. Then,* T_i *is a continuous and non-decreasing function on* $\mathbb{R}$*. Furthermore,* T_i *has at least one root in* $\mathcal{S}$*.*

Proof. See Appendix 3.6.9. □

Although we are not able to derive a closed-form solution for subproblem (3.2.6), the optimal $\mathbf{P}^*$ can be found as zeros of the functions $\{T_i\}_{i\in 1:N}$. Corollary 3.2.1 guarantees that the functions $\{T_i\}_{i\in 1:N}$ respect the assumptions required for the application of the bisection method [48, Chapter 2.1]. More refined root-finding methods, such as the Newton Method [48, Chapter 2.3], are not applicable in this instance given the requirement on the differentiability of T_i, which cannot be guaranteed, in general.

3.2.3 Gaussian RDPF in the "Perfect Realism" regime

The results of Theorem 3.2.3 characterize the optimal distortion vector $\mathbf{D}$ for a given perception vector $\mathbf{P}$. As an additional result, Theorem 3.2.3 gives the closed-form solution for the case where $\mathbf{P} = \mathbf{0}$ (all zeros vector), or equivalently $D(p_X \| p_Y) = 0$, referred to as *perfect realism* [34, 61].

Corollary 3.2.2. *Consider the optimization problem* (3.2.7) *for perception level* $\mathbf{P} = \mathbf{0}$. *Then, for a given Lagrangian multiplier* $s_D > 0$, *the optimal solution* $\mathbf{D}^* = [D_i^*]_{i \in 1:N}$ *is given by*

$$D_i^* = 2\lambda_{\Sigma_X,i} + \frac{1}{2s_D} - \sqrt{4\lambda_{\Sigma_X,i}^2 + \frac{1}{4s_D^2}} \tag{3.2.15}$$

such that the distortion level $D = \sum_{i=1}^N D_i^*$.

Proof. (3.2.15) is obtained from (3.2.12) for $P_i = 0$. □

We stress the following two technical remarks for Corollary 3.2.2.

Remark 3.2.3. *The optimal solution* $\mathbf{D}^*$ *is well defined in the limit* $s_D \to 0$, *since* $\lim_{s_D \to 0} D_i^* = 2\lambda_{\Sigma_X,i}$.

Remark 3.2.4. *In the water-filling solution of the classical multivariate Gaussian RDF, the optimal solution* $\mathbf{D}_{RD}^* = [D_{i,RD}^*]_{i \in 1:N}$ *for a Lagrangian multiplier* $s_D > 0$ *is given by*

$$D_{i,RD}^* = \min\left(w(s_D), \lambda_{\Sigma_X,i}\right) \qquad w(s_D) = \frac{1}{2s_D}$$

where water-level $w(s_D)$ *is dimension independent and the* $\min(\cdot)$ *operation is required to guarantee that* $D_{i,RD}^*$ *belongs to the constraint set. Heuristically, one can imagine that the* i*-th source component is discarded in the reconstruction whenever* $w(s_D) \geq \lambda_{\Sigma_X,i}$, *upper bounding the maximum distortion observed in the* i*-th component by* $\lambda_{\Sigma_X,i}$. *Conversely, the solution identified in* (3.2.15), *and in general the results of Theorem 3.2.3, can be interpreted as an adaptive water-level solution. In fact, in* (3.2.15), D_i^* *gets adapted to each dimension, guaranteeing that all source components are present in the reconstructed signal. However, as already observed in [9, Theorem 2],* (3.2.15) *suggests that the maximum distortion on the* i*-th dimension is greater than* $\lambda_{\Sigma_X,i}$ *and is upper bounded by* $2\lambda_{\Sigma_X,i}$.

3.3 Numerical Results

In this section, we provide numerical simulations based on the findings of Sections 3.1, 3.2.

3.3.1 Scalar-valued Gaussian Sources

Let $X \sim \mathcal{N}(0,1)$ be a scalar Gaussian source. Figure 3.2 shows $R^G(D,P)$ for X under MSE distortion measure and, respectively, direct Kullback–Leibler divergence (3.1(a)), re-

verse Kullback–Leibler divergence (3.1(b)), Geometric Jensen-Shannon divergence (3.1(c)), squared Hellinger distance (3.1(d)), and squared Wasserstein-2 distance (3.1(e)) perception measures. All the derived Gaussian RDPFs share similarities in the structure of the operating regions on the (D, P) plane. We distinguish three cases:

- **Case I**, where the $R^G(D, P)$ is identically similar to the associated RDF. In this regime, the perception constraint is not met with equality.
- **Case II**, where, due to the distortion constraint not met with equality, the $R^G(D, P)$ is identically zero.
- **Case III**, where both the distortion and perception constraints are met with equality.

In our derivations, we identify the operating regions $\{\textbf{Case I}\} \cup \{\textbf{Case II}\} = \mathcal{S}^c$ and $\{\textbf{Case III}\} = \mathcal{S}$, giving us the closed-form solutions of their the operating regions.

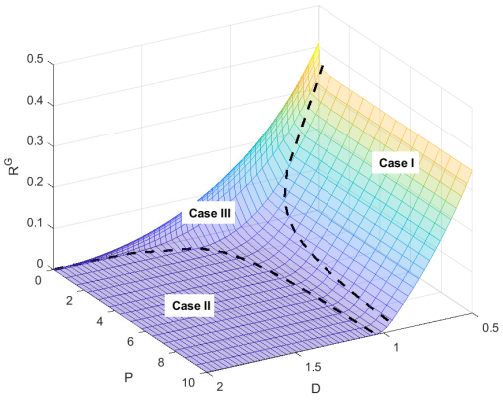
3.3.2 Multivariate Gaussian Sources

This section is devoted to the analysis of the numerical results of Algorithm 3. All the numerical experiments in this section have been conducted considering a multivariate Gaussian source $X \sim \mathcal{N}(0, \Sigma_X)$ with $\Sigma_X = \text{diag}([1, 3, 5])$ and setting the initialization point for Algorithm 3 to $(\mathbf{D}^{(0)}, \mathbf{P}^{(0)}) = (0, 0)$.

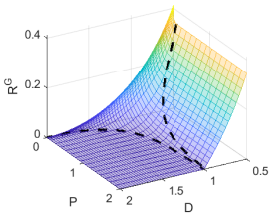
Gaussian RDPF Curves

In Figure 3.3, we illustrate $R^G(D, P)$ for the perception metrics introduced in Section 3.1.

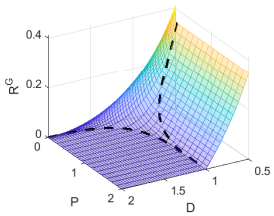
Focusing on Figure 3.2(a), we compare $R^G(D, P)$ under the squared Wasserstein-2 perception with the solution of the classical RDF problem using the reverse water-filling algorithm (black line) where for the classical case the perception measure has been computed a posteriori using the same divergence metric. The result confirms that for bounded divergence measure the RD solution can be obtained as an extreme case of $R^G(D, P)$ surface, retrieving the boundary between the regions of *Case I* and *Case III*. Moreover, the surface region of *Case I* can be retrieved by the rigid translation of the obtained boundary curve, since any (D, P) point in this region defines a Gaussian RDPF problem where the perception constraint is not active, and thus equivalent to the classical RDF problem. Similar remarks can be extended to the case of the $R^G(D, P)$ under H^2 perception metric. In Figs. 3.2(c), 3.2(d), and 3.2(e), we cannot proceed with the same kind of comparison due to properties of the Kullback–Leibler and Geometric Jensen-Shannon divergences; depending on the distortion level D, the classical RDF solution may induce the variance of one of the marginal distributions of the reconstructed source $\sigma^2_{Y_i} \to 0$. In this condition, the absolute continuity between p_X and p_Y is not guaranteed, causing the measured perception level $P \to +\infty$. Since Algorithm 3 presents identical behavior to the reverse water-filling algorithm for $s_P \to 0$ (equivalent to $P \to \infty$), for



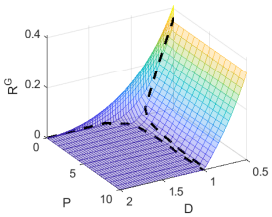
(a): Direct KL



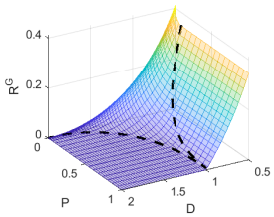
(b): Reverse KL



(d): HS

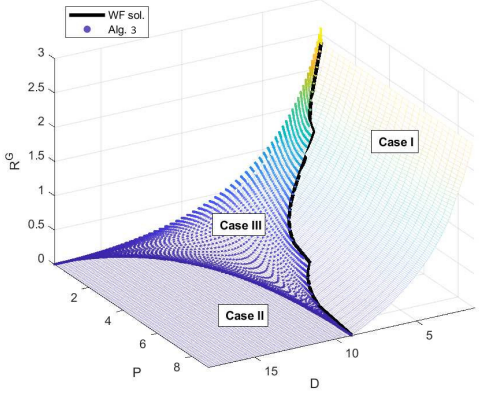


(c): GJS

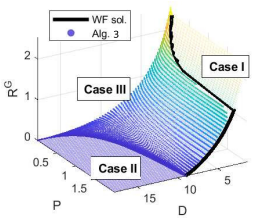


(e): W_2^2

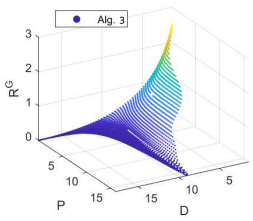
Figure 3.2: $R^G(D,P)$ **for a univariate Gaussian source** $X \sim \mathcal{N}(0,1)$ **under (a) direct** KL**, (b) reverse** KL**, (c)** GJS**, (d)** HS**, and (e)** $\mathcal{W}_2$ **perception constraints.**



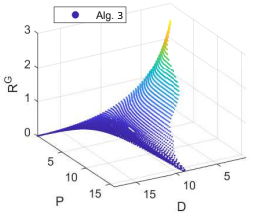
(a): W_2^2



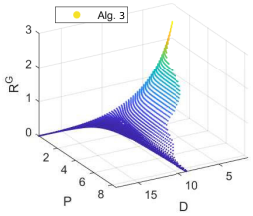
(b): HS



(d): Reverse KL



(c): Direct KL



(e): GJS

Figure 3.3: $R^G(D,P)$ **for a multivariate Gaussian source** $X \sim \mathcal{N}(0, \mathrm{diag}([1,3,5]))$ **under (a)** $\mathcal{W}_2$**, (b)** HS**, (c) direct** KL**, (d) reverse** KL**, and (e)** GJS **perception constraints.**

these cases, the various Gaussian RDPFs have been computed bounding the maximum perception level P imposing $s_P \geq 10^{-3}$.

Adaptive Water-Level

In Figure 3.4 we analyze the per-dimension levels of distortion D_i^* and perception P_i^* between the source X and its reconstruction Y, comparing Algorithm 3 with the classical RD reverse water-filling solution. The comparison has been conducted with a target distortion level $D = 6$ and varying the target perception level P. The most constrained case ($P \approx 0$) provides a clear example of the behavior mentioned in Remark 3.2.4, with

D_i^* not following a uniform water-level, but one that adapts to each marginal. Similar behaviors are also present, although less pronounced, in the remaining two cases ($P \approx 0.7$ and $P \approx 2$), where the looser perception level P induces a more uniform repartition of the per-dimension distortions D_i^*, thus demonstrating a behavior that is closer to the classical reverse water-filling solution.

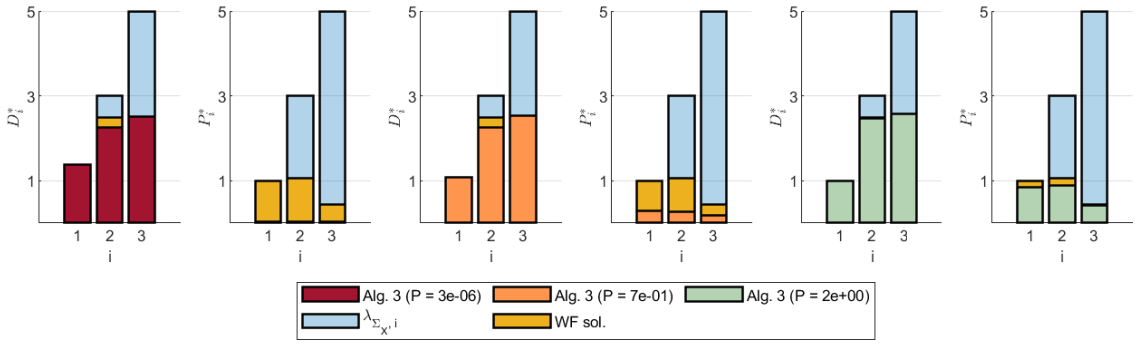


Figure 3.4: Comparison of the per-dimension distortion D_i^* and perception P_i^* measures for a fixed target distortion level $D = 6$ between the water-filling solution and Algorithm 3.

3.4 Useful Results Obtained via Tensorization

In this Appendix, we state and when needed prove certain useful results on tensorization of divergence functions and information measures.

First, we state a proposition which demonstrates how $\mathrm{D_{KL}}(\cdot\|\cdot)$ can be tensorized. For a proof, see for instance [58, Proposition 4].

Proposition 3.4.1. *Let $\Sigma_X \succ 0$ and $\Sigma_Y \succ 0$ on $\mathbb{R}^{N\times N}$. Moreover, let $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y \sim \mathcal{N}(0, \Sigma_Y)$. Then, $\mathrm{D_{KL}}(p_X\|p_Y)$ can be bounded from below as follows:*

$$\mathrm{D_{KL}}(p_X\|p_Y) \geq \sum_{i=1}^{N} \mathrm{D_{KL}}(p_{X_i}\|p_{Y_i})$$

and the inequality holds with equality iff Σ_X and Σ_Y are commuting matrices[3].

Similar to the case of KL divergence, the GJS divergence can also be tensorized. This is stated and proved next.

Proposition 3.4.2. *Let $\Sigma_X \succ 0$ and $\Sigma_Y \succ 0$ on $\mathbb{R}^{N\times N}$. Moreover, let $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y \sim \mathcal{N}(0, \Sigma_Y)$. Then, the GJS divergence $\mathrm{D_{GJS}}(p_X\|p_Y)$ can be bounded from below as follows:*

$$\mathrm{D_{GJS}}(p_X\|p_Y) \geq \sum_{i=1}^{N} \mathrm{D_{GJS}}(p_{X_i}\|p_{Y_i})$$

[3]The definition of commuting matrices can be found in [29, Section 0.7.7].

and the inequality holds with equality iff Σ_X and Σ_Y are commuting matrices.

Proof. The tensorization of the $\mathrm{D}_{\mathrm{GJS}}$ follows by observing that the geometric mean distribution p_g, as stated in Definition 1.4.4, is itself Gaussian. Hence, if Σ_X and Σ_Y commute, then Σ_g necessarily commutes by pairs [29, Section 0.7.7], thus allowing to directly apply [58, Proposition 4] to (1.4.4) and obtain the result in question. This concludes the proof. □

Similarly, one can show that W_2^2 and H^2 can be tensorized. These two results are stated next without any proof (for completeness one can check the supplementary material).

Proposition 3.4.3. *Let $\Sigma_X \succ 0$ and $\Sigma_Y \succ 0$ on $\mathbb{R}^{N\times N}$. Moreover, let $X \sim \mathcal{N}(0,\Sigma_X)$ and $Y \sim \mathcal{N}(0,\Sigma_Y)$. Then, $\mathrm{W}_2^2(p_X,p_Y)$ can be bounded from below as follows:*

$$\mathrm{W}_2^2(p_X,p_Y) \geq \sum_{i=1}^{N} \mathrm{W}_2^2(p_{X_i},p_{Y_i})$$

and the inequality holds with equality iff Σ_X and Σ_Y are commuting matrices.

Proposition 3.4.4. *Let $\Sigma_X \succ 0$ and $\Sigma_Y \succ 0$ on $\mathbb{R}^{N\times N}$. Moreover, let $X \sim \mathcal{N}(0,\Sigma_X)$ and $Y \sim \mathcal{N}(0,\Sigma_Y)$. Then, $\mathrm{H}^2(p_X,p_Y)$ can be bounded from below as follows:*

$$\mathrm{H}^2(p_X,p_Y) \geq 2\left(1-\prod_{i=1}^{N}\left(1-\frac{\mathrm{H}^2(p_{X_i},p_{Y_i})}{2}\right)\right) \tag{3.4.1}$$

and the inequality holds with equality iff Σ_X and Σ_Y are commuting matrices with increasing (or decreasing) eigenvalues order.

Next, we derive a tensorization result for mutual information of jointly Gaussian random vectors.

Lemma 3.4.1. *Let (X,Y) be jointly Gaussian random vectors on $\mathbb{R}^N$, such that $X \sim \mathcal{N}(0,\Sigma_X)$ with $\Sigma_X \succ 0$ and $Y = AX + W$, with $A \in \mathbb{R}^{N\times N}$ being invertible and diagonalizable, and $W \sim \mathcal{N}(0,\Sigma_W)$ with $\Sigma_W \succ 0$. Then, the mutual information $I(X,Y)$ can be bounded from below as follows:*

$$I(X,Y) = \log\left(\frac{|\Sigma_Y|}{|\Sigma_{Y|X}|}\right) \geq \sum_{i=1}^{N} \log\left(1+\frac{\lambda_{A,i}^2\lambda_{\Sigma_X,i}}{\lambda_{\Sigma_W,i}}\right) \tag{3.4.2}$$

and the inequality holds with equality iff A is a symmetric matrix and the matrices Σ_X, Σ_W and A commute by pairs.

Proof. Note that for jointly Gaussian random vectors (X,Y), mutual information between X and Y can be written as follows:

$$I(X,Y) = \log\left(\frac{|\Sigma_Y|}{|\Sigma_{Y|X}|}\right) = \log\left(\frac{|A\Sigma_X A^T+\Sigma_W|}{|\Sigma_W|}\right) = \log\left(|I+\Sigma_W^{-\frac{1}{2}}A\Sigma_X A^T\Sigma_W^{-\frac{1}{2}}|\right)$$

$$
\begin{aligned}
&= \log\left(\left|I + \left(\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right)\left(\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right)^T\right|\right) = \log\left(\left|I + \left(\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right)^T\left(\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right)\right|\right) \\
&\overset{(a)}{=} \sum_{i=1}^{N} \log\left(1 + \lambda_i\left[\left(\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right)^T\left(\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right)\right]\right) \\
&\overset{(b)}{=} \sum_{i=1}^{N} \log\left(1 + \left(s_i\left[\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right]\right)^2\right)
\end{aligned} \tag{3.4.3}
$$

where (a) follows from the fact that for any diagonalizable matrix M the eigenvalues $\lambda_{I+M} = 1 + \lambda_M$ and (b) follows from the definition of singular values of a matrix. In addition, we can also verify the following majorization inequality:

$$
\begin{aligned}
\log\left(s\left[\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right]\right) &\overset{(c)}{\succ} \log\left(s^{\downarrow}\left[\Sigma_W^{-\frac{1}{2}} A\right]\right) + \log\left(s^{\uparrow}\left[\Sigma_X^{\frac{1}{2}}\right]\right) \\
&\overset{(d)}{\succ} \log\left(s^{\downarrow}\left[\Sigma_W^{-\frac{1}{2}}\right]\right) + \log\left(s^{\uparrow}\left[A\right]\right) + \log\left(s^{\uparrow}\left[\Sigma_X^{\frac{1}{2}}\right]\right) \\
&\overset{(e)}{\succ}_w \log\left(\lambda^{\downarrow}\left[\Sigma_W^{-\frac{1}{2}}\right]\right) + \log\left(\left|\lambda^{\uparrow}\left[A\right]\right|\right) + \log\left(\lambda^{\uparrow}\left[\Sigma_X^{\frac{1}{2}}\right]\right) \\
&= \log\left(\frac{1}{\lambda^{\uparrow}\left[\Sigma_W^{\frac{1}{2}}\right]}\right) + \log\left(|\lambda_A^{\uparrow}|\right) + \log\left(\lambda^{\uparrow}\left[\Sigma_X^{\frac{1}{2}}\right]\right)
\end{aligned} \tag{3.4.4}
$$

where (c) and (d) hold from Lidskii's theorem [28, Theorem 3.4.6] and hold with equality for commuting matrices Σ_X, Σ_W and A with the correct singular value ordering, while (e) follows from Weyl's majorization theorem [28, Theorem 2.3.6] and holds with equality if and only if Σ_X, A, and Σ_W are a symmetric matrices.
Now, let $f(\cdot)$ to be the function defined as $f(x) \triangleq \log(1 + x^2)$. Although f is not convex, we can readily show that the composite function $t \to f(e^t)$ is convex and increasing on $\mathbb{R}$. Thus, we can invoke Weyl's majorization theorem, which states:

$$
\log(x) \succ_w \log(y) \implies f(x) \succ_w f(y). \tag{3.4.5}
$$

Therefore, we obtain:

$$
(3.4.3) = \sum_{i=1}^{N} f\left(s_i\left[\Sigma_W^{-\frac{1}{2}} A \Sigma_X^{\frac{1}{2}}\right]\right) \overset{(3.4.5)+(3.4.4)}{\geq} \sum_{i=1}^{N} f\left(\frac{|\lambda_{A,i}^{\uparrow}| \cdot \sqrt{\lambda_{\Sigma_X,i}}^{\uparrow}}{\sqrt{\lambda_{\Sigma_W,i}}^{\uparrow}}\right) = (3.4.2).
$$

This concludes the proof. □

We conclude this section by showing that the MSE admits a lower bound via tensorization for jointly Gaussian random vectors.

Lemma 3.4.2. *Let (X, Y) be jointly Gaussian random vectors on $\mathbb{R}^N$ such that $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y = AX + W$, with $A \in \mathbb{R}^{N \times N}$ being invertible and diagonalizable, and $W \sim \mathcal{N}(0, \Sigma_W)$. Moreover, let Σ_X and Σ_W be positive definite matrices. Then, the MSE*

can be bounded from below as follows:

$$\mathbb{E}\left[\|X-Y\|^2\right] \geq \sum_{i=1}^{N} \mathbb{E}\left[(X_i-Y_i)^2\right]$$

whereas the inequality holds with equality if A is symmetric and (A, Σ_W, Σ_X) commute by pairs.

Proof. Note that we can expand the MSE as follows:

$$\begin{aligned}\mathbb{E}\left[\|X-Y\|^2\right] &= \operatorname{Tr}\left(\Sigma_X+\Sigma_Y-2A\Sigma_X\right) \overset{(a)}{=} \operatorname{Tr}\left[(I-A)\Sigma_X(I-A)^T+\Sigma_W\right] \\ &= \operatorname{Tr}\left[\left((I-A)\Sigma_X^{\frac{1}{2}}\right)^T\left((I-A)\Sigma_X^{\frac{1}{2}}\right)+\Sigma_W\right] \overset{(b)}{=} \|(I-A)\Sigma_X^{\frac{1}{2}}\|_F^2+\operatorname{Tr}[\Sigma_W]\end{aligned} \tag{3.4.6}$$

where (a) follows from the substitution $\Sigma_Y = A\Sigma_X A^T + \Sigma_W$ and (b) follows from the definition of Frobenius norm $\|\cdot\|_F$. Moreover, from (3.4.6) we can derive

$$\begin{aligned}\|(I-A)\Sigma_X^{\frac{1}{2}}\|_F^2 = \sum_{i=1}^{N} s_i^2\left[(I-A)\Sigma_X^{\frac{1}{2}}\right] &\overset{(c)}{\geq} \sum_{i=1}^{N}\left(s_i^{\downarrow}[I-A]\right)^2\lambda_i^{\uparrow}(\Sigma_X) \\ &\overset{(d)}{\geq} \sum_{i=1}^{N}\left(|\lambda_i^{\downarrow}[I-A]|\right)^2\lambda_i^{\uparrow}(\Sigma_X) \geq \sum_{i=1}^{N}\left(|1-\lambda_i^{\uparrow}[A]|\right)^2\lambda_i^{\uparrow}(\Sigma_X)\end{aligned}$$

where (c) follows from Lidskii's theorem [28, Theorem 3.4.6] and hold with equality for commuting matrices (A, Σ_X) with the correct singular value ordering, while (d) follows from Weyl's majorization theorem [28, Theorem 2.3.6] and holds with equality if and only if A is a symmetric matrix. This concludes the proof. □

3.5 Mean Matching for Divergence Measures

Lemma 3.5.1. *(Mean matching) Let $D(\cdot\|\cdot)$ be a divergence function chosen between direct* $\mathrm{D_{KL}}$*, reverse* $\mathrm{D_{KL}}$*,* $\mathrm{D_{GJS}}$*,* H^2*, and* W_2^2*. Let $p \sim \mathcal{N}(\mu_p, \Sigma_p)$ and $q \sim \mathcal{N}(\mu_q, \Sigma_q)$ and let $\hat{q}$ be a shifted version of q such that the first moments of $\hat{q}$ and p match, i.e. $\hat{q} \sim \mathcal{N}(\mu_p, \Sigma_q)$. Then $d(p\|q) \geq d(p\|\hat{q})$.*

Proof. The proof follows from the substitutions of the analytical expressions of the considered divergences in the Gaussian case, for which we can verify that

$$\begin{aligned}
&(\textit{direct } \mathrm{D_{KL}}) && \mathrm{D_{KL}}(p\|q)-\mathrm{D_{KL}}(p\|\hat{q}) = (\mu_p-\mu_q)^T\Sigma_q^{-1}(\mu_p-\mu_q) \geq 0 \\
&(\textit{reverse } \mathrm{D_{KL}}) && \mathrm{D_{KL}}(q\|p)-\mathrm{D_{KL}}(\hat{q}\|p) = (\mu_q-\mu_p)^T\Sigma_p^{-1}(\mu_q-\mu_p) \geq 0 \\
&(\mathrm{H}^2) && \mathrm{H}^2(p\|q)-\mathrm{H}^2(p\|\hat{q}) = 2\frac{|\Sigma_p\Sigma_q|^{\frac{1}{4}}}{|\frac{\Sigma_p+\Sigma_q}{2}|^{\frac{1}{2}}}\left(1-e^{-\frac{1}{4}(\mu_p-\mu_q)^T(\Sigma_p+\Sigma_q)^{-1}(\mu_p-\mu_q)}\right) \geq 0 \\
&(\mathrm{W}_2^2) && \mathrm{W}_2^2(p\|q)-\mathrm{W}_2^2(p\|\hat{q}) = \|\mu_p-\mu_q\|^2 \geq 0
\end{aligned}$$

$$(\mathrm{D_{GJS}})\ \ \mathrm{D_{GJS}}(p\|q) - \mathrm{D_{GJS}}(p\|\hat{q}) = \frac{1}{2}(\mu_g - \mu_p)^T \Sigma_g^{-1}(\mu_g - \mu_p) + \frac{1}{2}(\mu_g - \mu_q)^T \Sigma_g^{-1}(\mu_g - \mu_q) \geq 0$$

where for the $\mathrm{D_{GJS}}$ divergence, μ_g and Σ_g are respectively the mean and covariance matrix of the geometric mean between the distributions p and q (see Section 1.4.2). This concludes the proof. □

3.6 Appendix

3.6.1 Proof of Theorem 3.1.1

First note that once we substitute the functional form of the direct $\mathrm{D_{KL}}(p_X\|p_Y)$ in (3.1.1), we obtain the following convex optimization problem:

$$R^G(D,P) = \min_{a\in\mathbb{R},\sigma_W^2\geq 0} \frac{1}{2}\log\left(1 + a^2\frac{\sigma_X^2}{\sigma_W^2}\right). \tag{3.6.1}$$

$$\text{s.t.} \quad (1-a)^2\sigma_X^2 + \sigma_W^2 \leq D \tag{3.6.2}$$

$$\frac{1}{2}\left[\log\left(\frac{a^2\sigma_X^2+\sigma_W^2}{\sigma_X^2}\right) + \frac{\sigma_X^2}{a^2\sigma_X^2+\sigma_W^2} - 1\right] \leq P. \tag{3.6.3}$$

To solve the specific optimization problem, we invoke Karush-Kuhn-Tucker (KKT) conditions [60]. Let $s_D \geq 0$ and $s_P \geq 0$ be the Lagrangian multipliers associated with the distortion and perception constraint, respectively. The Lagrangian function $L(a,\sigma_W^2,s_D,s_P)$ associated with (3.6.1)-(3.6.3) can give the unconstrained optimization problem as follows:

$$L(a,\sigma_W^2,s_D,s_P) = \frac{1}{2}\log\left(1 + a^2\frac{\sigma_X^2}{\sigma_W^2}\right) + s_D((1-a)^2\sigma_X^2 + \sigma_W^2 - D) + s_P\left\{\frac{1}{2}\left[\log\left(\frac{a^2\sigma_X^2+\sigma_W^2}{\sigma_X^2}\right) + \frac{\sigma_X^2}{a^2\sigma_X^2+\sigma_W^2} - 1\right] - P\right\}.$$

The *stationarity conditions* and *complementary slackness* are as follows:

$$\nabla_{(a,\sigma_W^2)} L(a,\sigma_W^2,s_D,s_P) = 0 \ \ \text{(stationarity condition)} \tag{3.6.4}$$

$$s_D((1-a)^2\sigma_X^2 + \sigma_W^2 - D) = 0 \tag{3.6.5}$$

$$s_P\left\{\frac{1}{2}\left[\log\left(\frac{a^2\sigma_X^2+\sigma_W^2}{\sigma_X^2}\right) + \frac{\sigma_X^2}{a^2\sigma_X^2+\sigma_W^2} - 1\right] - P\right\} = 0. \tag{3.6.6}$$

Next, we obtain the complete closed-form solution for the convex programming problem in (3.6.1)-(3.6.3) by breaking it into three distinct cases.

Case I: Suppose that $s_D > 0$ and $s_P = 0$, namely, only (3.6.2) is active in the optimization problem. Then, the problem is equivalent to the classical RDF (see, e.g., [62]), which has

known optimal realization $Y = aX + W$ with design variables (a, σ_W^2) given as follows:

$$a = \left(1 - \frac{D}{\sigma_X^2}\right), \ \sigma_W^2 = D - (1-a)^2\sigma_X^2. \tag{3.6.7}$$

Case II: Let $s_D = 0$ and $s_P > 0$, namely, only (3.6.3) is active in the optimization problem. Then, from (3.6.6), we obtain:

$$\log\left(\frac{\sigma_Y^2}{\sigma_X^2}\right) + \frac{\sigma_X^2}{\sigma_Y^2} = 2P + 1 \implies -\frac{\sigma_X^2}{\sigma_Y^2} \cdot e^{-\frac{\sigma_X^2}{\sigma_Y^2}} = -e^{-(2P+1)} \implies \frac{\sigma_X^2}{\sigma_Y^2} = -\mathcal{W}\left(-e^{-(2P+1)}\right) \tag{3.6.8}$$

where (3.6.8) is obtained through the application of the Lambert $\mathcal{W}$ function [30]. From (3.6.4) and (3.6.6), we can derive the optimal (a, σ_W^2), resulting in:

$$a = 0, \quad \sigma_W^2 = -\frac{\sigma_X^2}{\mathcal{W}\left(-e^{-(2P+1)}\right)} \tag{3.6.9}$$

which, once substituted in (3.6.1), shows that the inactive distortion constraint results in $R(D, P) = 0$. Furthermore, since the distortion level D is not met with equality, using (3.6.9) we can characterize the maximum achievable distortion as $\bar{D}(P) = \mathbb{E}\left[(X-Y)^2\right] = \left(1 - \frac{\sigma_X^2}{\mathcal{W}\left(-e^{-(2P+1)}\right)}\right)\sigma_X^2$. This allows us to express σ_W^2, similarly to **Case I**, as $\sigma_W^2 = \bar{D}(P) - (1-a)^2\sigma_X^2$.

Case III: Suppose that $s_D > 0$ and $s_P > 0$, namely, both (3.6.2) and (3.6.3) are active. Then, from (3.6.5) we have that $a^2\sigma_X^2 + \sigma_W^2 = D + (2a-1)\sigma_X^2$ and by substituting in (3.6.6) we obtain:

$$\begin{aligned}
&\log\left(\frac{D + (2a-1)\sigma_X^2}{\sigma_X^2}\right) + \frac{\sigma_X^2}{D + (2a-1)\sigma_X^2} = 2P + 1 \\
&\implies -\frac{\sigma_X^2}{D + (2a-1)\sigma_X^2} \cdot e^{-\frac{\sigma_X^2}{D+(2a-1)\sigma_X^2}} = -e^{-(2P+1)} \\
&\implies \frac{D + (2a-1)\sigma_X^2}{\sigma_X^2} = -\mathcal{W}(-e^{-(2P+1)})
\end{aligned}$$

that, upon further simplifications, results into the following choice of the design variables (a, σ_W^2):

$$a = \frac{1}{2}\left(1 - \frac{D}{\sigma_X^2} - \frac{1}{\mathcal{W}(-e^{-(2P+1)})}\right), \ \sigma_W^2 = D - (1-a)^2\sigma_X^2, \tag{3.6.10}$$

which once substituted in (3.6.1) give (3.1.2). It remains to determine domain boundaries for each of the three cases and to identify which branch of the Lambert $\mathcal{W}$ function makes the derived $R(D, P)$ compatible with the properties listed in Remark 1.4.1.

Note that the boundary between **Case I** and **Case III** can be identified by considering

only **Case I** and by substituting its solutions in (3.6.3), which yields

$$-\frac{\sigma_X^2}{\sigma_X^2 - D} \cdot e^{-\frac{\sigma_X^2}{\sigma_X^2 - D}} < -e^{-(2P+1)}. \tag{3.6.11}$$

Applying the Lambert $\mathcal{W}$ function to inequality (3.6.11) requires a decision between the distinct branches $\mathcal{W}_0$ and $\mathcal{W}_{-1}$, since the two have opposite monotonic behaviors (that is, monotonically increasing and monotonically decreasing, respectively). However, invoking the continuity of the RDPF, we select the branch of the Lambert $\mathcal{W}$ function resulting in $R(D, P)$ being continuous on the border. Of the two branches, only $\mathcal{W}_{-1}$ respects this continuity constraint. Therefore, (3.6.11) becomes

$$\frac{\sigma_X^2}{\sigma_X^2 - D} < -\mathcal{W}_{-1}\left(-e^{-(2P+1)}\right). \tag{3.6.12}$$

Similarly, one can identify the boundary between **Case II** and **Case III** by considering solely **Case II** and substituting its solution in (3.6.2), which results into the following strict inequality:

$$\frac{\sigma_X^2}{\sigma_X^2 - D} > \mathcal{W}_{-1}\left(-e^{-(2P+1)}\right). \tag{3.6.13}$$

From inequalities (3.6.12) and (3.6.13) we can identify the joint region of **Case I** and **Case II** on the (D, P) plane, denoted by the set $\mathcal{S}^c = \{(D, P) : (3.6.12) \cup (3.6.13)\}$. Subsequently, the region of **Case III** is identified as $\mathcal{S} = (\mathcal{S}^c)^c$, resulting in (3.1.4). The derived regions $\mathcal{S}$ and $\mathcal{S}^c$ are used to construct the function a and σ_W^2 in (3.1.5) and (3.1.6), where the term $\min\{D, \bar{D}(P)\}$ in (3.1.6) is justified by observing that $D \leq \bar{D}(P)$ in **Case I** and **Case III** while $D \geq \bar{D}(P)$ in **Case II**. This concludes the proof.

3.6.2 Proof of Lemma 3.1.1

We only sketch the proof as it uses the same logical structure of the proof of Theorem 3.1.1. The W_2^2 distance between two scalar Gaussian random variables $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y \sim \mathcal{N}(0, \sigma_Y^2)$ takes the form

$$\mathrm{W}_2^2(p_X, p_Y) = (\sigma_X - \sigma_Y)^2. \tag{3.6.14}$$

By substituting (3.6.14) in (3.1.1), we obtain the following convex optimization problem

$$R(D, P) = \min_{a, \sigma_W^2} \frac{1}{2} \log\left(1 + a^2 \frac{\sigma_X^2}{\sigma_W^2}\right) \tag{3.6.15}$$

$$\text{s.t.} \quad (1-a)^2 \sigma_X^2 + \sigma_W^2 \leq D \tag{3.6.16}$$

$$(3.6.14) \leq P. \tag{3.6.17}$$

Let $s_D \geq 0$ and $s_P \geq 0$ be the Lagrangian multipliers associated with the distortion and perception constraints, respectively. Similarly to Theorem 3.1.1, we will obtain the closed-form solution of the optimization problem in (3.6.15)-(3.6.17) by leveraging the KKT conditions [60], considering specific operating regions of the problem and identifying the sets where these solutions apply.

Case I: Suppose $s_D > 0$ and $s_P = 0$, namely, only (3.6.16) is active in the optimization problem. Then, the problem becomes equivalent to the classical RD, therefore, as in Theorem 3.1.1, the optimal design variables (a, σ_W^2) are equal to (3.6.7).

Case II: Let $s_D = 0$ and $s_P > 0$, namely only (3.6.17) holds with equality. As in Theorem 3.1.1, the inactive distortion constraint (3.6.16) implies that $R(D, P) = 0$ on this operating region. Moreover, we can design the design variables (a, σ_W^2) as:

$$a = 0, \ \sigma_W^2 = (\sigma_X - \sqrt{P})^2. \tag{3.6.18}$$

Furthermore, since the distortion level D is not met with equality, using (3.6.18) we can characterize the maximum attainable distortion as $\bar{D}(P) = \mathbb{E}\left[(X - Y)^2\right] = \sigma_X^2 + (\sigma_X - \sqrt{P})^2$. This allows us to express σ_W^2 as $\sigma_W^2 = \bar{D}(P) - (1 - a)^2\sigma_X^2$.

Case III: Suppose $s_D > 0$ and $s_P > 0$, namely both (3.6.16) and (3.6.17) are active and holding with equality. We can observe from (3.6.16) that:

$$\sigma_W^2 = D - (1 - a)^2\sigma_X^2 \tag{3.6.19}$$

$$\sigma_Y^2 = a^2\sigma_X^2 + \sigma_W^2 = D + (2a - 1)\sigma_X^2 \tag{3.6.20}$$

and from (3.6.17) we obtain:

$$\left|\sigma_X - \sqrt{a^2\sigma_X^2 + \sigma_W^2}\right| = \sqrt{P} \tag{3.6.21}$$

Note that (3.6.21) behaves as follows:

- if $\sigma_X \geq \sqrt{a^2\sigma_X^2 + \sigma_W^2}$:

$$a = \frac{1}{2}\frac{\sigma_X^2 + (\sigma_X - \sqrt{P})^2 - D}{\sigma_X^2} \tag{3.6.22}$$

- if $\sigma_X < \sqrt{a^2\sigma_X^2 + \sigma_W^2}$:

$$a = \frac{1}{2}\frac{\sigma_X^2 + (\sigma_X + \sqrt{P})^2 - D}{\sigma_X^2} \tag{3.6.23}$$

Between (3.6.22) and (3.6.23), the former holds the minimum once substituted in (3.6.15) with (3.6.19). Therefore, the optimal design variables $(a, \sigma_W^2) = ((3.6.22), (3.6.19))$.

Regarding the boundaries between the operating regions, similarly to Theorem 3.1.1 the boundary between **Cases I-III** can be obtained by considering **Case I** and substi-

tuting its solution (3.6.7) in (3.6.17), while the boundary between **Cases II-III** can be obtained by substituting (3.6.18) in (3.6.16). The derived boundaries define the regions of the piecewise functions a and σ_W^2 in (3.1.8), where the term $\min\{D, \bar{D}(P)\}$ is justified by observing that $D \leq \bar{D}(P)$ in **Case I** and **Case III** while $D \geq \bar{D}(P)$ in **Case II**. This concludes the proof.

3.6.3 Proof of Theorem 3.1.2

The proof follows the same approach as the proof of Theorem 3.1.1, therefore we highlight only the differences between the two while maintaining the same logical structure.
The $\mathrm{D}_{\mathrm{GJS}}$ divergence (1.4.4) between two scalar Gaussian random variables $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y \sim \mathcal{N}(0, \sigma_Y^2)$ is defined as

$$\mathrm{D}_{\mathrm{GJS}}(p_X \| p_Y) \triangleq \frac{1}{4}\left[-\log\left(\frac{1}{4}\frac{(\sigma_X^2+\sigma_Y^2)^2}{\sigma_X^2\sigma_Y^2}\right) + \frac{1}{2}\frac{(\sigma_X^2+\sigma_Y^2)^2}{\sigma_X^2\sigma_Y^2} - 2\right]. \tag{3.6.24}$$

Once (3.6.24) is substituted in (3.1.1), we obtain the following convex optimization problem:

$$R(D,P) = \min_{a,\sigma_W^2} \frac{1}{2}\log\left(1 + a^2\frac{\sigma_X^2}{\sigma_W^2}\right) \tag{3.6.25}$$

$$\text{s.t.} \quad (1-a)^2\sigma_X^2 + \sigma_W^2 \leq D \tag{3.6.26}$$

$$(3.6.24) \leq P. \tag{3.6.27}$$

Similarly to Theorem 3.1.1, we obtain the closed-form solution of the optimization problem in (3.6.25)-(3.6.27) by leveraging the KKT conditions, considering specific operating regions of the problem and identifying the sets where these solutions apply. Let $s_D \geq 0$ and $s_P \geq 0$ be the Lagrangian multipliers associated with the distortion and perception constraint, respectively.
Case I: Let $s_D > 0$ and $s_P = 0$, namely, only (3.6.26) is active in the optimization problem. Then, the problem becomes equivalent to the classical RD, therefore, as in Theorem 3.1.1, the optimal design variables (a, σ_W^2) are equal to (3.6.7).
Case II: Suppose $s_D = 0$ and $s_P > 0$, namely only (3.6.27) holds with equality. As in Theorem 3.1.1, the inactive distortion constraint (3.6.26) implies that $R(D,P) = 0$ on this operating region. Moreover, we can design the design variables (a, σ_W^2) as:

$$a = 0, \ \sigma_W^2 = f_L(P). \tag{3.6.28}$$

Furthermore, since the distortion level D is not met with equality, using (3.6.28) we can characterize the maximum attainable distortion as $\bar{D}(P) = \mathbb{E}\left[(X-Y)^2\right] = \sigma_X^2 + f_L(P)$. This allows us express σ_W^2 as $\sigma_W^2 = \bar{D}(P) - (1-a)^2\sigma_X^2$.
Case III: Let $s_D > 0$ and $s_P > 0$, meaning that both (3.6.26) and (3.6.27) are active

and hold with equality. We can observe from (3.6.26) that

$$\sigma_Y^2 = a^2\sigma_X^2 + \sigma_W^2 = D + (2a-1)\sigma_X^2 \tag{3.6.29}$$

and from (3.6.27) we obtain

$$\begin{aligned}\log\left(\frac{(\sigma_X^2+\sigma_Y^2)^2}{\sigma_Y^2\sigma_X^2}\right) - \frac{(\sigma_X^2+\sigma_Y^2)^2}{2\sigma_Y^2\sigma_X^2} &= -(4P+2-\log(4)) \\ -\frac{(\sigma_X^2+\sigma_Y^2)^2}{2\sigma_Y^2\sigma_X^2}\cdot e^{-\frac{(\sigma_X^2+\sigma_Y^2)^2}{2\sigma_Y^2\sigma_X^2}} &= -2\cdot e^{-(4P+2)} \\ \frac{(\sigma_X^2+\sigma_Y^2)^2}{\sigma_Y^2\sigma_X^2} &= -2\mathcal{W}(-2\cdot e^{-(4P+2))}).\end{aligned}$$

As in Theorem 3.1.1, it is necessary to understand which branch of the Lambert W respects the continuity property of RDPF. For the sake of simplicity, we will directly use the negative branch $\mathcal{W}_{-1}$ which can be later proved to be the correct one. Defining $g(P) = -2\mathcal{W}_{-1}(-2\cdot e^{-(4P+2))})$, the equation expands two a second degree polynomial in σ_Y^2:

$$(\sigma_Y^2)^2 + \sigma_Y^2\sigma_X^2(2-g(P)) + (\sigma_X^2)^2 = 0.$$

The solutions of the polynomial are both positive for $P \in [0,+\infty)$ and we select the one minimizing σ_Y^2, thus obtaining:

$$\sigma_Y^2 = \frac{1}{2}\sigma_X^2\left[-(2-g(P)) - \sqrt{g(P)(g(P)-4)}\right].$$

Enforcing (3.6.29) and solving for a and σ_W^2 results in

$$a = \frac{1}{2}\left[1 - \frac{D}{\sigma_X^2} - \frac{1}{2}\left((2-g(P)) + \sqrt{g(P)(g(P)-4)}\right)\right] \tag{3.6.30}$$

$$\sigma_W^2 = D - (1-a)^2\sigma_X^2 \tag{3.6.31}$$

which once substituted in (3.6.25) give (3.1.9).

Regarding the boundaries between the operating regions, similarly to Theorem 3.1.1, the boundary between **Cases I-III** can be obtained by considering **Case I** and substituting its solution (3.6.7) in (3.6.27), while the boundary between **Cases II-III** can be obtained by substituting (3.6.28) in (3.6.26). The derived boundaries define the regions of the piecewise functions a and σ_W^2 in (3.1.10), where the term $\min\{D, \bar{D}(P)\}$ is justified by observing that $D \leq \bar{D}(P)$ in **Case I** and **Case III** while $D \geq \bar{D}(P)$ in **Case II**. This concludes the proof.

3.6.4 Proof of Theorem 3.1.3

The proof follows the same approach as the proof of Theorem 3.1.1, therefore we highlight only the differences between the two while maintaining the same logical structure.
The H^2 distance (1.4.7) between two scalar Gaussian random variables $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y \sim \mathcal{N}(0, \sigma_Y^2)$ is defined as:

$$\mathrm{H}^2(p_X, p_Y) \triangleq 1 - \sqrt{\frac{2\sigma_X \sigma_Y}{\sigma_X^2 + \sigma_Y^2}}. \tag{3.6.32}$$

Once (3.6.32) is substituted in (3.1.1), we obtain the following convex optimization problem:

$$R(D, P) = \min_{a, \sigma_W^2} \frac{1}{2} \log \left(1 + a^2 \frac{\sigma_X^2}{\sigma_W^2}\right) \tag{3.6.33}$$

$$\text{s.t.} \quad (1-a)^2 \sigma_X^2 + \sigma_W^2 \leq D \tag{3.6.34}$$

$$(3.6.32) \leq P. \tag{3.6.35}$$

Similarly to Theorem 3.1.1, we obtain the closed-form solution of the optimization problem in (3.6.33)-(3.6.35) by leveraging the KKT conditions, considering specific operating regions of the problem and identifying the sets where these solutions apply. Let $s_D \geq 0$ and $s_P \geq 0$ be the Lagrangian multipliers associated with the distortion and perception constraint, respectively.
Case I: Let $s_D > 0$ and $s_P = 0$, namely, only (3.6.34) is active in the optimization problem. Then, the problem becomes equivalent to the classical RD, therefore, as in Theorem 3.1.1, the optimal design variables (a, σ_W^2) are equal to (3.6.7).
Case II: Suppose $s_D = 0$ and $s_P > 0$, namely only (3.6.35) holds with equality. As in Theorem 3.1.1, the inactive distortion constraint (3.6.34) implies that $R(D, P) = 0$ on this operating region. Moreover, we can design the design variables (a, σ_W^2) as:

$$a = 0, \;\; \sigma_W^2 = \sigma_X^2 \left(1 - 2g(P)\right). \tag{3.6.36}$$

Furthermore, since the distortion level D is not met with equality, using (3.6.36) we can characterize the maximum attainable distortion as $\bar{D}(P) = \mathbb{E}\left[(X-Y)^2\right] = 2\sigma_X^2 g(P)$. This allows us, similarly to **Case I**, to rewrite σ_W^2 as $\sigma_W^2 = \bar{D}(P) - (1-a)^2\sigma_X^2$.
Case III: Let $s_D > 0$ and $s_P > 0$, meaning that both (3.6.34) and (3.6.35) are active and hold with equality. We can observe from (3.6.34) that

$$\sigma_Y^2 = a^2\sigma_X^2 + \sigma_W^2 = D + (2a-1)\sigma_X^2 \tag{3.6.37}$$

and from (3.6.35) we obtain

$$\sigma_Y^2 = \sigma_X^2 \left(\frac{1 \pm \sqrt{1-(1-P)^4}}{(1-P)^2} \right)^2. \tag{3.6.38}$$

Now, solving the equation (3.6.37) − (3.6.38) = 0 for a holds two feasible solutions, of which we select the one minimizing (3.6.33). Hence, the optimal design variables (a, σ_W^2) are the following:

$$a = g(P) - \frac{D}{2\sigma_X^2},\ \sigma_W^2 = D - (1-a)^2\sigma_X^2. \tag{3.6.39}$$

Regarding the boundaries between the operating regions, similarly to Theorem 3.1.1 the boundary between **Cases I-III** can be obtained by considering **Case I** and substituting its solution (3.6.7) in (3.6.35), while the boundary between **Cases II-III** can be obtained by substituting (3.6.36) in (3.6.34). The derived boundaries define the regions of the piecewise functions a and σ_W^2 in (3.1.12), where the term $\min\{D, \bar{D}(P)\}$ is justified by observing that $D \leq \bar{D}(P)$ in **Case I** and **Case III** while $D \geq \bar{D}(P)$ in **Case II**. This concludes the proof.

3.6.5 Proof of Theorem 3.2.1

To prove this result, we use some prior results of Grippo and Sciandrone in [50] in the study of the 2-block non-linear Gauss-Seidel method under convex constraints.

Let $f(\mathbf{D}, \mathbf{P}) = \sum_{i=1}^N R_i^G(D_i, P_i)$ and let the set $\mathcal{L}_{(D,P)}$ be the set of level curves defined as follows:

$$\mathcal{L} = \left\{ (\mathbf{D}, \mathbf{P}) \in \mathcal{C}_{(D,P)} : f(\mathbf{D}, \mathbf{P}) \leq f(\mathbf{D}^{(0)}, \mathbf{P}^{(0)}) \right\} \tag{3.6.40}$$

where $\mathcal{C}_{(D,P)} = \{(\mathbf{D}, \mathbf{P}) \in \mathbb{R}^{2N} : \sum_{i=1}^N g(D_i) \leq D \wedge \sum_{i=1}^N r(P_i) \leq P\}$. Since the pair (D, P) is assumed to be finite, we remark that $\mathcal{C}_{(D,P)}$ is a compact set, and, as $f(\cdot)$ is a non-negative, continuous function in $\mathcal{C}_{(D,P)}$, the curve level set $\mathcal{L} = f^{-1}([0, f(\mathbf{D}^{(0)}, \mathbf{P}^{(0)})]) \cap \mathcal{C}_{(D,P)}$ is also a compact set. This result together with the convexity of $f(\cdot)$ satisfy the assumption of [50, Proposition 6], which guarantee that every limit point $(\mathbf{D}^*, \mathbf{P}^*)$ of the sequence $\{(\mathbf{D}^{(n)}, \mathbf{P}^{(n)}) :\ n = 1, 2, \ldots\}$ is a global minimizer of (3.2.3). This concludes the proof.

3.6.6 Proof of Theorem 3.2.2

Let $L(\mathbf{D}, \mathbf{P})$ be the Lagrangian function defined in (3.2.7) for fixed Lagrangian multipliers s. After T iterations, Algorithm 3 produces the sequences $\{\mathbf{D}^{(n)}\}_{n=0,\ldots,T}$ and $\{\mathbf{P}^{(n)}\}_{n=0,\ldots,T}$ applying the alternating minimization scheme $\mathbf{P}^{(n-1)} \to \mathbf{D}^{(n)} \to \mathbf{P}^{(n)}$ obtained by the solutions of the optimization problems (3.2.8) and (3.2.9). Therefore, by

construction, $L(\cdot,\cdot)$ is non-increasing along the produced sequence, i.e.,

$$L(\mathbf{D}^{(n)},\mathbf{P}^{(n)}) \geq L(\mathbf{D}^{(n+1)},\mathbf{P}^{(n)}) \geq L(\mathbf{D}^{(n+1)},\mathbf{P}^{(n+1)}). \tag{3.6.41}$$

Since $R_i^G(\cdot,\cdot)$ is continuous with Lipschitz continuous gradients in both arguments, there exists $K \in \mathbb{R}^+$ such that the Lagrangian $L(\mathbf{D},\mathbf{P})$ is continuous and its gradients $\nabla_{\mathbf{D}} L(\cdot,\mathbf{P})$ and $\nabla_{\mathbf{P}} f(\mathbf{D},\cdot)$ are K-Lipschitz continuous. Therefore, $\forall\ \mathbf{D}_1,\mathbf{D}_2,\mathbf{P}_1,\mathbf{P}_2 \in \mathbb{R}^d$, it holds

$$L(\mathbf{D}_2,\mathbf{P}_1) \leq L(\mathbf{D}_1,\mathbf{P}_1) + \nabla_{\mathbf{D}} L(\mathbf{D}_1,\mathbf{P}_1)^T(\mathbf{D}_2-\mathbf{D}_1) + \frac{K}{2}\|\mathbf{D}_2-\mathbf{D}_1\|_2^2. \tag{3.6.42}$$

Considering iteration n, for fixed $\mathbf{P}^{(n)}$ we define

$$\mathbf{D}^{(n+1)} = \arg\min_{\mathbf{D}}\ (3.2.8), \qquad \tilde{\mathbf{D}} = \mathbf{D}^{(n)} - \nabla_{\mathbf{D}} L(\mathbf{D}^{(n)},\mathbf{P}^{(n)})$$

for which, by construction, it holds

$$L(\mathbf{D}^{(n)},\mathbf{P}^{(n)}) \geq L(\tilde{\mathbf{D}},\mathbf{P}^{(n)}) \geq L(\mathbf{D}^{(n+1)},\mathbf{P}^{(n)}). \tag{3.6.43}$$

From (3.6.41), (3.6.42), and (3.6.43) we can derive

$$\begin{aligned} L(\mathbf{D}^{(n)},\mathbf{P}^{(n)}) - L(\mathbf{D}^{(n+1)},\mathbf{P}^{(n+1)}) &\geq L(\mathbf{D}^{(n)},\mathbf{P}^{(n)}) - L(\mathbf{D}^{(n+1)},\mathbf{P}^{(n)}) \\ &\geq L(\mathbf{D}^{(n)},\mathbf{P}^{(n)}) - L(\tilde{\mathbf{D}},\mathbf{P}^{(n)}) \\ &\geq -\nabla_{\mathbf{D}} f(\mathbf{D}^{(n)},\mathbf{P}^{(n)})^T(\tilde{\mathbf{D}}-\mathbf{D}^{(n)}) - \frac{K}{2}\|\tilde{\mathbf{D}}-\mathbf{D}^{(n)}\|_2^2 \\ &\geq \frac{1}{2K}\|\nabla_{\mathbf{D}} L(\mathbf{D}^{(n)},\mathbf{P}^{(n)})\|_2^2. \end{aligned} \tag{3.6.44}$$

Furthermore, considering the complete gradient operator $\nabla = (\nabla_D, \nabla_P)$, the identity $\|\nabla L(\mathbf{D}^{(n)},\mathbf{P}^{(n)})\|_2 = \|\nabla_{\mathbf{D}} L(\mathbf{D}^{(n)},\mathbf{P}^{(n)})\|_2$ holds, since $\|\nabla_P L(\mathbf{D}^{(n)},\mathbf{P}^{(n)})\|_2 = 0$ due to $\mathbf{P}^{(n)}$ being the minimizer for fixed $\mathbf{D}^{(n)}$ in the alternating minimization scheme. Therefore, we have that

$$L(\mathbf{D}^{(n)},\mathbf{P}^{(n)}) - L(\mathbf{D}^{(n+1)},\mathbf{P}^{(n+1)}) \geq \frac{1}{2K}\|\nabla L(\mathbf{D}^{(n)},\mathbf{P}^{(n)})\|_2^2. \tag{3.6.45}$$

Now, let $\mathbf{z}^{(n)} = (D_1^{(n)},\ldots,D_N^{(n)},P_1^{(n)},\ldots,P_N^{(n)})^T$ be the column vector resulting from the stacking $\mathbf{D}$ and $\mathbf{P}$ and let $L(\mathbf{z}^{(n)}) = L(\mathbf{D}^{(n)},\mathbf{P}^{(n)})$. Moreover, let $\mathcal{Z}^*$ be the set of minimizers of $L(\cdot)$ and, similarly to [63], let the maximal distance $M(\hat{\mathbf{z}})$ from the set $\mathcal{Z}^*$ be defined as

$$M(\hat{\mathbf{z}}) = \max_{\mathbf{z}\in\mathbb{R}^{2N}} \max_{\mathbf{z}^*\in\mathcal{Z}^*} \{\|\mathbf{z}-\mathbf{z}^*\|_2 : L(\mathbf{z}) \leq L(\hat{\mathbf{z}})\}.$$

From the definition $M(\hat{\mathbf{z}})$, we have that $L(\hat{\mathbf{z}}) \geq L(\tilde{\mathbf{z}})$ implies $M(\hat{\mathbf{z}}) \geq M(\tilde{\mathbf{z}})$. Then, from

the convexity of $L(\cdot)$, for all $z^* \in \mathcal{Z}^*$, we can derive

$$\begin{aligned} L(\mathbf{z}^{(n)}) - L(\mathbf{z}^*) &\leq \nabla L(\mathbf{z}^{(n)})^T(\mathbf{z}^{(n)} - \mathbf{z}^*) \\ &\leq \|\nabla L(\mathbf{z}^{(n)})\| \cdot \|\mathbf{z}^{(n)} - \mathbf{z}^*\| \\ &\leq \|\nabla L(\mathbf{z}^{(n)})\| M(\mathbf{z}^{(n)}) \\ &\leq \|\nabla L(\mathbf{z}^{(n)})\| M(\mathbf{z}^{(0)}). \end{aligned} \tag{3.6.46}$$

Furthermore, as a result of the *descent lemma* [60], the following holds

$$L(\mathbf{z}^{(0)}) - L(\mathbf{z}^*) \leq \frac{K}{2}\|\mathbf{z}^{(0)} - \mathbf{z}^*\|^2 \leq \frac{K}{2}M^2(\mathbf{z}^{(0)}). \tag{3.6.47}$$

Leveraging (3.6.41), (3.6.45), (3.6.46) and (3.6.47), and defining $A_n = L(\mathbf{z}^{(n)}) - L(\mathbf{z}^*)$ and $\gamma = \frac{1}{2KM^2(\mathbf{z}^{(0)})}$, we obtain

$$A_n \overset{(3.6.41)}{\geq} A_{n+1}, \qquad A_n - A_{n+1} \overset{(3.6.45)+(3.6.46)}{\geq} \gamma A_n^2, \qquad A_0 \overset{(3.6.46)}{\leq} \frac{1}{4\gamma}. \tag{3.6.48}$$

Applying the relationships obtained in (3.6.48), we derive

$$\frac{1}{A_n} - \frac{1}{A_{n-1}} = \frac{A_{n-1} - A_n}{A_n A_{n-1}} \geq \gamma \frac{A_{n-1}^2}{A_n A_{n-1}} = \gamma \frac{A_{n-1}}{A_n} \geq \gamma.$$

By summing for $n = 1, \ldots, T$, we have

$$\frac{1}{A_T} - \frac{1}{A_0} = \sum_{n=1}^{T}\left[\frac{1}{A_n} - \frac{1}{A_{n-1}}\right] \geq \sum_{n=1}^{T}\gamma = T\gamma \implies \frac{1}{A_T} \geq \frac{1}{A_0} + T\gamma \geq (T+4)\gamma \geq T\gamma$$

which further implies,

$$L(\mathbf{z}^{(T)}) - L(\mathbf{z}^*) \leq \frac{2K \cdot M^2(\mathbf{z}^{(0)})}{T} \leq \frac{2KC}{T} \tag{3.6.49}$$

with $M^2(\mathbf{z}^{(0)}) \leq C < \infty$. This concludes the proof.

3.6.7 Proof of Theorem 3.2.3

Let $L(\mathbf{D}, s_D)$ be the objective function of the unconstrained optimization problem defined in (3.2.8). Imposing the *stationarity condition* on $L(\mathbf{D}, s_D)$ we obtain

$$\left.\frac{\partial L(\mathbf{D}, s_D)}{\partial D_i}\right|_{D_i^*} = \left.\frac{\partial R_i^G(D_i, P_i)}{\partial D_i}\right|_{D_i^*} + s_D = 0 \tag{3.6.50}$$

with

$$\frac{\partial R_i^G(D_i,P_i)}{\partial D_i} = \begin{cases} -\frac{1}{2D_i} & (D_i,P_i)\in\mathcal{S}^c \\ -\frac{D_i-P_i-2\lambda_{\Sigma_X,i}+2\sqrt{P_i\lambda_{\Sigma_X,i}}}{(D_i-P_i)(D_i-(2\sqrt{\lambda_{\Sigma_X,i}}-\sqrt{P_i})^2} & (D_i,P_i)\in\mathcal{S} \end{cases}$$

where $\mathcal{S}$ is defined in Theorem 3.1.1. We can investigate the optimal solution of (3.6.50) in the two sets $\mathcal{S}$ and $\mathcal{S}^c$ as follows:

- for $(D_i^*,P_i)\in\mathcal{S}$ we have that

$$-\frac{1}{2D_i^*}+s_D=0 \implies D_i^*=\frac{1}{2s_D}$$

- for $(D_i^*,P_i)\in\mathcal{S}^c$ we have that

$$-\frac{D_i-P_i-2\lambda_{\Sigma_X,i}+2\sqrt{P_i\lambda_{\Sigma_X,i}}}{(D_i-P_i)(D_i-(2\sqrt{\lambda_{\Sigma_X,i}}-\sqrt{P_i})^2}+s_D=0$$

  which admits the following two solutions:

$$D_{i,1}^*=P_i+2\sigma_X(\sigma_X-\sqrt{P_i})+\left(\frac{1}{2s_D}-\sqrt{4\sigma_X^2(\sigma_X-\sqrt{P_i})^2+\frac{1}{4s_D^2}}\right)$$
$$D_{i,2}^*=P_i+2\sigma_X(\sigma_X-\sqrt{P_i})+\left(\frac{1}{2s_D}+\sqrt{4\sigma_X^2(\sigma_X-\sqrt{P_i})^2+\frac{1}{4s_D^2}}\right).$$

  We select the solution $D_i^*=D_{i,1}^*$ since $D_{i,1}^*<D_{i,2}^*$, minimizing $D=\sum_{i=1}^N D_i^*$, and this guarantees continuity of the function $D_i^*(s_D)$ on $\mathcal{S}\cup\mathcal{S}^c=\mathbb{R}^{2,+}$. This completes the proof.

3.6.8 Proof of Theorem 3.2.4

Let $L(\mathbf{P},s_P)$ be the objective function of the unconstrained optimization problem defined in (3.2.9). We can characterize the set of optimal solutions $\{\mathbf{P}^*\}$ imposing the *stationarity condition* on $L(\mathbf{P},s_P)$, thus obtaining

$$\left.\frac{\partial L(\mathbf{P},s_P)}{\partial P_i}\right|_{P_i^*}=\left.\frac{\partial R_i^G(D_i,P_i)}{\partial P_i}\right|_{P_i^*}+s_P=0.$$

We notice that set $\{\mathbf{P}^*\}$ and the set of the zeros of $T(\cdot)$ coincide. This concludes the proof.

3.6.9 Proof of Corollary 3.2.1

Since $R_i^G(D_i,P_i)$ is convex and differentiable in both arguments, $\frac{\partial R_i^G(D_i,P_i)}{\partial P_i}$ is necessarily continuous and non-decreasing on $\mathbb{R}_0^+$. From the definition of the set $\mathcal{S}$, we can charac-

terize $b_i \geq 0$ such that $\mathcal{S} = [0, b_i]$ and $\mathcal{S}^c = (b_i, +\infty)$. Due to continuity, $\frac{\partial R_i^G(D_i,b_i)}{\partial P_i} = 0$, thus implying $T_i(b_i) \geq 0$. On the other hand, $\lim_{P_i \to 0} \frac{\partial R_i^G(D_i,b_i)}{\partial P_i} = -\infty$ implies the existence of $a_i \in \mathcal{S}$ such that $T_i(a_i) \leq 0$. Hence, since on the extreme of the set $[a_i, b_i] \subseteq \mathcal{S}$ the continuous function T_i has opposite signs, by the Intermediate Value Theorem [48, Theorem 1.11] there exists at least one root in the set. This concludes the proof.

3.6.10 Proof of Proposition 3.4.3

For jointly Gaussian random variables, the squared Wasserstein-2 distance $\mathrm{W}_2^2(p_X, p_Y)$ has the following form [46]:

$$\mathrm{W}_2^2(p_X, p_Y) = \mathrm{Tr}\left[\Sigma_X + \Sigma_Y - 2\left(\Sigma_Y^{\frac{1}{2}}\Sigma_X\Sigma_Y^{\frac{1}{2}}\right)^{\frac{1}{2}}\right]. \tag{3.6.51}$$

By performing eigenvalue decomposition on the covariance matrices Σ_X and Σ_Y and writing them into their canonical form, i.e., $\Sigma_X = P\lambda_{\Sigma_X}P^T$ and $\Sigma_Y = Q\lambda_{\Sigma_Y}Q^T$, (3.6.51) can be expressed as follows:

$$\mathrm{W}_2^2(p_X, p_Y) = \mathrm{Tr}\left[\lambda_{\Sigma_X}\right] + \mathrm{Tr}\left[\lambda_{\Sigma_Y}\right] - 2\,\mathrm{Tr}\left[\left(Q\lambda_{\Sigma_Y}^{\frac{1}{2}}Q^TP\lambda_{\Sigma_X}P^TQ\lambda_{\Sigma_Y}^{\frac{1}{2}}Q^T\right)^{\frac{1}{2}}\right]. \tag{3.6.52}$$

Our objective is the minimization of (3.6.52) while maintaining the set eigenvalues λ_{Σ_X} and λ_{Σ_Y}. Therefore, our optimization variables will be the orthonormal matrices [29, Section 2.1] P and Q, whose columns are the set of eigenvectors of Σ_X and Σ_Y, respectively. Since the first two terms of (3.6.52) do not depend of P and Q, the minimization problem can be rewritten as a maximization problem as follows:

$$\max_{P,Q\in\mathcal{U}} \mathrm{Tr}\left[\left(Q\lambda_{\Sigma_Y}^{\frac{1}{2}}Q^TP\lambda_{\Sigma_X}P^TQ\lambda_{\Sigma_Y}^{\frac{1}{2}}Q^T\right)^{\frac{1}{2}}\right] \tag{3.6.53}$$

where $\mathcal{U}$ is the set of orthonormal matrices on $\mathbb{R}^{N\times N}$. The optimization objective can be expressed as:

$$\mathrm{Tr}\left[\left(Q\lambda_{\Sigma_Y}^{\frac{1}{2}}Q^TP\lambda_{\Sigma_X}P^TQ\lambda_{\Sigma_Y}^{\frac{1}{2}}Q^T\right)^{\frac{1}{2}}\right] = \|P\lambda_{\Sigma_X}^{\frac{1}{2}}P^TQ\lambda_{\Sigma_Y}^{\frac{1}{2}}Q^T\|_1 \overset{(a)}{=} \|\lambda_{\Sigma_X}^{\frac{1}{2}}P^TQ\lambda_{\Sigma_Y}^{\frac{1}{2}}\|_1$$

where (a) is justified by the unitary invariance property of the norm. By defining $S \triangleq \lambda_{\Sigma_X}^{\frac{1}{2}}P^TQ\lambda_{\Sigma_Y}^{\frac{1}{2}}$, we obtain

$$\|S\|_1 \overset{(b)}{=} \mathrm{Tr}(|S|) \overset{(c)}{=} \sup_{U\in\mathcal{U}} \mathrm{Tr}(US)$$

where (b) stems from the definition of Schatten-1 norm and by defining $|S| = \sqrt{S^TS}$ with $|S|$ being the unique positive semidefinite root of S^TS; (c) stems from the fact that using polar decomposition [29, Theorem 7.3.1], there exists a unique matrix $U \in \mathcal{U}$ such

that $|S| = US$ and the fact that $|S|$ maximizes $\mathrm{Tr}(US)$ over the set of unitary matrices $U \in \mathcal{U}$ [64, Proposition 5]. Consequentially, we can reformulate the optimization problem as follows:

$$\max_{P,Q\in\mathcal{U}} \sup_{U\in\mathcal{U}} \mathrm{Tr}\left(U\lambda_{\Sigma_X}^{\frac{1}{2}} P^T Q \lambda_{\Sigma_Y}^{\frac{1}{2}}\right). \tag{3.6.54}$$

Now, let us define the function $f : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ as

$$f(U,V) \triangleq \mathrm{Tr}(U\lambda_{\Sigma_X}^{\frac{1}{2}} V^T \lambda_{\Sigma_Y}^{\frac{1}{2}}).$$

We can verify that f satisfies the properties of an inner product on $\mathcal{U}$ [29, Definition 5.1.3], therefore using Cauchy–Schwarz inequality [29, Theorem 5.1.4], we obtain

$$\sup_{U\in\mathcal{U}} \mathrm{Tr}\left(U\lambda_{\Sigma_X}^{\frac{1}{2}} P^T Q \lambda_{\Sigma_Y}^{\frac{1}{2}}\right) = \sup_{U\in\mathcal{U}} f(U, Q^T P) \leq f(Q^T P, Q^T P)$$

Using the above inequality in (3.6.54) results into the following:

$$\begin{aligned}(3.6.54) &\overset{(d)}{\leq} \max_{P,Q\in\mathcal{U}} \mathrm{Tr}\left(Q^T P \lambda_{\Sigma_X}^{\frac{1}{2}} P^T Q \lambda_{\Sigma_Y}^{\frac{1}{2}}\right) = \max_{R=Q^T P\in\mathcal{U}} \mathrm{Tr}\left(R\lambda_{\Sigma_X}^{\frac{1}{2}} R^T \lambda_{\Sigma_Y}^{\frac{1}{2}}\right) \\ &= \max_{R=Q^T P\in\mathcal{U}} \langle \lambda_{\Sigma_Y}^{\frac{1}{2}}, R\lambda_{\Sigma_X}^{\frac{1}{2}} R^T\rangle_F \overset{(e)}{\leq} \langle \lambda_{\Sigma_Y}^{\frac{1}{2}}, \lambda_{\Sigma_X}^{\frac{1}{2}}\rangle_F\end{aligned}$$

where (d) and (e) hold with equality for $Q^T P = I$, implying that the maximum of (3.6.53), and consequentially the minimum of (3.6.52), is attained by commuting matrices Σ_X and Σ_Y. Therefore,

$$\mathrm{W}_2^2(p_X, p_Y) \geq \mathrm{Tr}\left[\lambda_{\Sigma_X}\right] + \mathrm{Tr}\left[\lambda_{\Sigma_Y}\right] - 2\,\mathrm{Tr}\left[\lambda_{\Sigma_Y}^{\frac{1}{2}} \lambda_{\Sigma_X}^{\frac{1}{2}}\right] = \sum_{i=1}^{N} \mathrm{W}_2^2(p_{X_i}, p_{Y_i}).$$

This concludes the proof.

3.6.11 Proof of Proposition 3.4.4

For jointly Gaussian random variables, the squared $\mathrm{H}^2(p_X, p_Y)$ has the following form:

$$\mathrm{H}^2(p_X, p_Y) = 2\left(1 - 2^{\frac{N}{2}} \frac{|\Sigma_X \Sigma_Y|^{\frac{1}{4}}}{|\Sigma_X + \Sigma_Y|^{\frac{1}{2}}}\right). \tag{3.6.55}$$

By performing eigenvalue decomposition on the covariance matrices Σ_X and Σ_Y and writing them into their canonical form, i.e., $\Sigma_X = P\lambda_{\Sigma_X} P^T$ and $\Sigma_Y = Q\lambda_{\Sigma_Y} Q^T$, (3.6.55) can be expressed as follows:

$$\mathrm{H}^2(p_X, p_Y) = 2\left(1 - 2^{\frac{N}{2}} \frac{|P\lambda_{\Sigma_X} P^T Q \lambda_{\Sigma_Y} Q^T|^{\frac{1}{4}}}{|P\lambda_{\Sigma_X} P^T + Q\lambda_{\Sigma_Y} Q^T|^{\frac{1}{2}}}\right) = 2\left(1 - 2^{\frac{N}{2}} \frac{|\lambda_{\Sigma_X}\lambda_{\Sigma_Y}|^{\frac{1}{4}}}{|\lambda_{\Sigma_X} + R\lambda_{\Sigma_Y} R^T|^{\frac{1}{2}}}\right) \tag{3.6.56}$$

with $R = P^T Q \in \mathcal{U}$, where $\mathcal{U}$ is the set of orthonormal matrices [29, Section 2.1]. Our objective is the minimization of (3.6.56) while maintaining the set eigenvalues λ_{Σ_X} and λ_{Σ_Y}. Therefore, our optimization variables will be the orthonormal matrix R. Since only the denominator of the second term in (3.6.56) depends on R, finding the minimizing R is equivalent to

$$\arg\min_{R\in\mathcal{U}} |I + \lambda_{\Sigma_X}^{-\frac{1}{2}} R \lambda_{\Sigma_Y} R^T \lambda_{\Sigma_Y}^{-\frac{1}{2}}|$$

which is itself equivalent to

$$\arg\min_{R\in\mathcal{U}} \log\left(|I + S^T S|\right) \tag{3.6.57}$$

where $S = \lambda_{\Sigma_X}^{-\frac{1}{2}} R \lambda_{\Sigma_Y}^{\frac{1}{2}}$. Now, expanding (3.6.57) we obtain

$$\log\left(|I + S^T S|\right) = \sum_{i=1}^{N} \log(\lambda[I + S^T S]) \overset{(a)}{=} \sum_{i=1}^{N} \log(1 + \lambda_i[S^T S]) \overset{(b)}{=} \sum_{i=1}^{N} \log(1 + s_{S,i}^2) \tag{3.6.58}$$

where (a) follows from the fact that for any diagonalizable matrix M the eigenvalues $\lambda_{I+M} = 1 + \lambda_M$ and (b) follows from the definition of singular values of a matrix. In addition, we can also verify the following majorization inequality:

$$\begin{aligned}
\log\left(s\left[\lambda_{\Sigma_X}^{-\frac{1}{2}} R \lambda_{\Sigma_Y}^{\frac{1}{2}}\right]\right) &\overset{(c)}{\succ} \log\left(s^{\downarrow}\left[\lambda_{\Sigma_X}^{-\frac{1}{2}} R\right]\right) + \log\left(\lambda^{\uparrow}\left[\Sigma_X^{\frac{1}{2}}\right]\right) \\
&\overset{(d)}{\succ} \log\left(\lambda^{\downarrow}\left[\Sigma_X^{-\frac{1}{2}}\right]\right) + \log\left(s^{\uparrow}[R]\right) + \log\left(\lambda^{\uparrow}\left[\Sigma_X^{\frac{1}{2}}\right]\right) \\
&\overset{(e)}{=} \log\left(\frac{1}{\lambda^{\uparrow}\left[\Sigma_X^{\frac{1}{2}}\right]}\right) + \log\left(\lambda^{\uparrow}\left[\Sigma_Y^{\frac{1}{2}}\right]\right)
\end{aligned} \tag{3.6.59}$$

where (c) and (d) hold from Lidskii's Theorem [28, Theorem 3.4.6] and hold with equality for commuting matrices λ_{Σ_X}, λ_{Σ_Y} and R with the correct singular value ordering, while (e) follows from R being orthonormal, hence $s_{R,i} = 1\ \forall i \in 1 : N$. Now, let $f(\cdot)$ to be the function defined as $f(x) \triangleq \log(1 + x^2)$. Although f is not convex, we can readily show that the composite function $t \to f(e^t)$ is convex and increasing on $\mathbb{R}$. Thus, we can invoke Weyl's majorization theorem, which states:

$$\log(x) \succ_w \log(y) \implies f(x) \succ_w f(y). \tag{3.6.60}$$

Therefore, we obtain

$$(3.6.57) = \sum_{i=1}^{N} f\left(s_i\left[\lambda_{\Sigma_X}^{-\frac{1}{2}} R \lambda_{\Sigma_Y}^{\frac{1}{2}}\right]\right) \overset{(3.6.59)+(3.6.60)}{\geq} \sum_{i=1}^{N} f\left(\frac{\sqrt{\lambda_{\Sigma_Y,i}}^{\uparrow}}{\sqrt{\lambda_{\Sigma_X,i}}^{\uparrow}}\right)$$

where the lower bound is achievable iff the matrix R induces the proper ordering between λ_{Σ_X} and λ_{Σ_Y}, including the case where Σ_X and Σ_Y are commuting matrices with eigenvalues in increasing (or decreasing) order. Therefore, assuming that both Σ_X and Σ_Y have the same eigenvalues order, (3.6.56) is lower bounded by

$$\mathrm{H}^2(p_X, p_Y) \geq 2\left(1 - \prod_{i=1}^{N} \sqrt{\frac{2\sqrt{\lambda_{\Sigma_X,i}}\sqrt{\lambda_{\Sigma_Y,i}}}{\lambda_{\Sigma_X,i} + \lambda_{\Sigma_Y,i}}}\right) = 2\left(1 - \prod_{i=1}^{N}\left(1 - \frac{\mathrm{H}^2(p_{X_i}, p_{Y_i})}{2}\right)\right)$$

where $\mathrm{H}^2(p_{X_i}, p_{Y_i})$ is the squared Hellinger distance between the marginals $X_i \sim \mathcal{N}(0, \lambda_{\Sigma_X,i})$ and $Y_i \sim \mathcal{N}(0, \lambda_{\Sigma_Y,i})$. This concludes the proof.

Chapter 4

Gaussian Processes RDPF

In this chapter, we build on the results established in Chapter 3 for finite dimensional Gaussian measures by extending the RDPF framework to sources modeled as Gaussian processes (GPs). Specifically, we study the case of GP-RDPF over a measure space Ω, under Mean Squared Error (MSE) distortion and the squared Wasserstein-2 distance as the perception metric. To this end, we show that the optimal reconstruction of the source is itself a GP, designed such that its covariance operator shares the same set of eigenvectors with the source. The common set of eigenvalues allows the formulation of the RDPF problem as a function of the Karhunen–Loève (KL) transform coefficients of the involved GPs, similar to the classical RDF for GPs [65, 66]. Noticing the similarities with the finite-dimensional Gaussian RDPF, we formulate an analytical upper bound to GP-RDPF able to recover the optimal solution in the *"perfect realism"* regime. Moreover, focusing on the specific case where the source is a stationary GP and Ω is the interval $[0, T]$ equipped with the Lebesgue measure, we characterize a tight upper bound on the rate and distortion levels for a fixed perceptual level and for $T \to \infty$ as a function of the power spectral density of the source process. We validate our theoretical results through illustrative numerical simulations.

4.1 Preliminaries

4.1.1 Gaussian Processes

A GP X is a collection of real random variables indexed by an index set $\mathcal{T}$, such that for any finite subset $\mathcal{T}' \subset \mathcal{T}$, the collection $\{f(t)\}_{t\in\mathcal{T}'}$ has a joint Gaussian distribution. A GP is parameterized by $m(t) = \mathbb{E}[X(t)]$ and $k(t,s) = \mathbb{E}[(X(t) - m(t))(X(s) - m(s))]$, where $m : \mathcal{T} \to \mathbb{R}$ and $k(\cdot,\cdot) : \mathcal{T}^2 \to \mathbb{R}$ denote the *mean* and *covariance* functions, respectively. From its definition it follows that k is a positive semidefinite symmetric function. For a strictly positive k, the associated GP $f \sim \mathcal{GP}(m, k)$ is referred to as *non-degenerate*. Furthermore, in the case where $\mathcal{T}$ is the Euclidean space $\mathbb{R}^d$, a GP is said to be *stationary* if the covariance function $k(t,s) = k(\tau)$ is a function of the difference vector $\tau = t - s$. In

this case, the GP can be alternatively characterized by its power spectral density $S(f)$, i.e., the Fourier transform of its covariance function $k(\tau)$, see, e.g., [65, Chapter 8].

4.1.2 Separable Hilbert spaces and L^2-spaces

A Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is said *separable* if there exists a countable orthonormal set of vectors $\mathcal{B} = \{e_i\}_{i=1}^{\infty}$, i.e., a Schauder basis, such that the closed linear hull of $\mathcal{B}$ spans $\mathcal{H}$ [67]. For a separable Hilbert space, the trace $\mathrm{Tr}(\cdot)$ of a bounded linear operator T on $\mathcal{H}$ is defined as $\mathrm{Tr}(T) \triangleq \sum_{\mathcal{B}'} \langle Te_i, e_i\rangle$, independently of the chosen basis $\mathcal{B}'$. Furthermore, a bounded operator T is said *trace class* if and only if $\mathrm{Tr}[(T^*T)^{\frac{1}{2}}] < \infty$, where T^* indicates the adjoint operator of T. Let $\Omega = (\mathcal{X}, \Sigma_{\mathcal{X}}, \mu)$ be a measure space and let $L^2(\Omega)$ be the space of L^2-integrable functions from $\mathcal{X}$ to $\mathbb{R}$. Equipping $L^2(\Omega)$ with the inner product $\langle f, g\rangle = \int_\Omega f(x)g(x)\mu(dx)$, for $f, g \in L^2(\Omega)$, it becomes a Hilbert space[1]. Throughout this chapter, we assume that the underlying measure space $\mathcal{X}$ is such that $(L^2(\Omega), \langle\cdot,\cdot\rangle)$ is separable.

4.1.3 Covariance Operators and Mercer's Theorem

Given a covariance function $k \in L^2(\Omega \times \Omega)$, we can define the associated Hilbert–Schmidt (HS) integral operator $K : L^2(\Omega) \to L^2(\Omega)$ as

$$[K\phi](t) = \int_\Omega k(x, s)\phi(x)\mu(dx) \qquad t \in \mathcal{X}.$$

Then, K is a self-adjoint, compact, positive, and trace-class operator, and the space of such covariance operators is a convex space. Moreover, since the mapping $k \to K$ is an isometric isomorphism from $L^2(\Omega \times \Omega)$ to the space of Hilbert-Schmidt operators in $L^2(\Omega)$, we use interchangeably the notations $\mathcal{GP}(m, k)$ and $\mathcal{GP}(m, K)$.
The constructed HS operator K also satisfies the conditions of Mercer's Theorem [68]. Let $\{\phi_i\}_{i=1}^{\infty}$ and $\{\lambda_i\}_{i=1}^{\infty}$ be the sets of eigenfunctions and associated eigenvalues of K. Then, the covariance function k can be represented by the expansion

$$k(t, s) = \sum_{i=1}^{\infty} \lambda_i \phi_i(t)\phi_i(s) \qquad (t, s) \in \mathcal{X}^2,$$

with absolute and uniform convergence on $\Omega \times \Omega$, implying that K can be expressed as

$$[K\psi](t) = \sum_{i=1}^{\infty} \lambda_i \langle\psi, \phi_i\rangle\phi_i(t) \qquad t \in \mathcal{X},$$

[1] Formally, $\langle\cdot,\cdot\rangle$ is a semi-inner product. However, introducing the equivalence classes of functions that differ only in μ-negligible sets, i.e., $f = g \iff f - g = 0\ \mu - a.e.$, $\langle\cdot,\cdot\rangle$ becomes an inner product.

i.e., K is a diagonizable operator. Furthermore, a zero-mean stochastic process X can be represented as

$$X(t) = \sum_{i=1}^{\infty} X_i \phi_i(t) \qquad t \in \mathcal{X}$$

where $X_i = \langle X, \phi_i \rangle$ is a random variable with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i X_j] = \lambda_i \delta_{i,j}$. Additionally, if $X \sim \mathcal{GP}(0, K)$, then X_i is Gaussian distributed and, consequently, $\forall (i,j)\ X_i \perp X_j$. Therefore, given a suitable base of $L^2(\Omega)$, GP X can be represented by a countable set $\{X_i\}_{i=1}^{\infty}$ of independent Gaussians. This representation is commonly referred to as the KL transform [69] and we refer to $\{X_i\}_{i=1}^{\infty}$ as the KL coefficients of the process X.

4.1.4 Squared Wasserstein-2 distance for GP

Squared Wasserstein-2 distance was originally introduced in [46] as a specific instance of the optimal transport problem (see, e.g., [45, Chapter 7]). In particular, squared Wasserstein-2 distance is defined as follows

$$\mathrm{W}_2^2(P_X, P_Y) \triangleq \min_{\Pi(P_X, P_Y)} \mathbb{E}\left[\|X - Y\|^2\right] \tag{4.1.1}$$

where $\Pi(P_X, P_Y)$ is the set of all joint distributions $P_{X,Y}$ with marginals P_X and P_Y. Following [70, Definition 1], the Wasserstein-2 distance can be extended to GPs. Let $X \sim \mathcal{GP}(m_X, k_X)$ and $Y \sim \mathcal{GP}(m_Y, k_Y)$ with $m_X, m_Y \in L^2(\Omega)$ and $k_X, k_Y \in L^2(\Omega \times \Omega)$, then the squared Wasserstein-2 distance between GPs X and Y is given by

$$\mathrm{W}_2^2(X, Y) = d^2(m_X, m_Y) + \mathrm{Tr}\left[K_X + K_Y - 2\left(K_X^{\frac{1}{2}} K_Y K_X^{\frac{1}{2}}\right)^{\frac{1}{2}}\right]$$

where $d^2(\cdot,\cdot)$ is the canonical distance in $L^2(\Omega)$ and K_X, K_Y are the HS operators associated with k_X and k_Y, respectively.

4.2 Main Results

We start this section by providing a formal characterization of the RDPF problem for sources modeled as GPs.

Theorem 4.2.1. *(GP-RDPF) Let $D \geq 0$, $P \geq 0$, and $X \sim \mathcal{GP}(0, K_X)$ be a source modeled by a GP. Then, the associated RDPF under MSE distortion and squared Wasserstein-2 divergence is achieved by a reconstruction $Y \sim \mathcal{GP}(0, K_Y)$ (i.e., is itself a GP), such that K_X and K_Y share the same set of eigenvectors. Additionally, the RDPF can be*

expressed as

$$\begin{aligned} R_{GP}(D,P) = \min_{\{P_{Y_i|X_i}\}_{i=1}^{\infty}} & \sum_{i=1}^{\infty} I(X_i;Y_i) \\ s.t. \quad & \sum_{i=1}^{\infty} \mathbb{E}\left[\|X_i - Y_i\|^2\right] \leq D \\ & \sum_{i=1}^{\infty} \mathrm{W}_2^2(X_i,Y_i) \leq P \end{aligned} \tag{4.2.1}$$

where $\{X_i\}_{i=1}^{\infty}$ *and* $\{Y_i\}_{i=1}^{\infty}$ *are the KL coefficients of the GP* X *(source) and the GP* Y *(reconstruction), respectively.*

Proof. See Appendix 4.4.1 □

Theorem 4.2.1 provides a structural characterization of the RDPF problem for GPs and serves as an optimization problem on the set of joint random variables $\{(X_i, Y_i)\}_{i=1}^{\infty}$ defining the statistics of the source and reconstruction GPs. This process can be seen as selecting the proper set of orthonormal vectors, i.e., the eigenvectors of K_X, so that the problem "diagonalizes", similarly to the finite-dimensional Gaussian RDPF [23].
The following corollary further simplifies (4.2.1) leveraging knowledge of the analytic solution of the scalar Gaussian RDPF under MSE distortion and $\mathrm{W}_2^2(\cdot,\cdot)$ perception.

Corollary 4.2.1. *The optimization problem in* (4.2.1) *can be cast as follows*

$$\begin{aligned} R_{GP}(D,P) = \min_{\{(D_i,P_i)\}_{i=1}^{\infty}} & \sum_{i=1}^{\infty} R_{X_i}(D_i,P_i) \\ s.t. \quad & \sum_{i=1}^{\infty} D_i \leq D \quad \sum_{i=1}^{\infty} P_i \leq P \end{aligned} \tag{4.2.2}$$

where $R_{X_i}(\cdot,\cdot)$ *is the RDPF under MSE distortion and squared Wasserstein-2 perception for the Gaussian* $X_i \sim \mathcal{N}(0,\lambda_i)$*.*

Proof. See Appendix 4.4.2 □

In Chapter 3, we analyzed the finite-dimensional version of (4.2.2), designing a solving algorithm based on alternating minimization for the computation of the optimal allocations $\{(D_i, P_i)\}_{i=1}^{\infty}$. In that case, the solution leveraged an alternating minimization scheme, where $\{D_i\}_{i=1}^{\infty}$ and $\{P_i\}_{i=1}^{\infty}$ are alternatively optimized while fixing the other set of variables. Alas, this computational approach cannot be implemented in (4.2.2) due to the cardinality of the set of optimization variables. However, fixing the perceptual levels $\{P_i\}_{i=1}^{\infty}$ allows us to characterize the optimal distortion allocation and associated per-dimension rates, as shown in the following theorem.

Theorem 4.2.2. *Let the perception allocations* $\{P_i\}_{i=1}^{\infty}$*, such that* $\sum_{i=1}^{\infty} P_i \leq P$*, be given. Then, the associated optimal distortion allocations* $\{\tilde{D}_i\}_{i=1}^{\infty}$ *and the per-dimension rates*

$\{\tilde{R}_{X_i}\}_{i=1}^{\infty}$ *are*

$$\tilde{D}_i = \begin{cases} \min\{\gamma, \lambda_i\} & \text{if } \sqrt{P_i} \geq \sqrt{\lambda_i} - \sqrt{\lambda_i - \min\{\lambda_i, \gamma\}} \\ P_i + 2\sqrt{\lambda_i}\left(\sqrt{\lambda_i} - \sqrt{P_i}\right) & \\ \quad +\gamma - \sqrt{4\lambda_i\left(\sqrt{\lambda_i} - \sqrt{P_i}\right)^2 + \gamma^2} & \text{otherwise.} \end{cases}$$

$$\tilde{R}_{X_i} = \begin{cases} \frac{1}{2}\log^+\left(\frac{\lambda_i}{\gamma}\right) & \text{if } \sqrt{P_i} \geq \sqrt{\lambda_i} - \sqrt{\lambda_i - \min\{\lambda_i, \gamma\}} \\ \frac{1}{2}\log\left(\frac{2\lambda_i(\sqrt{\lambda_i}-\sqrt{P_i})^2}{\gamma\left(\sqrt{4\lambda_i(\sqrt{\lambda_i}-\sqrt{P_i})^2+\gamma^2}-\gamma\right)}\right) & \text{otherwise.} \end{cases}$$

where $\log^+(x) = \max\{x, 0\}$ *and* $\gamma \geq 0$ *is chosen such that* $\sum_{i=1}^{\infty} \tilde{D}_i \leq D$.

Proof. The proof of the optimal distortion allocation $\{\tilde{D}_i\}$ follows as a limit case of [23, Theorem 7] for the finite-dimensional case. The additional assumption $\sum_{i=1}^{\infty} P_i \leq P$ is required to ensure that $\sum_{i=0}^{\infty} \tilde{D}_i < \infty$ independently of γ, since

$$\sum_{i=1}^{\infty} \tilde{D}_i < \sum_{i=1}^{\infty} P_i + 2\sum_{i=1}^{\infty} \lambda_i \overset{(a)}{<} +\infty$$

where (a) derives from $\{\lambda_i\}_{i=1}^{\infty}$ being the eigenvalues of the trace-class operator K_X. The associated per-dimension rates $\{R_{X_i}\}_{i=1}^{\infty}$ result from the scalar Gaussian RDPF [23] evaluated at $(\tilde{D}_i, P_i)$. This concludes the proof. □

We stress the following technical remark for Theorem 4.2.2.

Remark 4.2.1. *Fixing the perceptual levels* $\{P_i\}_{i=1}^{\infty}$*, the parametric solutions recovered in Theorem 4.2.2 can be interpreted as an extension of the classical water-filling solution for GP-RDF. In fact, we note that the first branch of the function* $\tilde{D}_i$ *and* $\tilde{R}_{X_i}$ *recovers the classical RDF solutions [66]. Conversely, the second branch of the functions can be interpreted as an adaptive water-level solution;* $\tilde{D}_i$*, through the dependence on the* i^{th} *dimension second moment* λ_i*, gets adapted to each dimension, thus guaranteeing that all source components are present in the reconstructed signal.*

It should be noted that the analytical characterization of optimal perception levels $\{P_i^*\}_{i=1}^{\infty}$ remains an open problem, even in the finite-dimensional setting. However, for any convergent series $\{\tilde{P}_i\}_{i=1}^{\infty}$, Theorem 4.2.2 characterizes the associated minimizing distortion allocation $\{\tilde{D}_i\}_{i=1}^{\infty}$ and the per-dimension rates $\{\tilde{R}_{X_i}\}_{i=1}^{\infty}$, which identify an analytical upper bound to the GP-RDPF, i.e.,

$$R_{GP}(D, P) = \min_{\{D_i, P_i\}_{i=1}^{\infty}} \sum_{i=1}^{\infty} R_{X_i}(D_i, P_i) \leq \sum_{i=1}^{\infty} \tilde{R}_{X_i}.$$

Nevertheless, the numerical results for the finite-dimensional Gaussian RDPF from [23, Alg. 1] hint at a relative proportionality in the perceptual level assignment. Based on this observation, we propose a heuristic perceptual levels allocation proportional to the second moments of $\{X_i\}_{i=1}^{\infty}$, i.e.,

$$\tilde{P}_i = \alpha\lambda_i \quad \Longrightarrow \quad \alpha\sum_{i=1}^{\infty}\lambda_i = P \tag{4.2.3}$$

where α acts as a proportionality constant to enforce the desired perception level P. The advantage of this allocation is that it recovers the optimal solution for $P \to 0$, i.e., for all $i = 1, 2, \ldots$ $P_i \to 0$, while still providing an excellent bound in the general case.

4.2.1 GP-RDPF for Stationary Sources

We devote this section to characterizing the particular case of GP-RDPF for stationary processes. To this end, we consider $\Omega = (\mathcal{X}, \Sigma_{\mathcal{X}}, \mu)$ such that $\mathcal{X} = [0, T]$, $\Sigma_{\mathcal{X}}$ is the σ-algebra of all Lebesgue measurable subsets $\mathcal{U} \subseteq [0, T]$, and μ is the Lebesgue measure. Similarly to the classical RDF [65, 66], in this setting, we formulate the RDPF considering normalized versions of main quantities of interest, i.e.,

$$R = \frac{1}{T}\sum_{i=1}^{\infty} R_{X_i}(D_i, P_i), \quad D \geq \frac{1}{T}\sum_{i=1}^{\infty} D_i, \quad P \geq \frac{1}{T}\sum_{i=1}^{\infty} P_i.$$

Under the assumption that the source X is a stationary GP, this normalization allows us to extend the results of Theorem 4.2.2 to the case where $T \to \infty$, as shown in the following theorem.

Theorem 4.2.3. *Let $X \sim \mathcal{GP}(0, k_X)$ be a stationary process defined on $[0, T] \subset \mathbb{R}$ and let S_X be the associated power spectral density. Then, for $T \to \infty$, the GP-RDPF is upper bounded by the parametric curve $(\tilde{R}, \tilde{D}, \tilde{P})$ parameterized by $\gamma > 0$ and $0 \leq \alpha \leq 1$ as*

$$\begin{aligned}\tilde{R} = &\frac{1}{2}\int_{\mathcal{S}_{RDP}} \log\left(\frac{2S_X^2(f)(1-\sqrt{\alpha})^2}{\gamma(\sqrt{\gamma^2 + 4S_X^2(f)(1-\sqrt{\alpha})^2} - \gamma)}\right) df \\ &+ \frac{1}{2}\int_{\mathcal{S}_{RD}} \log^+\left(\frac{S_X(f)}{\gamma}\right) df\end{aligned} \tag{4.2.4}$$

$$\begin{aligned}\tilde{D} &= \int_{\mathcal{S}_{RDP}} \hat{D}(S_X(f))df + \int_{\mathcal{S}_{RD}} \min\{\gamma, S_X(f)\}df \\ \hat{D}(x) &= \gamma + x(1 + (1-\sqrt{\alpha})^2) - \sqrt{4x^2(1-\sqrt{\alpha})^2 + \gamma^2}\end{aligned} \tag{4.2.5}$$

and $\tilde{P} = \alpha \int S_X(f)df$, where the sets $\mathcal{S}_{RD}$ and $\mathcal{S}_{RDP}$ are defined as

$$\mathcal{S}_{RD} \triangleq \{f \in \mathbb{R} : S_X(f) \geq \gamma + (1-\sqrt{\alpha})^2 S_X(f)\}$$

$$\mathcal{S}_{RDP} \triangleq \mathbb{R}/\mathcal{S}_{RD}.$$

Proof. Considering the proportional perceptual level allocation in (4.2.3), the proof follows as a direct application of [65, Lemma 8.5.3] to the results of Theorem 4.2.2. □

4.3 Numerical Example

In this section, we provide a numerical example to better illustrate the results of Theorem 4.2.3. Let the source X be modeled as a stationary GP with power spectral density S_X as reported in Figure 4.1.

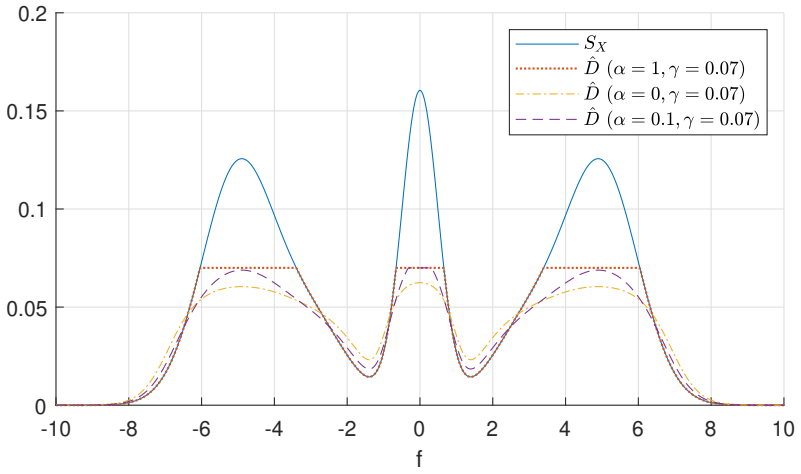
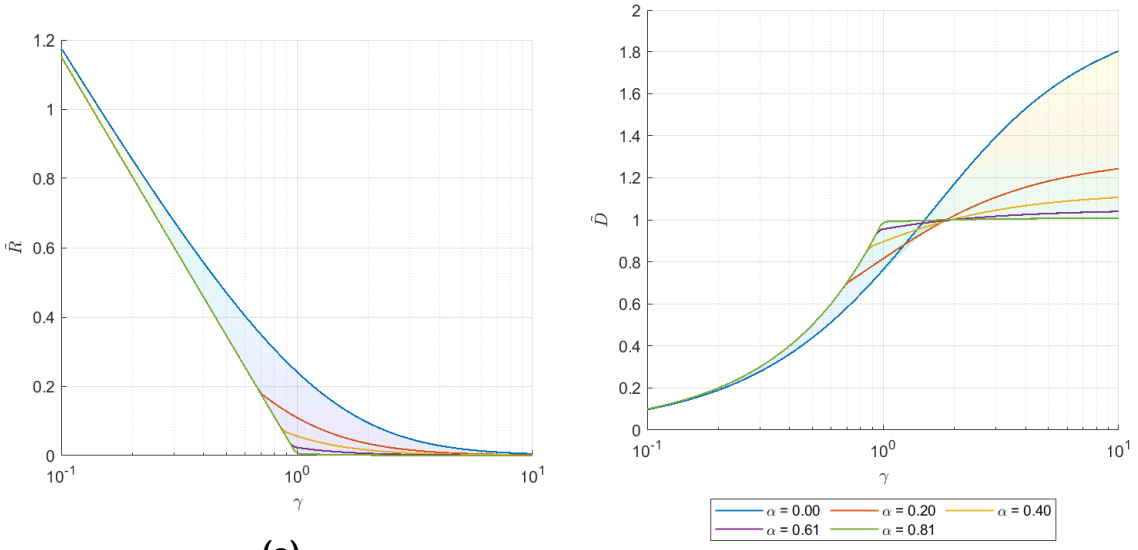


Figure 4.1: Source Power Spectrum $S_X(f)$ vs. per-frequency distortion $\hat{D}(S_X(f))$ for $\gamma = 0.7$ and varying α.

We first investigate the per-frequency distortion $\hat{D}(S_X(f))$, defined as the integral argument of (4.2.5), for fixed γ and varying $\alpha \in [0, 1]$. For $\alpha = 1$, i.e., no perceptual constraint is enforced, distortion allocation remains constant and independent of the source power spectrum $S_X(f)$. In this case, it adheres to the classical RDF water-filling solution [65]. For lower values of α, i.e., stricter perceptual requirements, distortion allocation is adapted to the structure of the source power spectrum $S_X(f)$, extending the observations in Remark 4.2.1.

Figure 4.2 shows the numerically estimated curves of the rate $\tilde{R}$ (Figure 4.1(a)) and distortion $\tilde{D}$ (Figure 4.1(b)) defined in (4.2.4) and (4.2.5), respectively. The curves are calculated considering the parameters $\gamma \in [0.01, 1]$ and $\alpha \in [0, 1]$. Similarly to the Gaussian RPDF in the finite-dimensional setting, the rate penalty due to the perceptual constraint diminishes in significance in the low distortion regime, i.e., as $\gamma \to 0$. However, in the moderate to high distortion regime, the impact of the perceptual constraint controlled by α becomes increasingly pronounced.



(a) **(b)**

Figure 4.2: Rate $\tilde{R}$ (a) and distortion $\tilde{D}$ (b) curves parameterized by (α, γ).

4.4 Appendix

4.4.1 Proof of Theorem 4.2.1

Before we delve into the technicalities of the proof, we give some useful notation. We denote with $\{\phi_i\}_{i=0}^{+\infty}$ and $\{\lambda_i\}_{i=0}^{+\infty}$ the set of eigenvectors and eigenvalues of K_X. Moreover, let $\{\eta_i\}_{i=0}^{+\infty} \subset L^2(\Omega)$ be any countable orthonormal set of eigenvectors. Then, we can construct the HS operator K_Y and the associated process Y as

$$[K_Y \psi] = \sum_{i=1}^{\infty} \nu_i \langle \psi, \eta_i \rangle \eta_i \qquad Y = \sum_{i=1}^{\infty} Y_i \eta_i, \tag{4.4.1}$$

with $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}[Y_i Y_j] = \nu_i \delta_{i,j}$. As a result, the mutual information between processes X and Y can be expressed as $I(\{X_i\}_{i=1}^{\infty}; \{Y_i\}_{i=1}^{\infty})$. Note that until now, we did not assume that $\{Y_i\}_{i=0}^{\infty}$ is necessarily Gaussian distributed.

We now show that the assumption $\{\eta_i\}_{i=1}^{\infty} = \{\phi_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ Gaussian distributed is optimal. Leveraging the equivalence between GP and non-degenerate Gaussian measures, [71, Proposition 2.4] allows to lower bound the $\mathrm{W}_2^2(\cdot,\cdot)$ perception as follows

$$\begin{aligned} \mathrm{W}_2^2(f_X, f_Y) &\overset{(a)}{\geq} \sum_{i=1}^{\infty} \mathrm{W}_2^2(X_i, Y_i) \\ &\overset{(b)}{\geq} \sum_{i=1}^{\infty} |\mu_{X_i} - \mu_{Y_i}|^2 + \sum_{i=1}^{\infty} (\sqrt{\lambda_i} - \sqrt{\eta_i})^2 \end{aligned}$$

where (a) and (b) hold with equality iff $\{\eta_i\}_{i=1}^{\infty} = \{\phi_i\}_{i=1}^{\infty}$ and $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ are Gaussian distributed, respectively. The optimality of the assumptions considered

with respect to the mutual information and MSE distortion is derived from their proven optimality in the classical RDF [66]. Consequently, the MSE of the two processes can be expressed as

$$\begin{aligned}\mathbb{E}\left[\|X-Y\|^2\right] &= \mathbb{E}\left[\left\|\sum_{i=0}^{\infty}(X_i-Y_i)\phi_i\right\|^2\right] \\ &= \sum_{i=0}^{\infty}\mathbb{E}\left[|X_i-Y_i|^2\right].\end{aligned}$$

Furthermore, since $\{Y_i\}_{i=0}^{\infty}$ is a set of uncorrelated Gaussian random variables, and therefore independent, the mutual information $I\left(\{X_i\}_{i=1}^{\infty};\{Y_i\}_{i=1}^{\infty}\right)$ becomes

$$I\left(\{X_i\}_{i=1}^{\infty};\{Y_i\}_{i=1}^{\infty}\right) = \sum_{i=1}^{\infty} I(X_i,Y_i).$$

This concludes the proof.

4.4.2 Proof of Corollary 4.2.1

Introducing the additional constraints $\mathbb{E}\left[|X_i-Y_i|^2\right] \leq D_i$ and $\mathrm{W}_2^2(X_i,Y_i) \leq P_i$, we can express (4.2.1) as

$$R_{GP}(D,P) = \min_{\substack{\{(D_i,P_i)\}_{i=1}^{\infty} \\ \sum_{i=1}^{\infty} D_i \leq D \\ \sum_{i=1}^{\infty} P_i \leq P}} \sum_{i=1}^{\infty} \min_{\substack{P_{Y_i|X_i} \\ \mathbb{E}\left[|X_i-Y_i|^2\right]\leq D_i \\ \mathrm{W}_2^2(X_i,Y_i)\leq P_i}} I(X_i,Y_i)$$

where each element in the summation can be recognized to be the definition of the RDPF $R_{X_i}(D_i,P_i)$ for a Gaussian source X_i. This concludes the proof.

Chapter 5

Copula-based Estimation of Constrained Rate-Distortion Functionals

In this chapter, we take a step back from the distributional assumptions of previous chapters and adopts a fully general perspective. Rather than restricting the analysis to information source belonging a specific family of distributions, such as the Gaussian, we propose a novel copula-based estimation method for computing the Perfect Realism RDPF (PR-RDPF) for arbitrary multivariate continuous sources, subject to a single-letter average distortion constraint. The proposed framework is notably general in scope, as it also enables the computation of the Entropic Optimal Transport (EOT) and the Output-Constrained Rate-Distortion Function (OC-RDF), of which the PR-RDPF emerges as a special case.

The main contributions of this chapter 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 5.1.1), 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 divergence, i.e., I-projection, on a convex constraint set (Problem 5.2.1). However, although this class of projection has been extensively studied in [72], 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 5.2.2) that we subsequently show that it can be made arbitrarily tight (Theorem 5.2.2). **(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 5.2.3). **(iv)** We propose an algorithmic approach via a stochastic gradient descent method, to estimate the strictly convex optimization problem of Theorem 5.2.3 (see Alg. 4). **(v)** We derive a Shannon lower bound (SLB) for the PR-RDPF under MSE distortion (Theorem 5.2.4). We supplement our theoretical results with various numerical evaluations aiming to estimate the PR-RDPF

under various sources and different distortion measures via Alg. 4, and to demonstrate the efficacy of our algorithmic approach compared to the obtained SLB.

5.1 Preliminaries

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

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

Definition 5.1.1. *(PR-RDPF) Let $P_X \in \mathcal{P}(\mathcal{X})$. Then, the PR-RDPF for the source $X \sim P_X$ under a distortion measure $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$ is given as follows*

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

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

It should be noted that the *perfect realism* regime represents a limit case of the general problem of the RDPF [9], where one constrains the reconstruction Y to have the same distribution as the source X. Although PR-RDPF became quite popular through [9], similar ideas were previously explored by Li *et. al.* in [73], 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 [34] 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 [61], 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 [33] (see also [73]), for which the mathematical definition is stated next.

Definition 5.1.2. *(OC-RDF) Let $P_X \in \mathcal{P}(\mathcal{X})$. Then, the OC-RDF for the source $X \sim P_X$ under a distortion measure $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$ and a target reconstruction distribution $P_Y \in \mathcal{P}(\mathcal{Y})$ is given as follows*

$$R_{OC}(D) = \min_{\substack{P_{Y|X} \in \hat{\Pi}(P_X, P_Y) \\ \mathbb{E}[d(X,Y)] \leq D}} I(X,Y) \tag{5.1.1}$$

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

The main difference between the problems of PR-RDPF and OC-RDF lies in how the constraint on the reconstruction distribution p_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 5.1.1. *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. $P_Y = P_X$).*

Additionally, the OC-RDF highlights an interesting connection to the EOT problem (see [74, 75]), of which the mathematical definition is stated as follows.

Definition 5.1.3. *(EOT) Let $P_X \in \mathcal{P}(\mathcal{X})$ and $P_Y \in \mathcal{P}(\mathcal{Y})$. Then, the EOT for $\epsilon > 0$ and distortion measure $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$, is given as follows*

$$D_{EOT}(\epsilon) = \min_{P_{X,Y} \in \bar{\Pi}(P_X, P_Y)} \mathbb{E}\left[d(X,Y)\right] + \epsilon I(X,Y) \tag{5.1.2}$$

where the minimization is on the convex set of joint distributios $\bar{\Pi}(P_X, P_Y) \triangleq \{P_{X,Y} : m_X(P_{X,Y}) = P_X, m_Y(P_{X,Y}) = P_Y\}$.

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 [76, 77], this observation has not been documented in this specific setting, hence we formalized it in the following theorem.

Theorem 5.1.1. *(Connection of OC-RDF and EOT) Let $P_X \in \mathcal{P}(\mathcal{X})$ and $P_Y \in \mathcal{P}(\mathcal{Y})$. Then, for any $D > 0$, there exists an $\epsilon > 0$ such that the problems of OC-RDF and EOT are equivalent.*

Proof. See Appendix 5.4.1. □

In view of Theorem 5.1.1, we can treat the OC-RDF and EOT problems as equivalent problems. As a result, the computational schemes derived in Section 5.2 applicable to the OC-RDF problem can be adapted *mutatis mutandis* to the EOT problem.

5.1.2 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 chapter. The following definitions and theorems are taken from [78].

Definition 5.1.4. *(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 5.1.2. *(Sklar's Theorem) Let P be a d-dimensional d.f. with marginal d.f. $P_1, P_2, \ldots, P_d$. Let A_j denote the range of P_j, $A_j \triangleq P_j\left(\bar{\mathbb{R}}\right)$ $(j = 1, 2, \ldots, d)$. Then, there exists a d-copula d.f. C such that for all $(x_1, x_2, \ldots, x_d) \in \bar{\mathbb{R}}^d$,*

$$P(x_1, \ldots, x_d) = C\left(P_1(x_1), \ldots, P_d(x_d)\right). \tag{5.1.3}$$

Such a C is uniquely determined on $A_1 \times A_2 \times \cdots A_d$ and, hence, it is unique when $P_1, P_2, \ldots, P_d$ are continuous.

Corollary 5.1.1. *Let $p : \bar{\mathbb{R}}^d \to \mathbb{R}^+$ be the pdf associated with* (5.1.3)*. Then, p can be uniquely decomposed as*

$$p(x_1, \ldots, x_d) = c\left(P_1(x_1), \ldots, P_d(x_d)\right) \prod_{j=1}^{d} p_j(x_j) \tag{5.1.4}$$

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

Corollary 5.1.2. *Let $P_1, P_2, \ldots, P_d$ be univariate d.f.'s and C be a copula d.f.. Then, the function $P : \bar{\mathbb{R}}^d \to [0,1]$ defined in* (5.1.3) *is a d-dimensional d.f. with marginals $P_1, P_2, \ldots, P_d$.*

It is worth noticing that Corollary 5.1.1 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 5.1.2 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 5.1.5. *(Quantile function) Let $X \sim P_X$ be a univariate RV on $\mathcal{X} \subseteq \mathbb{R}$. We define the quantile function $Q_X : [0,1] \to \mathbb{R}$ as $Q_X(u) \triangleq \sup\{x \in \mathcal{X} : P_X(x) \leq u\}$. If P_X is continuous and strictly increasing, then $Q_X = P_X^{-1}$. However, even if P_X may fail to have an inverse function, Q_X guaranties that $Q_X\left(P_X(X)\right) = X$ almost surely (a.s.).*

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

5.2 Main Results

In this section, we derive our main results.

5.2.1 Copula Lower Bound

First, we prove a lemma with which the functionals in the mathematical formulations of Definitions 5.1.2 and 5.1.3 can be redefined using copula distributions[1].

Lemma 5.2.1. *Let $(X,Y) \sim p_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ be a $2d$-variate RV with marginal pdfs $p_X \in \mathcal{P}(\mathcal{X})$ and $p_Y \in \mathcal{P}(\mathcal{Y})$. Then, the mutual information $I(X,Y)$ can be equivalently written as follows*

$$I(X,Y) = \mathrm{D}_{\mathrm{KL}}(C_{X,Y} \| C_X \times C_Y) \tag{5.2.1}$$

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

$$\mathbb{E}_{P_{X,Y}}\left[d(X,Y)\right] = \mathbb{E}_{C_{X,Y}}\left[d\left(\Psi_X(U_X), \Psi_Y(U_Y)\right)\right] \tag{5.2.2}$$

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

Proof. See Appendix 5.4.2. □

Leveraging Lemma 5.2.1, we can provide an alternative formulation of the mathematical expression in (5.1.1), which will be the subject of our estimation analysis. This is stated next as Problem 5.2.1.

Problem 5.2.1. *(Copula-based OC-RDF) The mathematical expression* (5.1.1) *can be reformulated as follows*

$$R_{OC}(D) = \min_{C\in\mathcal{C}_{2d}} \mathrm{D}_{\mathrm{KL}}(C \| C_X \times C_Y) \tag{5.2.3}$$

$$s.t.\ \mathbb{E}_C\left[d(\Psi_X(U_X), \Psi_X(U_Y))\right] = D \tag{5.2.4}$$

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

Remark 5.2.1. *(On Problem 5.2.1) Problem 5.2.1 is a convex program in the space of copula d.f. Moreover, the problem is equivalent to finding the I-projection of $C_X \times C_Y$ on the set $\mathcal{B} \subset \mathcal{C}_{2d}$ of copula d.f. satisfying the modified distortion constraint* (5.2.4).

Problem 5.2.1 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 $C_X \times C_Y$ while respecting a linear set of constraints. This

[1] An alternative link between the mutual information $I(X,Y)$ and the associated copula entropy $h(C_{X,Y})$ can be found in [79].

class of projection problems has been thoroughly studied by Csiszár in [72], where the analytical form of the optimal projection for the considered case has been characterized. Using [72], we derive the following theorem.

Theorem 5.2.1. *(Analytical solution of Problem 5.2.1) Let $R = C_X \times C_Y$ and assume there exists a copula d.f. P such that $\mathrm{D_{KL}}(P\|R) < \infty$ and (5.2.4) is satisfied. Then, Problem 5.2.1 admits a minimizing copula Q with Radon–Nikodym derivative with respect to the measure R of the form*

$$\frac{dC}{dR}(\mathbf{u}) = e^{\zeta+\theta[d(\Psi_X(\mathbf{u}_x),\Psi_Y(\mathbf{u}_y))]} \prod_{i=1}^{2d} g_i(u_i) \tag{5.2.5}$$

for some constants (ζ, θ), and nonnegative uni-variate functions g_i such that $\log(g_i(s)) \in \mathcal{L}_1([0,1])$ for $i = 1, \ldots, 2d$.

Proof. See Appendix 5.4.3. □

Although Theorem 5.2.1 provides a characterization of the solution of Problem 5.2.1, 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 (5.2.5). Following an idea of [80], we circumvent this technical issue by introducing a relaxation on the constraint set of Problem 5.2.1, that results into a lower bound on OC-RDF. This is demonstrated next in Problem 5.2.2.

Problem 5.2.2. *(Lower bound to Problem 5.2.1) For any integer N, Problem 5.2.1 can be lower bounded as follows*

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

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

Remark 5.2.2. *(Problem 5.2.1 vs Problem 5.2.2) The main technical difference between Problems 5.2.1 and 5.2.2 concerns their constraint sets. Particularly, in Problem 5.2.1 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 5.2.2 only require to respect up to N moments of a uniform distribution. This in turn implies that the constraint set of Problem 5.2.1 is a proper subset of the constraint set of Problem 5.2.2, justifying the lower bound of the latter.*

In the following theorem, we show that, for $N \to \infty$, Problem 5.2.2 recovers the solution of Problem 5.2.1.

Theorem 5.2.2. *Let Q^* be the optimal solution of Problem 5.2.1 and $\hat{Q}_N^*$ be the optimal solution of Problem 5.2.2. Then, as $N \to \infty$,*

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

Proof. See Appendix 5.4.4. □

We now provide the analytical form of the solution of Problem 5.2.2. Unlike Theorem 5.2.1, 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 5.2.2 obtained as result of its dual problem.

Theorem 5.2.3. *(Analytical solution of Problem 5.2.2) Let $R = C_X \times C_Y$ and assume there exists a d.f. P on $[0,1]^{2d}$ such that $\mathrm{D}_{\mathrm{KL}}(P\|R) < \infty$ and (5.2.4) is satisfied. Then, Problem 5.2.2 admits minimizing copula Q with Radon–Nikodym derivative with respect to the measure R of the form*

$$\frac{dQ}{dR}(\mathbf{u}) = e^{\zeta + \theta d(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y))} \prod_{i=1}^{2d} e^{\sum_{n=0}^{N} \nu_{i,n} u_i^n} \tag{5.2.6}$$

where the constants $(\zeta, \theta, \{\nu_{i,n}\}_{(i,n)\in I})$ are the Lagrangian multipliers of Problem 5.2.2 obtained as a result of the following dual program

$$\min_{(\zeta,\theta,\{\nu_{i,n}\}_{(i,n)\in I})} \mathbb{E}_R\left[\frac{dQ}{dR}\right] - \zeta - \theta D - \sum_{(i,n)\in I} \nu_{i,n}\alpha_n \tag{5.2.7}$$

Proof. See Appendix 5.4.5. □

The following result is a consequence of Theorem 5.2.3.

Corollary 5.2.1. *Let Q be the minimizing copula d.f. characterized in Theorem 5.2.3. Then, the mutual information $I(X,Y)$ of the joint distribution (X,Y) defined by marginals d.f. $\{P_{X_i}\}_{i=1,\ldots,d}$ and $\{P_{Y_i}\}_{i=1,\ldots,d}$ and copula Q is given by*

$$I(X,Y) = \mathrm{D}_{\mathrm{KL}}(Q\|R) = -\zeta - \theta D - \sum_{(i,n)\in I} \nu_{i,n}\alpha_n. \tag{5.2.8}$$

5.2.2 Copula Estimation

As anticipated in Theorem 5.2.3, the Lagrangian multipliers $(\zeta, \theta, \{\nu_{i,n}\}_{(i,n)\in I})$ defining the optimal solution of Problem 5.2.2 can be obtained by solving (5.2.7). Although not available in closed form, the solution of (5.2.7) can be optimally computed using numerical methods, given the properties of the problem.

Lemma 5.2.2. *The optimization problem* (5.2.7) *is strictly convex, hence it has a unique solution.*

Proof. See Appendix 5.4.6. □

To compute (5.2.7), 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 (5.2.7), 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 (5.2.8)), we can approximate the integral using Monte Carlo method [81]. The resulting iterative scheme can be considered as a *mini-batch stochastic gradient descent algorithm* on a convex objective [82]. The algorithm is given in Alg. 4.

Algorithm 4 Copula Estimation $R_{OC}(D)$

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

1: **for** i **do** $= 1, \dots, T$
2: Sample $\{\mathbf{u}_i\}_{i=1\dots M}$ with $u_i \sim U([0,1]^{2d})$
3: $f(\mathbf{l}) \approx (5.2.8) + \left(\frac{1}{M}\sum_{i=1}^{M} \frac{dQ}{dR}(\mathbf{l}, \mathbf{u}_i) dR(\mathbf{u}_i)\right)$
4: $\mathbf{l^{(i)}} = \texttt{GradientMethod}(\mathbf{l}^{(i-1)}, f)$
5: **end for**

Ensure: Lagrangian multipliers $\mathbf{l}^{(T)}$; $I(X,Y) = (5.2.8)$.

5.2.3 SLB for PR-RDPF

In this subsection, we prove a generalization of the well-known SLB on the classical RDF with MSE distortion [62] to the case of PR-RDPF, denoted hereinafter by R_{PR}^{SLB}. The bound is stated in the following theorem.

Theorem 5.2.4. *(SLB for PR-RDPF) Let* $\mathcal{S} \triangleq \{p_X : \mathbb{E}_{p_X}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right] \preceq \Sigma\}$ *be the set of source distribution with a fixed covariance matrix* Σ*. Then, for all* $X \sim p_X$ *with* $p_X \in \mathcal{S}$*, the PR-RDPF under MSE distortion constraint admits the following lower bound*

$$R_{PR}(D) \geq R_{PR}^{SLB}(D) = h(X) - h(X^*) + R_{PR}^{G}(D) \tag{5.2.9}$$

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

Proof. See Appendix 5.4.7. □

We stress the following technical remark on Theorem 5.2.4.

Remark 5.2.3. *(On Theorem 5.2.4) For the scalar case of the PR-RDPF, let* $\mathcal{S} \triangleq \{p_X : \mathbb{E}_{p_X}\left[(X - \mathbb{E}[X])^2\right] \leq \sigma^2]\}$ *for a finite variance value* σ^2*. Then,* (5.2.9) *can be further simplified to*

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

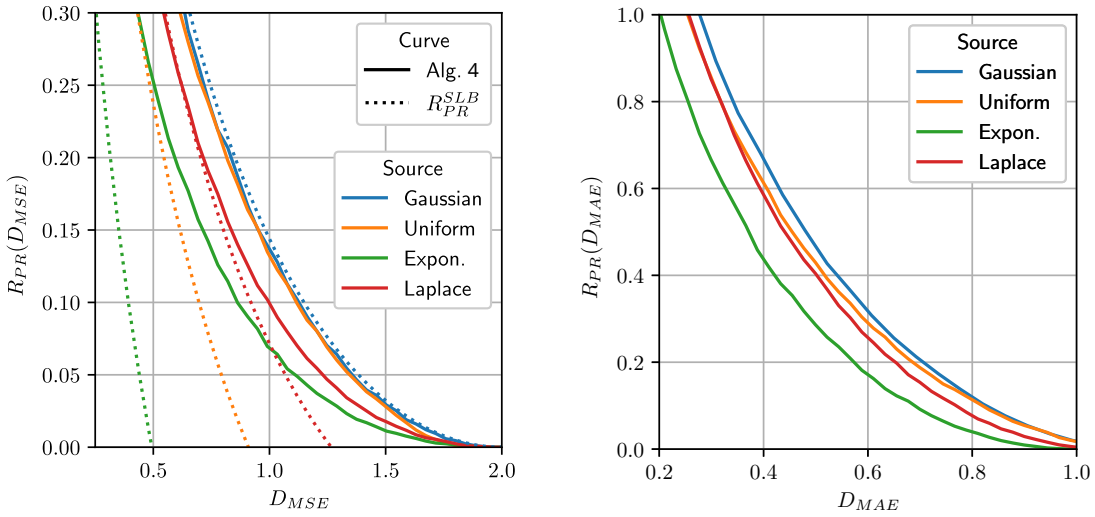
with $N(X)$ denoting the entropy power of source X. For the general vector case, the lower bound depends on the vector Gaussian PR-RDPF, R_{PR}^{G}, which can be easily computed using the adaptive reverse-water-filling solution developed in [23, Corollary 3].

5.3 Numerical Results

In this section, we provide a numerical estimation of the PR-RDPF for both scalar and vector sources using Alg. 4.

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. 5.0(a)), and (b) the $\mathcal{L}_1$ norm i.e. the mean-absolute-error (MAE) distortion (see Fig. 5.0(b)). 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. 5.0(a), we also compare the estimated result with the SLB derived in Theorem 5.2.4. In Fig. 5.0(a), the Gaussian source case allows us to quantify the



(a) **(b)**

Figure 5.1: PR-RDPF for various source distributions under (a) MSE distortion metric and (b) MAE distortion metric.

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. 5.1(a)) or exponentially (see Fig. 5.1(b)) distributed with zero mean and variance $\sigma^2 = 1$. In both cases, the multivariate distribution is constructed by imposing a Gaussian coupling [78] 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.

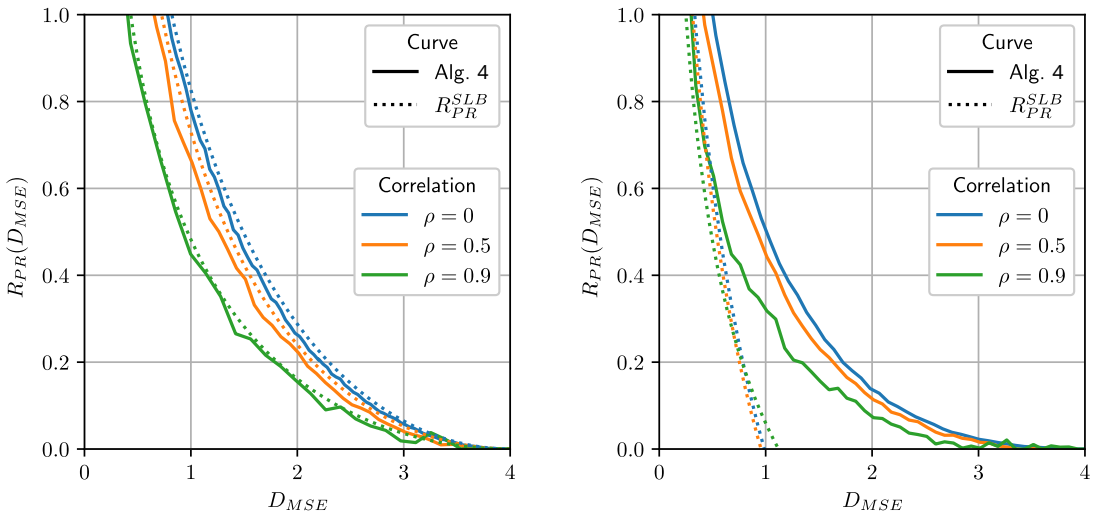


Figure 5.2: PR-RDPF under MSE distortion metric for a (a) Gaussian and (b) exponential bivariate source.

5.1(a), we demonstrate a comparison between the Gaussian PR-RDPF estimate obtained via Alg. 4 with the R_{PR}^{SLB} obtained in (5.2.9) with the term $R_{PR}^{G}(D)$ computed via the optimal adaptive reverse-water-filling solution of [23, Corollary 3], which results into a tight $R_{PR}^{SLB}(D)$. We observe that Alg. 4 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. 4 remains stable in the low to moderate correlation cases while showing a slightly noisier behavior (fluctuations) in the high correlation case. Contrary to Fig. 5.1(a), in Fig. 5.1(b) we observe that beyond high resolution (low distortion), the exponential PR-RDPF estimate obtained via Alg. 4 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.

5.4 Appendix

5.4.1 Proof of Theorem 5.1.1

We start by showing that $\hat{\Pi}$ and $\bar{\Pi}$ in Definitions 5.1.2 and 5.1.3, define the same set, i.e., there exists a bijection between two sets. Assuming p_X to be the source distribution in OC-RDF, then for any Markov kernel $p_{Y|X} \in \hat{\Pi}$, the joint pdf $p_{Y|X} \cdot p_X$ lies in $\bar{\Pi}$. Conversely, for any joint distribution $p_{X,Y} \in \bar{\Pi}$ the Markov kernel $p_{Y|X} = \frac{p_{X,Y}}{p_X}$ belongs to $\hat{\Pi}$. Hence, there is a one-to-one mapping between the optimization variables of Definitions 5.1.2 and 5.1.3.

Let (D, λ) be the pair composed by the distortion level in the constraint of (5.1.1) and the associated Lagrangian multiplier. Then, the Lagrangian functional of Definition 5.1.2 for distortion level D is defined as

$$\mathcal{L}_{RD}(p_{Y|X}, \lambda) \triangleq I(X,Y) + \lambda \mathbb{E}_{p_{Y|X}p_X}\left[d(X,Y)\right]. \tag{5.4.1}$$

Similarly, the Lagrangian functional associated with Definition 5.1.3 is defined as

$$\mathcal{L}_{EOT}(p_{X,Y}, \epsilon) \triangleq \mathbb{E}_{p_{X,Y}}\left[d(X,Y)\right] + \epsilon I(X,Y). \tag{5.4.2}$$

Based on (5.4.1), (5.4.2), we observe that the following relation holds

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

hence

$$\begin{aligned}\arg\min_{p_{X,Y}\in\bar{\Pi}} \mathcal{L}_{EOT}\left(p_{X,Y}, \frac{1}{\lambda}\right) &= \arg\min_{p_{X,Y}\in\bar{\Pi}} \mathcal{L}_{RD}\left(\frac{p_{X,Y}}{p_X}, \lambda\right) \\ &= p_X \cdot \arg\min_{p_{Y|X}\in\hat{\Pi}} \mathcal{L}_{RD}(p_{Y|X}, \lambda).\end{aligned} \tag{5.4.3}$$

As a result, (5.4.3) shows that the solution of Definition 5.1.3 for $\epsilon = \frac{1}{\lambda}$ is uniquely determined by the solution of Definition 5.1.2 for the pair (D, λ). This completes the proof.

5.4.2 Proof of Lemma 5.2.1

From the definitions of $I(X,Y)$ and $\mathbb{E}_{P_{X,Y}}\left[d(X,Y)\right]$, (5.2.1) and (5.2.2) can be derived as

$$\begin{aligned}I(X,Y) &= \int_{\mathbb{R}^{2d}} p_{XY}(\mathbf{x},\mathbf{y}) \log\left(\frac{p_{XY}(\mathbf{x},\mathbf{y})}{p_X(\mathbf{x})p_Y(\mathbf{y})}\right) d\mathbf{x}d\mathbf{y} \\ &\overset{(a)}{=} \int_{\mathbb{R}^{2d}} c_{X,Y}(\Phi_X(\mathbf{x}), \Psi_Y(\mathbf{y})) \cdot \log\left(\frac{c_{X,Y}(\Phi_X(\mathbf{x}), \Psi_Y(\mathbf{y}))}{c_X(\Phi_X(\mathbf{x}))c_Y(\Psi_Y(\mathbf{y}))}\right) \prod_{i=1}^{d} dP_{X_i}(x_i) dP_{Y_i}(y_i)\end{aligned}$$

$$\overset{(b)}{=} \int_{[0,1]^{2d}} c_{X,Y}(\mathbf{u}_x, \mathbf{u}_y) \log \left(\frac{c_{X,Y}(\mathbf{u}_x, \mathbf{u}_y)}{c_X(\mathbf{u}_x) c_Y(\mathbf{u}_y)} \right) d\mathbf{u}_x d\mathbf{u}_y$$
$$= \mathrm{D}_{\mathrm{KL}}(C_{X,Y} \| C_X \times C_Y)$$

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

where (a) follows from the application of Corollary 5.1.1 on $p_{X,Y}$,p_X, and p_Y, and (b) follows from the change of variables $\mathbf{u}_x = \Phi_X(\mathbf{x})$ and $\mathbf{u}_y = \Phi_Y(\mathbf{y})$.

5.4.3 Proof of Theorem 5.2.1

The proof follows similar steps to the proof of [80, 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 [80, 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 [83, 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 [80, 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 (5.2.4). 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 [72, Theorem 3.1, (Case A)], we obtain that $R_{\mathcal{A}}$ is the unique I-projection of R onto $\mathcal{A}$ with density

$$\frac{dR_{\mathcal{A}}}{dR}(\mathbf{u}) = e^{\zeta + \theta d(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y))} \tag{5.4.4}$$

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

$$\mathrm{D}_{\mathrm{KL}}(P\|R) = \mathrm{D}_{\mathrm{KL}}(P\|R_{\mathcal{A}}) + \mathrm{D}_{\mathrm{KL}}(R_{\mathcal{A}}\|R). \tag{5.4.5}$$

Moreover, under the result of [72, 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 (5.4.5) 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 [72, Theorem 3.1, (Case B)], there exist non negative scalar functions g_i with $\log(g_i) \in \mathcal{L}_1([0,1])$ 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 $\frac{dQ}{dR} = \frac{dQ}{dR_{\mathcal{A}}} \frac{dR_{\mathcal{A}}}{dR} =$ (5.2.5). This completes the proof.

5.4.4 Proof of Theorem 5.2.2

Let $\hat{Q}_i$ and $\mathcal{A}_i$ be, respectively, the copula distribution solution of Problem 5.2.2 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 5.2.1 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 [72, Theorem 2.3], we can characterize $\hat{Q}_{i+1}$ and Q^* as the unique 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 [72, Equation 3.1])

$$\mathrm{D_{KL}}(\hat{Q}_i \| Q^*) = \mathrm{D_{KL}}(\hat{Q}_i \| \hat{Q}_{i+1}) + \mathrm{D_{KL}}(\hat{Q}_{i+1} \| Q^*). \tag{5.4.6}$$

Recursively, applying (5.4.6) $k+1$ times leads to

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

from which we immediately obtain

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

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 $\{\hat{Q}_i\}_{i=0,1,\ldots}$ in KL-divergence.

To prove that the limit of the sequence is Q^*, we transform Problem 5.2.1 into a specific instance of Problem 5.2.2 under an infinite number of marginals moments constraints. As a consequence of the uniqueness of solutions of the Hausdorff moments problem [84], 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 5.2.1 into a set

of countably infinite marginal constraints.

In such form, the only difference between Problems 5.2.1 and 5.2.2 resides in the finite number i of the moment constraints of the latter. Hence, since Q^* exists and is unique and Q^* and $\hat{Q}_\infty$ coincide, also $\hat{Q}_\infty$ exists and is unique and the limit $\lim_{i\to\infty} \mathrm{D_{KL}}(\hat{Q}_i\|Q^*) = \lim_{i\to\infty} \mathrm{D_{KL}}(\hat{Q}_i\|\hat{Q}_\infty) = 0$ holds. This completes the proof.

5.4.5 Proof of Theorem 5.2.3

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 5.2.2 is defined by linear constraints in the copula d.f. $C_{X,Y}$. The results of [72, 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 5.2.2 as follows

$$\min_{g\in\mathcal{M}^+([0,1]^{2d})} \int_{[0,1]^{2d}} \log\left(g(\mathbf{u})\right) g(\mathbf{u})dR(\mathbf{u}). \tag{5.4.7}$$

$$\text{s.t.} \quad \int_{[0,1]^{2d}} g(\mathbf{u})dR(\mathbf{u}) = 1 \tag{5.4.8}$$

$$\int_{[0,1]^{2d}} d(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y))g(\mathbf{u})dR(\mathbf{u}) = D \tag{5.4.9}$$

$$\int_{[0,1]^{2d}} u_i^n g(\mathbf{u})dR(\mathbf{u}) = \alpha_n \qquad (i,n) \in I \tag{5.4.10}$$

Indicating with $\zeta', \theta, \{\nu_{i,n}\}$ the Lagrangian multipliers associated with constraints (5.4.8)-(5.4.10), we define the Lagrangian functional of the problem as

$$L(g, \zeta', \theta, \{\nu_{i,n}\}) = \int S(g(\mathbf{u}), \zeta', \theta, \{\nu_{i,n}\})dR(\mathbf{u}) + V(\zeta', \theta, \{\nu_{i,n}\}) \tag{5.4.11}$$

where

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

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

By applying the Euler-Lagrange equation [85], we characterize necessary conditions for the function g to be an extreme point for (5.4.11). If g^* is an extreme point of (5.4.11), then the necessary *stationarity condition* holds, i.e.,

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

from which we obtain

$$g^*(\mathbf{u}) = \frac{dQ}{dR}(\mathbf{u}) = \exp\Big[(\zeta' - 1) + \sum_{(i,n)\in I} \nu_{i,n} u_i^n + \theta d(\Psi_X(\mathbf{u}_x), \Psi_Y(\mathbf{u}_y))\Big], \tag{5.4.12}$$

which is equivalent to (5.2.6) by considering $\zeta = \zeta' - 1$.

To determine the optimal values of the Lagrangian multipliers $\zeta', \theta, \{\nu_{i,n}\}$, we leverage Lagrangian duality theorem [56] and define the dual problem as

$$\begin{aligned}
&\max_{(\zeta',\theta,\{\nu_{i,n}\}_{(i,n)\in I})} L(g^*, \zeta', \theta, \{\nu_{i,n}\}_{(i,n)\in I}) \\
&= \min_{(\zeta,\theta,\{\nu_{i,n}\}_{(i,n)\in I})} -L(g^*, \zeta + 1, \theta, \{\nu_{i,n}\}_{(i,n)\in I}) \\
&= (5.2.7).
\end{aligned}$$

This concludes the proof.

5.4.6 Proof of Lemma 5.2.2

The proof makes use of the Hessian matrix of the optimization problem in (5.2.7). Specifically, define the vector $\mathbf{l} \triangleq ((\zeta, \theta, \{\nu_{i,n}\}_{(i,n)\in I})) \in \mathbb{R}^{2+2d\cdot N}$ and the mapping $\omega(\mathbf{u}) \triangleq (1, d(\Psi_X(\mathbf{u}_x), \Psi_X(\mathbf{u}_y)), u_1, \ldots, u_1^N, \ldots, u_{2d}^N)$. Then, the Hessian of the optimization problem (5.2.7) 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 (5.2.7). This completes the proof.

5.4.7 Proof of Theorem 5.2.4

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

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

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

$$\begin{aligned}
R_{PR}(D) &= h(X) - \max_{p_{Y|X}\in\mathcal{H}(p_X,D)} h(X|Y) \\
&\overset{(a)}{=} h(X) - \max_{p_{Y|X}\in\mathcal{H}(p_X,D)} h(Y|X) \\
&\overset{(b)}{\geq} h(X) - \max_{p_Z\in\mathcal{S}} \max_{p_{Y|Z}\in\mathcal{H}(p_Z,D)} h(Y|Z)
\end{aligned} \tag{5.4.13}$$

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 $\sigma^2_{Y|Z} = \sigma^2_Z(1-\rho^2)$, hence the constraint set $\mathcal{H}(p_Z, D)$ can be simplified to

$$\begin{aligned} &\mathbb{E}_{p_Z p_{Y|Z}}\left[\|Z-Y\|^2\right] = 2\sigma_Z^2(1-\rho) \leq D \\ \implies &\sigma^2_{Y|Z} \leq \sigma_Z^2\left(1-\left(1-\frac{D}{\sigma_Z^2}\right)^2\right) = D'. \end{aligned} \tag{5.4.14}$$

Since (5.4.14) constraints only the second moment of the distribution $p_{Y|Z}$, we can infer that the maximum $h(Y|Z)$ is attained by $p_{Y|Z} \sim N(0, D')$. Furthermore, since (5.4.14) 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 $p_Z \sim N(0, \sigma^2)$, 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

$$\max_{p_Z \in \mathcal{S}} \max_{p_{Y|Z} \in \mathcal{H}(p_Z, D)} h(Y|Z) = \max_{p_{Y|X^*} \in \mathcal{H}(p_{X^*}, D)} h(Y|Z) \tag{5.4.15}$$

$$= R^G_{PR}(D) - h(X^*) \tag{5.4.16}$$

which, together with (5.4.13) yields (5.2.9).

For the general vector case, we consider that for every source $Z \sim p_Z \in \mathcal{S}$, we have marginals $\{Z_i\}_{i=1,\dots,N}$ with variance $\sigma^2_{Z_i} = \lambda_i$, where $\{\lambda_i\}_{i=1,\dots,N}$ is the set of eigenvalues of Σ. Then, from (5.4.13) we obtain

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

$$= -h(X^*) + \max_{D_i : \sum_i^N D_i = D} \sum_{i=1}^{N} R_{PR}^{G,i}(D_i) \tag{5.4.17}$$

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 (5.4.16), 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 [23, Corollary 3], we see that the second term of (5.4.17) can be shown to be equivalent to the PR-RDPF for the source X^*, therefore showing that (5.4.13), and consequently (5.2.9), also hold in the general vector case. This completes the proof.

Part III:

Distributionally Robust Lossy Source Coding

Chapter 6

Distributionally Robust Lossy Source Coding

This chapter broadens the scope of the RDP tradeoff by relaxing a fundamental assumption made in previous chapters: that the statistical distribution of the information source is known. In practice, this assumption is rarely satisfied, and the true source distribution may only be known up to a certain level of precision. Distributionally robust source coding addresses precisely this challenge, by studying the problem of source coding under statistical uncertainty, where the true source distribution is unknown but assumed to belong to a set of plausible candidates. This setting is highly relevant in a wide range of real-world applications, including the compression of IoT sensor data under variable operating conditions [86] and distributionally resilient deep neural network codecs [87], where ambiguity in the source statistics can significantly degrade coding performance.

Universal variable-length codes that adjust their encoding rate based on the observed source statistic have been a long-standing focus of research in information theory. In the context of lossless compression, Lempel-Ziv codes [88, 89] asymptotically achieve the entropy rate by learning the statistical structure of the data through an expanding dictionary, capturing all the redundancies in the source without requiring prior knowledge of the source distribution. Similar ideas have also been employed to some degree in the context of lossy compression, see e.g. [90, 91]. However, in many applications, the worst-case performance over a class of possible sources is a more representative description of the operative conditions. In this setting, the fundamental compression limit is characterized as a zero-sum game [92] between the designer choosing a coding strategy and nature selecting the worst possible source distribution given the chosen code. This idea is expressed as a min-max optimization problem, known as the compound or robust rate-distortion function (R-RDF) [93]. Depending on the considered class of sources, various characterizations of the R-RDF have been studied in the literature, see e.g., [93–96]. However, although novel extensions of the classical rate-distortion have been introduced for various

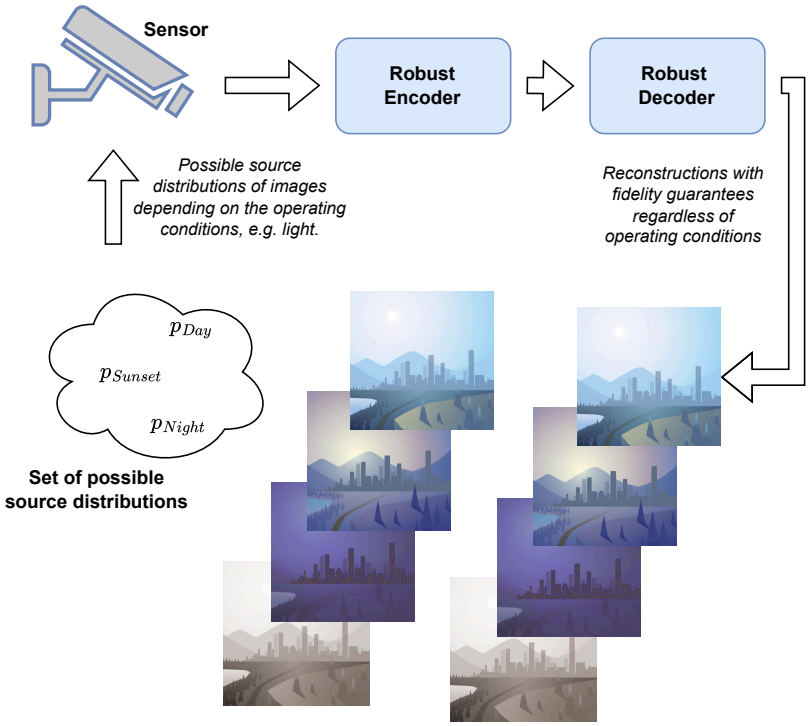


Figure 6.1: A distributionally robust scenario in image compression. Depending on the environmental conditions (weather, light, etc.), the distribution of images observed by the sensor changes, requiring the coding procedure to be resilient to distributional shifts.

applications, their distributionally robust counterparts have remained largely unexplored. Notably, neither the RDPF nor its generalization, the Information Rate Function (IRF) [31], have been extended to the robust setting. As shown in [31], both problems have operational significance when considering an encoding-decoding scheme with access to a (possibly infinite) source of common randomness. The proposed coding theorems leverage the common randomness to construct coding schemes based on the so-called *Strong Functional Representation Lemma (SFRL)* [32]. However, the extension to the robust setting of this code construction has not yet been investigated.

In this chapter, we propose a distributional robust formulation of the IRF (and, by extension, of the RDPF), discussing both coding and computational aspects of the problem. To this end, we first propose two novel extensions of the SFRL. For a fixed conditional distribution $Q_{Y|X}$, we consider the couplings (X, Y) induced by $X \sim \mu_X$ with μ_X belonging to a set $\mathcal{S}$ of possible distributions. Considering the cases where $\mathcal{S}$ is either a finite set of distributions (Theorem 6.2.1) or a Kullback–Leibler divergence sphere centered at a fixed nominal distribution (Theorem 6.2.2), we show that, for any $\mu_X \in \mathcal{S}$, Y can be expressed as a deterministic function of X and an auxiliary random variable Z. Then,

considering the operational setup in [31], we use the derived robust SFRL extensions to characterize distributionally robust coding schemes for both the one-shot (Theorem 6.3.1) and asymptotic regime (Theorem 6.3.3), generalizing previous results present in the literature [31, 93]. Focusing on the R-RDF defined on abstract spaces, we then consider the specific case where μ_X belongs to a given KL-Sphere. We first show that, in this setting, the R-RDF is equivalent to a more tractable max-min problem (Theorem 6.4.1), which allows us to derive an implicit characterization of the points attaining the R-RDF curve (Theorem 6.4.2). The derived solutions are later used to implement a novel algorithm for the computation of the R-RDF in the specific case of finite alphabet sources (Algorithm 5). We corroborate our theoretical results using numerical simulations, comparing the estimation capabilities of the proposed algorithm against the closed form R-RDF derived for binary sources (Theorem 6.4.3).

Notation Given a measurable space $(\mathcal{X}, \Sigma_{\mathcal{X}})$, let $\mathcal{M}_1(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. For $\mu \in \mathcal{M}_1(\mathcal{X})$, a function $f \in L_p(\mu)$ if f is a real-valued measurable function and $(\int_{\mathcal{X}} |f|^p d\mu)^{\frac{1}{p}} < \infty$. Given a second measurable space $(\mathcal{Y}, \Sigma_{\mathcal{Y}})$, we indicate the set of conditional distributions with $\mathcal{L}(\mathcal{X} \times \mathcal{Y}) \triangleq \{Q \mid Q : \mathcal{X} \to \mathcal{M}_1(\mathcal{Y}) \text{ s.t. } \forall A \in \Sigma_{\mathcal{Y}},\ Q(A|\cdot) \text{ is measurable}\}$. Given $\mu_X \in \mathcal{M}_1(\mathcal{X})$ and a conditional measure $Q_{Y|X} \in \mathcal{L}(\mathcal{X} \times \mathcal{Y})$ describing the coupling $(X, Y) \sim \mu_X \otimes Q_{Y|X}$, we indicate their mutual information as either $I(X, Y)$ or $I(\mu_X, Q_{Y|X})$, depending on the context. Lastly, given $\mathcal{S} \subset \mathcal{M}_1(\mathcal{X})$ and a conditional measure Q, we indicate with $\bar{I}(\mathcal{S}, Q) = \sup_{\mu \in \mathcal{S}} I(\mu, Q)$.

6.1 Preliminaries

In this section, we provide the main assumptions regarding our setup, as well as the operational definitions identifying "good" distributionally robust source codes. Although not necessary, throughout this paper, we assume that the source alphabet $\mathcal{X}$ and reconstruction alphabet $\mathcal{Y}$ are Polish spaces. This choice derives from the observation that the majority of source coding applications consider sources defined on a metrizable space. In the following, we first define source code.

Definition 6.1.1. *(Source code) For sets $\mathcal{X}$ and $\mathcal{Y}$, we define a (possibly stochastic) N-block encoder as any function in $\mathcal{F}_N = \{f : \mathcal{X}^N \times \mathcal{Z} \to \mathbb{N}_0\}$, where $\mathcal{Z}$ denotes the domain of the auxiliary source of randomness. Similarly, a (stochastic) N-block decoder is a function in $\mathcal{G}_N = \{g : \mathbb{N}_0 \times \mathcal{Z} \to \mathcal{Y}^N\}$. An element of $\mathcal{F}_N \times \mathcal{G}_N$ is called an N-block code.*

In this paper, we focus on the case of *distributionally robust source coding*, i.e., the statistic of the information source is not perfectly known a priori, but is assumed to belong to a set of possible sources $\mathcal{S}$. Therefore, in what follows, we provide the definitions of an *achievable* code, differentiating between the one-shot and block coding settings.

Definition 6.1.2. *(One-Shot Achievability) Given a set of sources $\mathcal{S}$ and a set of fidelity constraints, the rate R is robust one-shot achievable if there exist an encoder $f \in \mathcal{F}_1$, a decoder $g \in \mathcal{G}_1$ and a random variable Z valued on $\mathcal{Z}$ such that, for any $\mu_X \in \mathcal{S}$ defining the source $X \sim \mu_X$,*

$$M = f(X, Z) \quad \textit{and} \quad Y = g(M, Z)$$

guarantee that the conditional entropy $H(M|Z) \leq R$ while the induced joint distribution $P_{X,Y}$ satisfies the given constraints.

Definition 6.1.3. *(Asymptotic Achievability) Given a set of sources $\mathcal{S}$ and a set of fidelity constraints, the rate R is asymptotically achievable if there exist an sequence of codes $(f_N, g_N) \in \mathcal{F}_N \times \mathcal{G}_N$ and a random variable Z valued on $\mathcal{Z}$ such that, for any $\mu_X \in \mathcal{S}$ inducing the independent and identically distributed sequence $X^N = (X_1, \dots, X_N)$ with $X_i \sim \mu_X$,*

$$M = f_N(X^N, Z) \quad \textit{and} \quad Y^N = g_N(M, Z)$$

guarantee that, for N large enough, the conditional entropy $\frac{H(M|Z)}{N} \leq R$ while the induced joint distribution P_{X_i,Y_i} satisfies the given constraints for all $i \in 1 : N$.

6.2 Robust Strong Functional Representation Lemmas

Our first technical results focus on the extension of the SFRL [32] to the distributionally robust setting, where, for a fixed conditional distribution Q, the uncertainty on the real distribution of the coupling (X, Y) derives from the uncertainty on the distribution of the marginal X, i.e., μ_X can be any of the distributions belonging to a set $\mathcal{S}$. We call this extension *Robust SFRL* (R-SFRL) and it will be the main tool for the derivation of the new robust source coding theorems in Section 6.3.

With the following theorem, we start by characterizing the case where $\mathcal{S}$ is a finite collection of distributions.

Theorem 6.2.1. *(Robust SFRL) Let $\mathcal{S} = \{\mu_1, \dots, \mu_{|\mathcal{S}|}\}$ be a finite set of distributions and, given a conditional distribution Q, let $(X_\alpha, Y_\alpha) \sim \mu_\alpha \otimes Q$ with $I(X_\alpha, Y_\alpha) < \infty$, for all $\alpha \in 1 : |\mathcal{S}|$. Then, there exists a random variable Z such that, for all $\alpha \in 1 : |\mathcal{S}|$, Z is independent of X_α ($X_\alpha \perp Z$) and Y_α can be expressed as a deterministic function $g(X_\alpha, Z)$ of (X_α, Z) such that*

$$H(Y_\alpha|Z) \leq \bar{\mathrm{I}}(\mathcal{S}, Q) + \log\Big(|\mathcal{S}| \cdot (\bar{\mathrm{I}}(\mathcal{S}, Q) + 1)\Big) + 4. \tag{6.2.1}$$

Proof. See Appendix 6.6.1. □

A second interesting scenario is the case where $\mathcal{S}$ is a KL divergence sphere of radius Γ centered in some nominal distribution μ_0, where the belief on the nominal distribution

being the real distribution is inversely proportional to Γ. The following theorem extends the SFRL to this case.

Theorem 6.2.2. *(KL-Robust SFRL) Given a nominal distribution μ_0 and $\Gamma \geq 0$, let $\mathcal{S}(\Gamma) = \{\mu \in \mathcal{M}_1(\mathcal{X}) : D_{KL}(\mu\|\mu_0) \leq \Gamma\}$. Assume that, for a given conditional distribution Q and for all $\mu \in \mathcal{S}$ such that $(X,Y) \sim \mu \otimes Q$, $I(X,Y) < \infty$. Then, there exist a random variable Z such that, for any $X \sim \mu$ with $\mu \in \mathcal{S}(\Gamma)$, Z is independent of X $(X \perp Z)$ and Y can be expressed as a function of $g(X,Z)$ of (X,Z) such that*

$$H(Y|Z) \leq \bar{\mathrm{I}}(\mathcal{S},Q) + \Gamma + \log(\bar{\mathrm{I}}(\mathcal{S},Q) + \Gamma + 1) + 4. \tag{6.2.2}$$

Proof. See Appendix 6.6.2. □

We conclude this section with the following observation.

Remark 6.2.1. *(Special cases) If in Theorem 6.2.1 we consider a singleton set $S = \{\mu_0\}$ or in Theorem 6.2.2 a KL-sphere with radius $\Gamma = 0$, then both Theorems 6.2.1 and 6.2.2 recover the original SFRL bound [32].*

6.3 Robust Lossy Source Coding Theorems

In this section, we study the application of the R-SFRL to the robust source coding problem. To this end, we first introduce a robust version of the information rate function (IRF), first described in [31].

Definition 6.3.1. *(Robust Information Rate Function (R-IRF)) Let $\mathcal{S} \subset \mathcal{M}_1(\mathcal{X})$ be a set of probability distributions, $\{\ell_i\}_{i=1}^M$ be a finite set of real-valued functions on the space $\mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$ of joint distributions $P_{X,Y}$, and $\theta \in \mathbb{R}^M$. Then, the robust IRF is defined as*

$$\begin{aligned} \bar{R}(\theta) &\triangleq \inf_{Q_{Y|X} \in \mathcal{L}(\theta)} \bar{\mathrm{I}}(S, Q_{Y|X}) \\ &= \inf_{Q_{Y|X} \in \mathcal{L}(\theta)} \sup_{\mu_X \in \mathcal{S}} I(\mu_X, Q_{Y|X}) \end{aligned} \tag{6.3.1}$$

where the set of conditional distributions $\mathcal{L}(\theta) \subseteq \mathcal{L}(\mathcal{X} \times \mathcal{Y})$ is defined as

$$\mathcal{L}(\theta) \triangleq \left\{ Q_{Y|X} : \ \ell_i[\mu_X Q_{Y|X}] \leq \theta_i \ , \forall \mu_X \in \mathcal{S}, i \in 1:M \right\}. \tag{6.3.2}$$

The definition of R-IRF is based on a min-max formulation of the IRF. For target constraint levels θ, the uncertainty set $\mu_X \in \mathcal{S}$ can be seen as the nature trying to maximize the rate, while the designer selects a code $Q_{Y|X} \in \mathcal{L}(\theta)$ trying to minimize it. The following remark elaborates on the generality of the R-IRF definition.

Remark 6.3.1. *Similarly to the IRF, the R-IRF provides a general description for a wide variety of information-theoretic quantities. For example, let ℓ_1 and ℓ_2 be defined as*

$$\ell_1(P_{X,Y}) = \mathbb{E}[d(X,Y)] \quad \ell_2(P_{X,Y}) = d(\mu_X, q_Y) \tag{6.3.3}$$

for a proper choice of distortion metric d and divergence function d. Then, the classical R-RDF [93] can be obtained by considering the set $\{\ell_1\}$, whereas the set $\{\ell_1, \ell_2\}$ can be seen as a novel robust extension of the RDPF.

Due to its generality, the R-IRF is not guaranteed to have a non-empty feasible solution for all constraint levels $\theta \in \mathbb{R}_+^M$. To ease our subsequent analysis, we will operate under the following assumption, leaving the investigation of the feasibility conditions of the R-IRF for future work.

Assumption 6.3.1. *We consider the set of possible constraint levels*

$$\Theta = \{\theta \in \mathbb{R}^M \mid \exists Q_{Y|X} \in \mathcal{L}(\theta) \text{ s.t. } \bar{\mathrm{I}}(S, Q_{Y|X}) < \infty\},$$

thus considering only θ for which $\mathcal{L}(\theta)$ has at least a feasible point. Hence, for $\theta \in \Theta$, the optimization problem (6.3.1) *defining the R-IRF is feasible, i.e., $\bar{R}(\theta) < \infty$.*

Under Assumption 6.3.1, we are now ready to present our achievability and converse results relative to the one-shot case.

Theorem 6.3.1. *Let $\mathcal{S} = \{p_1, \dots, p_{|\mathcal{S}|}\}$ be a finite set of distributions, $\{\ell_i\}_i^M$ be a set of constraints and θ the associated constraint levels. Then, R is robust one-shot achievable if*

$$R \geq \bar{R}(\theta) + \log\Big(|\mathcal{S}| \cdot (\bar{R}(\theta) + 1)\Big) + 4 \tag{6.3.4}$$

Proof. See Appendix 6.6.3. □

Theorem 6.3.2. *Let $\mathcal{S} = \{p_1, \dots, p_{|\mathcal{S}|}\}$ be a finite set of distributions, $\{\ell_i\}_i^M$ be a set of constraints and θ the associated constraint levels. Then, if R is robust one-shot achievable, then $R \geq \bar{R}(\theta)$.*

Proof. See Appendix 6.6.4. □

Unfortunately, the results of Theorems 6.3.1 and 6.3.2 do not provide a tight characterization of the one-shot rate region. However, when extending the analysis to the asymptotic setting, the rate region becomes well characterized, as shown in the following theorem.

Theorem 6.3.3. *Let $\mathcal{S} = \{p_1, \dots, p_{|\mathcal{S}|}\}$ be a finite set of distributions, $\{\ell_i\}_i^M$ be a set of constraint and θ the associated constraints levels. Then, R is asymptotically achievable if and only if $R \geq \bar{R}(\theta)$.*

Proof. See Appendix 6.6.5. □

Theorem 6.3.3 recovers classical results in robust source coding, extending them to the case of variable-length coding with common randomness. On the other hand, Sakrison

in [93] recovered the same rate region by considering fixed-rate codes and average single-letter distortion constraint for either finite or compact uncertainty sets $\mathcal{S}$. Given the similarities with [93], we believe that Theorem 6.3.3 could also be extended to consider source sets $\mathcal{S}$ satisfying the same notion of compactness.

6.4 Case Study: Robust RDF

In this section, we focus our attention on a specific instance of the R-IRF, considering the classical problem of the distributionally robust RDF (R-RDF). More specifically, we will consider the case where the unknown true source probability measure is assumed to belong to the set

$$\mathcal{S}(\Gamma) \triangleq \{\mu \in \mathcal{M}_1(A) : D_{KL}(\mu\|\mu_0) \leq \Gamma\}, \quad \Gamma \in [0, +\infty)$$

i.e., a KL-sphere of radius Γ centered $\mu_0 \in \mathcal{M}_1(A)$.

Given the set $\mathcal{S}(\Gamma) \subset \mathcal{M}_1(\mathcal{X})$ and a distortion function $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$, the set of stochastic kernels $\mathcal{L}(\theta)$ defined in (6.3.2) becomes

$$\begin{aligned} \mathcal{L}_\mu(D) &\triangleq \left\{Q \in \mathcal{L}(\mathcal{X} \times \mathcal{Y}) : \int d(x,y)Q(dy|x)\mu(dx) \leq D\right\} \\ \mathcal{L}(D) &\triangleq \bigcap_{\mu \in \mathcal{S}(\Gamma)} \mathcal{L}_\mu(D) \end{aligned} \tag{6.4.1}$$

where $\mathcal{L}_\mu(D)$ is equivalent to the constraint set of the classical RDF for a fixed source μ. Consequently, the constraint set of the R-RDF can be expressed as the intersection of the constraint set associated with each source present in the considered class $\mathcal{S}(\Gamma)$. Therefore, following Definition 6.3.1 with the described constraint set, the R-RDF is defined as

$$R^+(D) \triangleq \inf_{Q \in \mathcal{L}(D)} \sup_{\mu \in \mathcal{S}(\Gamma)} I(\mu, Q). \tag{6.4.2}$$

The min-max structure of (6.4.2), and of the R-IRF in general, can present numerous challenges in terms of computational complexity. For this reason, it is often preferable to consider an alternative max-min formulation of the problem, i.e.,

$$R^-(D) \triangleq \sup_{\mu \in \mathcal{S}(\Gamma)} \inf_{Q \in \mathcal{L}(D)} I(\mu, Q) \tag{6.4.3}$$

which has the added advantage of sharing a similar structure to the classical RDF, i.e., the inner inferior over $Q \in \mathcal{L}(D)$. However, in general, $R^+(D,P) \geq R^-(D,P)$, where equality holds if the conditions of the Min-Sup Theorem[1] hold for the considered problem [97]. In the following theorem, we provide general conditions under which (6.4.2) satisfies the min-sup conditions, i.e., (6.4.2) is equivalent to (6.4.3).

[1] Sometimes referred to also as Von Neumann condition [97].

Theorem 6.4.1. *Let $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_0^+$ be a measurable distortion function such that, for $x \in \mathcal{X}$, $d(x, \cdot)$ is continuous on $\mathcal{Y}$, μ-almost everywhere. Then, there exists $Q^* \in \mathcal{L}(D)$ such that*

$$R^-(D) = \sup_{\mu \in \mathcal{S}(\Gamma)} I(\mu, Q^*) = R^+(D)$$

Proof. See Appendix 6.6.6. □

Using the result of Theorem 6.4.1, we devote the rest of this section to the characterization of the minimizers achieving $R^-(D)$. To this end, we first recall a known result to characterize the RDF achieving conditional distribution for a fixed source on abstract spaces [98].

Lemma 6.4.1. *For a fixed $\mu \in \mathcal{S}(\Gamma)$ and any $s \geq 0$, the inner infimum in (6.4.3) is attained by $Q^* \in \mathcal{L}(D)$ given by*

$$Q^*(A|x) = \frac{\int_A e^{-sd(x,y)} q^*(dy)}{\int_{\mathcal{Y}} e^{-sd(x,y)} q^*(dy)} \tag{6.4.4}$$

where $A \subseteq \mathcal{Y}$ and $q^* \in \mathcal{M}_1(\mathcal{Y})$ is the marginal on $\mathcal{Y}$ of the measure $\mu \otimes Q$. Furthermore, $\inf_{Q \in \mathcal{L}(D)} I(\mu, Q) = F(q^*)$, where

$$F(q^*) = \sup_{s \geq 0} \left[-sD - \int_{\mathcal{A}} \log \left(\int_{\mathcal{Y}} e^{-sd(x,y)} q^*(dy) \right) \mu(dx) \right].$$

From Lemma 6.4.1, (6.4.3) can be expressed as

$$R^-(D) = \sup_{s \geq 0} \sup_{\mu \in \mathcal{S}(\Gamma)} \left[-sD - \int_{\mathcal{A}} \log \left(\int_{\mathcal{Y}} e^{-sd(x,y)} q^*(dy) \right) \mu(dx) \right] \tag{6.4.5}$$

where the suprema over $\mu \in \mathcal{S}(\Gamma)$ and $s \geq 0$ are interchangeable. We can now introduce the Lagrangian functional for the constraint set $\mathcal{S}(\Gamma)$, i.e.,

$$F_{s,\lambda}(\mu, q^*) \triangleq \Big[-sD - \lambda \left(D_{KL}(\mu \| \mu_0) - \Gamma \right) - \int_{\mathcal{A}} \log \left(\int_{\mathcal{Y}} e^{-sd(x,y)} q^*(dy) \right) \mu(dx) \Big]$$

where $\lambda \geq 0$ is the associated Lagrangian multiplier. Let $\mu^* = \arg\sup_{\mu \in \mathcal{M}_1(\mathcal{X})} F_{s,\lambda}(\mu, q^*)$. It can be shown, via Lagrange duality, that

$$R^-(D) = (6.4.5) = \sup_{s \geq 0} \inf_{\lambda \geq 0} F_{s,\lambda}(\mu^*, q^*). \tag{6.4.6}$$

In the next theorem, we use (6.4.6) to provide the complete implicit solution of the $R^-(D)$ problem.

Theorem 6.4.2. *Assume* $e^{\frac{g}{\lambda}} \in L_1(\mu_0)$ *and* $ge^{\frac{g}{\lambda}} \in L_1(\mu_0)$*, where* $g : \mathcal{X} \to \mathbb{R}$*,* $g(x) = -\log\left(\int_{\mathcal{Y}} e^{sd(x,y)} q^*(dy)\right)$*. Then, the solution of* (6.4.6) *is given by*

$$R^-(D) = sD + \lambda\Gamma + \lambda \log\left(\int_{\mathcal{X}} \left(\int_{\mathcal{Y}} e^{sd(x,y)} q^*(dy)\right)^{-\frac{1}{\lambda}} \mu_0(dx)\right) \tag{6.4.7}$$

where $s \geq 0, \lambda \geq 0$ *are, respectively, the Lagrangian multipliers associated with the distortion constraint and uncertainty set constraint. Furthermore, the infimum over* $Q \in \mathcal{L}(D)$ *is attained by* (6.4.4)*, while the supremum over the class of sources* $\mu \in \mathcal{S}(\Gamma)$ *is attained by*

$$\mu^*(dx) = \frac{\left(\int_{\mathcal{Y}} e^{sd(x,y)} q^*(dy)\right)^{-\frac{1}{\lambda}} \mu_0(dx)}{\int_{\mathcal{X}} \left(\int_{\mathcal{Y}} e^{sd(x,y)} q^*(dy)\right)^{-\frac{1}{\lambda}} \mu_0(dx)}. \tag{6.4.8}$$

Proof. See Appendix 6.6.7. □

We stress the following remark on the relation between the Lagrangian multipliers (s, λ) and the constraint levels (D, Γ).

Remark 6.4.1. *Similarly to the classical RDF [40, 47], assuming fixed the Lagrangian multipliers* (s, λ) *rather then the constraint levels* (D, Γ) *greatly simplifies the derivation of the optimal solutions* (μ^*, Q^*)*. Although it can be proven the existence of an inversely proportional relation between Lagrangian multiplies and constraint levels, the absence of an analytical mapping implies the necessity of search routines to identify the* (s, λ) *inducing the specific levels* (D, Γ)*.*

6.4.1 Computation of the Robust RDF for Discrete Sources

We devote this section to the application of the results of Theorem 6.4.2 for the computation of the R-RDF, focusing on the case of discrete sources, i.e., $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. To this end, we exploit the fact that a pair (μ^*, Q^*) solving the system of equations (6.4.4)-(6.4.8) is guaranteed to achieve $R^-(D)$. However, to simplifying the derivation of an algorithmic solution, we first introduce an alternative to (6.4.4) characterizing the optimal marginal distribution q^*. Considering the source μ^* fixed, it can be shown that the marginal q^* must satisfy

$$c(y) = \sum_{x \in \mathcal{X}} \frac{e^{-sd(x,y)}}{\sum_{y \in \mathcal{Y}} e^{-sd(x,y)} q^*(dy)} \mu^*(dx) \leq 1, \tag{6.4.9}$$

holding with equality $\forall y \in \mathcal{Y}$ such that $q^*(y) > 0$ [47]. Substituting (6.4.8) in (6.4.9), we can obtain an implicit equation depending only on the marginal distribution q^*, i.e.,

$$c(y) = \frac{1}{Z} \sum_{x \in \mathcal{X}} \frac{e^{-sd(x,y)}}{\left(\sum_{y \in \mathcal{Y}} e^{-sd(x,y)} q^*(y)\right)^{1+\frac{1}{\lambda}}} \mu_0(x)$$

$$Z = \sum_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{Y}} e^{sd(x,y)} q^*(dy) \right)^{-\frac{1}{\lambda}} \mu_0(x).$$

Unfortunately, verifying the inequality condition in (6.4.9), hence ensuring the optimality of q^*, remains challenging from a computational standpoint. To partially solve this issue, we introduce the following simplified implicit function

$$F : \mathcal{M}_1(\mathcal{Y}) \to \mathbb{R}^{|\mathcal{Y}|}, \; F(q) = \left[q(y)(1 - c(y)) \right]_{y \in \mathcal{Y}}. \tag{6.4.10}$$

It is easy to check that $F(q) = 0$ represents a necessary optimality condition, meaning that the set of roots of F contains the optimal marginal q^*. The advantage of this formulation lies in the possibility of applying a root-finding method to characterize a smaller set of possible q, which can then be tested for optimality using (6.4.9). Algorithm 5 implements the ideas described so far, considering Newton's method [48] as a root-finding routine.

Algorithm 5 Discrete Robust RDF Estimation

Require: Nominal distribution μ_0; Set of initial points $\{q_i\}_{i=1}^N$; Lagrangian multipliers $s \geq 0, \lambda \geq 0$; Jacobian $J_F(\cdot)$; Error tolerance $\epsilon \geq 0$;

```
Op = {}
for i = 1, ..., N do
    do                                              ▷ Newton's Method
        η ← (J_F(q_i))^{-1} F(q_i)
        q_i ← q_i − η
    while ‖η‖ ≥ ε
    if q_i satisfies (6.4.9) then
        Op ← {q_i} ∪ Op
    end if
end for
```

Ensure: Optimal Marginals Op.

Remark 6.4.2. *The implicit function F can be shown to be non-monotonic for some values of (s, λ). In these cases, the convergence of Algorithm 5 heavily relies on the choice of the initialization points used in the root-finding routine. Although the implementation of discretization-based schemes, e.g., [99], can guarantee the convergence of the algorithm, we empirically found that considering a single initial point, i.e.,*

$$q(y) \propto \sum_{x \in \mathcal{X}} \frac{e^{-sd(x,y)} (\mu_0(x))^{\frac{\lambda}{\lambda+1}}}{\left(\sum_{y \in \mathcal{Y}} e^{-sd(x,y)}\right)^{\frac{\lambda+1}{\lambda}}}$$

provides good results in most cases, as shown in Section 6.5.

To test the efficacy of Algorithm 5, we use as a benchmark the following theorem, which provides the explicit characterization of the R-RDF under Hamming distortion for a class of Bernoulli sources.

Theorem 6.4.3. *(R-RDF for Bernoulli sources) Let* $\mathcal{X} = \{0,1\}$ *and* $\mathcal{S}(\Gamma) = \{\mu \in \mathcal{M}_1(\mathcal{X}) : D_{KL}(\mu\|\mu_0) \leq \Gamma\}$ *for some* $\mu_0 = Bern(\beta_0)$*. Then, the R-RDF* $R^-(D)$ *under Hamming distortion* $d(x,y) = \mathbb{1}_{(x\neq y)}$ *becomes*

$$R^-(D) = h_b(\beta^*) - h_b(D), \quad \beta^* = \left(1 + \left(\frac{1-\beta_0}{\beta_0}\right)^{\frac{\lambda}{1+\lambda}}\right)^{-1}$$

where $h_b(\cdot)$ *is the binary entropy function and* $\lambda \geq 0$ *is chosen such that* $D_{KL}(\mu\|\mu_0) \leq \Gamma$ *with* $\mu = Bern(\beta^*)$*.*

Proof. See Appendix 6.6.8. □

6.5 Numerical Examples

In this section, we provide empirical evidence of the performance of Algorithm 5 via numerical simulations.

Example 6.5.1. *Suppose* $\mathcal{X} = \{0,1\}$ *and let* $\mu_0 = Bern(0.1)$ *be the given nominal distribution. In Fig. 6.2, we display the point estimate of the* $R^-(D)$ *under Hamming distortion* $d(x,y) = \mathbb{1}_{(x\neq y)}$ *resulting from Algorithm 5, comparing it to the analytical expression recovered in Theorem 6.4.3. The recovered surface shows the expected convex-concave behavior in* (D,Γ)*. Furthermore, we remark that the estimate of Algorithm 5 perfectly recovers the analytical solution.*

Example 6.5.2. *Suppose* $|\mathcal{X}| = 10$ *and let*

$$\mu_0 = [0.0006, 0.0003, 0.09, 0.4262, 0.0002, 0.1409, 0.0019, 0.0029, 0.3358, 0.0012]$$

be the given nominal distribution. Fig. 6.3 shows the point estimate of the $R^-(D)$ *under Hamming distortion resulting from Algorithm 5 and the associated interpolated surface. We remark that, with the considered initialization conditions, Algorithm 5 fails to identify suitable solutions for larger values of* Γ*, i.e.,* $\lambda \to 0$*. However, as pointed out in Remark 6.4.2, testing a larger pool of initial points would solve this issue at the expense of a higher computational complexity.*

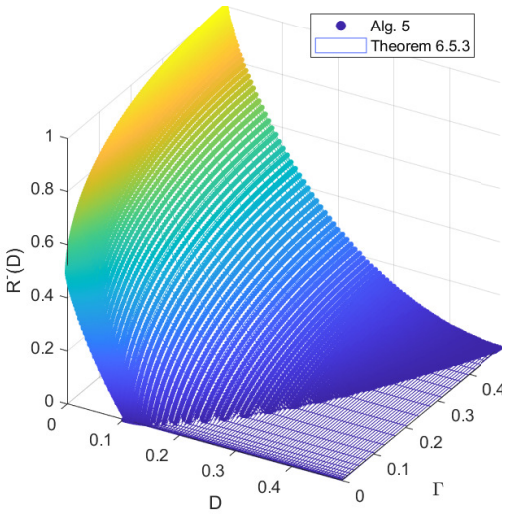


Figure 6.2: Robust RDF for with nominal distribution $\mu_0 = Bern(0.1)$.

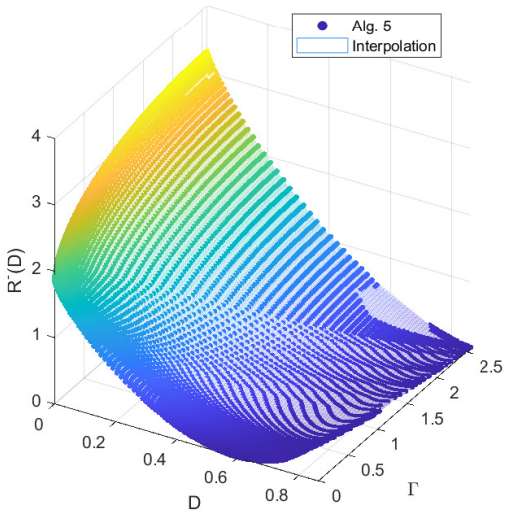


Figure 6.3: Robust RDF for with nominal distribution μ_0.

6.6 Appendix

6.6.1 Proof of Theorem 6.2.1

To ease the notation, for every $\mu_\alpha \in \mathcal{S}$ inducing $(X_\alpha, Y_\alpha) \sim \mu_\alpha \otimes Q$, we will indicate with q_α the associated marginal distribution of Y_α. Furthermore, we will indicate with $\mathcal{S}' = \{q_\alpha\}_{\alpha=1}^{|\mathcal{S}|}$ the set of such marginals.

Let $\Theta = \{\Theta_\alpha\}_{\alpha=1}^{|\mathcal{S}|}$ where $\Theta_\alpha = \{(T_{\alpha,i}, \tilde{Y}_{\alpha,i})\}_{i=1}^{\infty}$ is a non-homogeneous Poisson process (P. P.) with intensity measure $q_\alpha \times \lambda$, where λ is the Lebesgue measure. For any fixed $x \in \mathcal{X}$ and any $\alpha = 1, \dots, |\mathcal{S}|$, define the mapping

$$f_\alpha : (\tilde{y}, t) \to \left(y, t \cdot \frac{dq_\alpha}{dQ(\cdot|x)}(\tilde{y}) \right).$$

Similarly to the construction of the SFRL [32, Theorem 1], we apply the mapping f_α to the P.P. Θ_α, resulting in the point process

$$\left\{ \left(\tilde{Y}_{\alpha,i}, T_{\alpha,i} \cdot \frac{dq_\alpha}{dQ(\cdot|x)}(\tilde{Y}_{\alpha,i}) \right) \right\}$$

which can be shown to be a P.P. with intensity measure $Q(\cdot|x) \times \lambda$. We define $\{\nu_\alpha\}$,$\{K_\alpha\}$, K, and A as follows:

$$\begin{aligned}
\nu_\alpha &= \min_i \left\{ T_{\alpha,i} \cdot \frac{dq_\alpha}{dQ(\cdot|x)}(\tilde{Y}_{\alpha,i}) \right\} \\
K_\alpha &= \arg\min_i \left\{ T_{\alpha,i} \cdot \frac{dq_\alpha}{dQ(\cdot|x)}(\tilde{Y}_{\alpha,i}) \right\} \\
K &= \min_\alpha K_\alpha
\end{aligned}$$

$$A = \arg\min_{\alpha} K_\alpha$$

We remark that, for all α, $\nu_\alpha \sim \text{Exp}(1)$ and, conditioned on the realization of ν_α, $\tilde{Y}_{\alpha,K_\alpha} \sim Q(\cdot|x)$.

Now, consider the following coding scheme. Assume Θ to be the available both at the encoder and decoder as *common randomness*, and let $M = (K, A)$ be the encoded variable. The decoder can isolate the P.P. associated with A and output $Y = \tilde{Y}_{A,K} \sim Q(\cdot|x)$. Since $\tilde{Y}_{A,K}$ is a function of (A, K) and the common randomness Θ, we have that:

$$H(Y|\Theta) \leq H(K, A) \leq H(K) + H(A). \tag{6.6.1}$$

We are interested in bounding $H(Y|\Theta)$. Since A takes value on the index set of the possible source distributions $\mathcal{S}$, we can bound $H(A)$ as:

$$H(A) \leq \log(|\mathcal{S}|) \tag{6.6.2}$$

On the other hand, we can derive

$$\begin{aligned}
\mathbb{E}[\log(K)|X = x] &= \mathbb{E}\left[\min_{\alpha} \log(K_\alpha)|X = x\right] \\
&\overset{(a)}{\leq} \min_{\alpha} \mathbb{E}\left[\log(K_\alpha)|X = x\right] \\
&\overset{(b)}{\leq} \min_{q_\alpha \in \mathcal{S}'} D_{KL}(Q(\cdot|x)\|q_\alpha) + c
\end{aligned}$$

where $c = e^{-1}\log(e) + 1$, (a) follows from Jensen's inequality, and (b) leverages the results contained in the proof of [32, Theorem 1]. It follows that, for any $\mu_\beta \in \mathcal{S}$, it holds

$$\begin{aligned}
\mathbb{E}[\log(K)] &\leq \mathbb{E}_{X\sim\mu_\beta}\left[\min_{q_\alpha \in \mathcal{S}'} D_{KL}(Q(\cdot|x)\|q_\alpha) + c\right] \\
&\leq \min_{q_\alpha \in \mathcal{S}'} D_{KL}(\mu_\beta Q\|\mu_\beta q_\alpha) + c \\
&\overset{(c)}{\leq} \max_{\mu_\beta \in \mathcal{S}} \min_{q_\alpha \in \mathcal{S}'} D_{KL}(\mu_\beta Q(\cdot|X)\|\mu_\beta q_\alpha) + c \\
&\overset{(d)}{\leq} \max_{\mu_\beta \in \mathcal{S}} I(\mu_\beta, Q) + c
\end{aligned}$$

where (c) follows from the maximization on all possible sources $\mu_\beta \in \mathcal{S}$ and (d) derives from

$$\begin{aligned}
D_{KL}(\mu_\beta Q(\cdot|X)\|\mu_\beta q_\alpha) &= I(\mu_\beta, Q) + D_{KL}(q_\beta\|q_\alpha) \\
\Rightarrow \min_{q_\alpha} D_{KL}(\mu_\beta Q(\cdot|X)\|\mu_\beta q_\alpha) &= I(\mu_\beta, Q).
\end{aligned} \tag{6.6.3}$$

Leveraging the results in [32, Appendix B] characterizing the maximum entropy distribu-

tion of K under constraint $\mathbb{E}[\log(K)] \leq \bar{\mathrm{I}}(\mathcal{S}, Q) + c$, we obtain that

$$H(K) \leq \bar{\mathrm{I}}(\mathcal{S}, Q) + \log(\bar{\mathrm{I}}(\mathcal{S}, Q) + 1) + 4 \tag{6.6.4}$$

Substituting (6.6.2) and (6.6.4) in (6.6.1), we recover (6.2.1). This concludes the proof.

6.6.2 Proof of Theorem 6.2.2

Similarly to the proof of Theorem 6.2.1, for every $\mu \in \mathcal{S}$ inducing $(X, Y) \sim \mu \otimes Q$, we will indicate with q the marginal distribution associate with Y, while $\mathcal{S}'$ identifies the set of such marginals. Furthermore, let q_0 indicate the marginal distribution associated with the nominal distribution μ_0.

Let $\Theta = \{(T_i, \tilde{Y}_{0,i})\}_{i=1}^{\infty}$ is a non-homogeneous Poisson process (P.P.) with intensity measure $q_0 \times \lambda$, where λ is the Lesbegue measure. For any fixed $x \in \mathcal{X}$, define the mapping

$$f : (\tilde{y}, t) \to \left(y, t \cdot \frac{dq_0}{dQ(\cdot|x)}(\tilde{y}) \right).$$

Similarly to the proof of [32, Theorem 1], we can apply the mapping f_α to the P.P. Θ_α, resulting in the point process

$$\left\{ \left(\tilde{Y}_{0,i}, T_{0,i} \cdot \frac{dq_0}{dQ(\cdot|x)}(\tilde{Y}_{0,i}) \right) \right\}$$

which can be show to be a P.P. with intensity measure $Q(\cdot|x) \times \lambda$. Let ν and K be defined as:

$$\nu = \min_i \left\{ T_{0,i} \cdot \frac{dq_0}{dQ(\cdot|x)}(\tilde{Y}_{0,i}) \right\}$$
$$K = \arg\min_i \left\{ T_{0,i} \cdot \frac{dq_0}{dQ(\cdot|x)}(\tilde{Y}_{0,i}) \right\}$$

We note that, due to the properties of the mapped P.P., $\nu \sim \mathrm{Exp}(1)$ and, conditioned on the realization of ν, $\tilde{Y}_{0,K} \sim Q(\cdot|x)$. Similarly to [32, Theorem 1], the coding scheme assumes the access to Θ at both encoder and decoder. The encoder then will transmit K, while the decoder will output $Y = \tilde{Y}_{0,K}$. Since Y is a function of K and the common randomness Θ, we have that $H(Y|\Theta) \leq H(K)$.
From the results of [32, Theorem 1], we have that

$$\mathbb{E}[\log(K)|X = x] \leq D_{KL}(Q(\cdot|x)\|q_0) + c$$

where $c = e^{-1}\log(e)$. Hence, for any $\mu \in \mathcal{S}$

$$\begin{aligned}\mathbb{E}[\log(K)] &\leq D_{KL}(\mu \otimes Q \| \mu \times q_0) + c \\ &\leq \max_{\mu \in \mathcal{S}} D_{KL}(\mu \otimes Q \| \mu \times q_0) + c \\ &\overset{(a)}{=} \max_{\mu \in \mathcal{S}} I(\mu, Q) + D_{KL}(q \| q_0) + c \\ &\overset{(b)}{\leq} \max_{\mu \in \mathcal{S}} I(\mu, Q) + \Gamma + c\end{aligned}$$

where (a) results from (6.6.3) and (b) stems from from the data processing inequality, i.e., $D_{KL}(q\|q_0) \leq D_{KL}(\mu\|\mu_0) \leq \Gamma$. Leveraging the results in [32, Appendix B] characterizing the maximum entropy distribution of K under constraint $\mathbb{E}[\log(K)] \leq \bar{\mathrm{I}}(\mathcal{S}, Q) + \Gamma + c$, we obtain that

$$H(K) \leq \bar{\mathrm{I}}(\mathcal{S}, Q) + \Gamma + \log(\bar{\mathrm{I}}(\mathcal{S}, Q) + \Gamma + 1) + 4, .$$

This concludes the proof.

6.6.3 Proof of Theorem 6.3.1

The proof follows from the results of Theorem 6.2.1. Under Assumption 6.3.1, the IRF is feasible and thus there exist a $Q_{Y|X}$ such that

$$\forall i : \ell_i(P_{X,Y}) \leq \theta_i, \text{ and } \bar{\mathrm{I}}(S, Q) \leq \bar{R}(\theta) + \epsilon$$

for $\epsilon > 0$. Following the construction of random variables (A, K) detailed in the proof of Theorem 6.2.1, we have that

$$H(A, K) \leq \bar{R}(\theta) + \log\Big(|\mathcal{S}| \cdot (\bar{R}(\theta) + \epsilon + 1)\Big) + 4 + \epsilon \leq R$$

for $\epsilon \to 0$. Since in our encoding scheme $M = (A, K)$, it follows that

$$H(M|\Theta) \leq H(M) = H(A, K) \leq R,$$

thus concluding the proof.

6.6.4 Proof of Theorem 6.3.2

The converse theorem follows the same structure of [31, Theorem 2], observing that the rate R must satisfy the worst-case scenario among all source distributions $\mu_X \in \mathcal{S}$, i.e.,

$$\begin{aligned}R &\geq \max_{\mu_X \in \mathcal{S}} H(f(X, Z)|Z) \\ &\geq \max_{\mu_X \in \mathcal{S}} I(X, f(X, Z)|Z)\end{aligned}$$

$$\geq \max_{\mu_X \in \mathcal{S}} I(X, f(X,Z)|Z) + I(X,Z)$$
$$\geq \max_{\mu_X \in \mathcal{S}} I(X,(Y,Z)) \geq \max_{\mu_X \in \mathcal{S}} I(X,Y) \geq \bar{R}(\theta)$$

This concludes the proof.

6.6.5 Proof of Theorem 6.3.3

Under Assumption 6.3.1, there exists a $Q_{Y|X}$ such that $Q_{Y,X} \in \mathcal{L}(\theta)$ and $\bar{\mathrm{I}}(\mathcal{S}, Q_{Y,X}) < \bar{R}(\theta) + \frac{1}{N}$. To leverage the results of Theorem 6.2.1, we can construct the conditional distribution as the product measure

$$Q_{Y^N|X^N} = \prod_{i=i}^{N} Q_{Y|X}$$

guaranteeing that, for all $\mu_X \in \mathcal{S}$, $(X_n, Y_n) \sim \mu_X \otimes Q_{Y|X}$ respects the IRF constraints. Furthermore the product structure implies

$$I(X^N, Y^N) \leq N\, \bar{\mathrm{I}}(\mathcal{S}, Q_{Y|X}). \tag{6.6.5}$$

It follows that, similarly to Theorem 6.3.1, the average entropy of the message $M = (K_N, A_N)$ is at most

$$\begin{aligned}\frac{H(M)}{N} &\leq \frac{1}{N}\left(N\,\bar{\mathrm{I}}(\mathcal{S}, Q_{Y|X}) + \log\left(|\mathcal{S}| \cdot (N\,\bar{\mathrm{I}}(\mathcal{S}, Q_{Y|X}) + 1)\right) + 4\right) \\ &\leq \bar{R}(\theta) + \frac{1}{N}\log\left(|\mathcal{S}| \cdot (N\bar{R}(\theta) + 2)\right) + \frac{5}{N}\end{aligned}$$

Therefore, for $N \to \infty$, $\frac{H(M)}{N} \to \bar{R}(\theta)$, proving the achievability of the coding scheme.

The converse of the coding theorem can be recovered with a similar strategy to the proof of Theorem 6.3.2, i.e.,

$$R \geq \lim_{N\to\infty} \max_{\mu_X \in \mathcal{S}} \frac{H(f_N(X^N|Z))}{N} \geq \lim_{N\to\infty} \max_{\mu_X \in \mathcal{S}} \frac{I(X^N, Y^N)}{N} \geq \max_{\mu_X \in \mathcal{S}} I(X,Y) \geq \bar{R}(\theta)$$

since we can easily observe that, for all N,

$$\begin{aligned} I(X^N, Y^N) &= h(X^N) - H(X^N|Y^N) \\ &\overset{(a)}{\geq} \sum_{i=1}^{N} h(X_i) - h(X_i|Y^N, X_1, \ldots, X_{i-1}) \\ &\overset{(b)}{\geq} \sum_{i=1}^{N} h(X_i) - h(X_i|Y_i) \\ &\geq \sum_{i=1}^{N} I(X_i, Y_i) = N I(X,Y) \end{aligned}$$

where (a) derives from the fact that X^N is a sequence of independent symbols and (b) derives from the properties of conditional entropy. This concludes the proof.

6.6.6 Proof of Theorem 6.4.1

To provide a complete and general proof, we require the introduction of appropriate topologies and function spaces on which the functions R^+ and R^- are evaluated. We recall that the spaces $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be Polish space.

Let $BC(\mathcal{Y})$ denote the vector space of bounded and continuous real valued function defined on $\mathcal{Y}$. Equipped with the sup norm topology, $(BC(\mathcal{Y}), \|\cdot\|_\infty)$ is a Banach space. Let $(BC(\mathcal{Y}))^*$ denote the topological dual of $BC(\mathcal{Y})$. Then, following [100, Theorem IV.6.2], $(BC(\mathcal{Y}))^*$ is isometrically isomorphic to the Banach space of finitely additive regular sign measures on $\mathcal{Y}$, i.e., $(BC(\mathcal{Y}))^* \simeq \mathcal{M}_{rba}(\mathcal{Y})$. Let $\Pi_{rba}(\mathcal{Y}) \subset \mathcal{M}_{rba}$ denote the set of regular bounded finitely additive probability measures on $\mathcal{Y}$. For $\mu \in \mathcal{M}_1(\mathcal{X})$, we indicate with $L_1(\mu, BC(\mathcal{Y}))$ the space of all μ-integrable functions defined on $\mathcal{X}$ an valued in $BC(\mathcal{Y})$. Equipping $L_1(\mu, BC(\mathcal{Y}))$ with the norm topology induced by the norm

$$\phi \in L_1(\mu, BC(\mathcal{Y})) \quad \|\phi\|_\mu = \int_{\mathcal{X}} \|\phi(x)(\cdot)\|_{BC(\mathcal{Y})} \mu(dx) < \infty$$

defines the Banach space $(L_1(\mu, BC(\mathcal{Y})), \|\cdot\|_\mu)$. Unfortunately, without any additional assumption, $(BC(\mathcal{Y}))^* \simeq \mathcal{M}_{rba}(\mathcal{Y})$ does not satisfy the Radon Nikodym property (RNP) [101, Theorems III.5 and IV.1], implying that the dual of $L_1(\mu, BC(\mathcal{Y}))$ is not $L_\infty(\mu, \mathcal{M}_{rba}(\mathcal{Y}))$[2]. However, it follows from the theory of "lifting" [102, Theorems VII.7 and VII.9] that the dual of $L_1(\mu, BC(\mathcal{Y}))$ can be identified with $L^w_\infty(\mu, \mathcal{M}_{rba}(\mathcal{Y}))$, the space of all functions defined on $\mathcal{X}$ and valued on $\mathcal{M}_{rba}(\mathcal{Y})$ which are weak* measurable, i.e., $Q \in L^w_\infty(\mu, \mathcal{M}_{rba}(\mathcal{Y}))$ if, for each $\phi \in BC(\mathcal{Y})$, $x \to \int_{\mathcal{Y}} \phi(y) Q(dy; x)$ is μ-measurable and μ-essentially bounded. Now, we define the following set

$$\mathcal{L}_{ad} \triangleq L^w_\infty(\mu, \Pi_{rba}(\mathcal{Y})) \subset L^w_\infty(\mu, \mathcal{M}_{rba}(\mathcal{Y}))$$

which represent the unit sphere in the space $L^w_\infty(\mu, \mathcal{M}_{rba}(\mathcal{Y}))$. Following Alaoglu Theorem [103, Theorem 2.6.18], $\mathcal{L}_{ad}$ is ensured to be weak* compact. For each $\phi \in L_1(\mu, BC(\mathcal{Y}))$, we can define the linear functional on $L^w_\infty(\mu, \mathcal{M}_{rba}(\mathcal{Y}))$

$$l_\phi(Q) \triangleq \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} \phi(x, y) Q(dy; x) \right) \mu(dx)$$

which is certainly bounded linear weak* continuous.

Under this function space formulation, we can consider a distortion function $d : \mathcal{X} \times$

[2]Under the additional assumption of $\mathcal{Y}$ being a compact space, $(BC(\mathcal{Y}))^* \simeq \mathcal{M}_{ca}(\mathcal{Y})$ is the space of countably additive sign measures, which satisfy the RNP.

$\mathcal{Y} \to \mathbb{R}_0^+$ as a measurable function from the class $L_1(\mu, BC(\mathcal{Y}))$ satisfying the constraint

$$\mathcal{L}_\mu(D) = \{Q \in \mathcal{L}_{ad} : l_d(Q) \leq D\} \tag{6.6.6}$$

As a subset of $\mathcal{L}_{ad}$ the set $\mathcal{L}_\mu(D)$ is convex, bounded and weak* closed, hence weak* compact (being a closed subset of a weak* compact set).

The assumption on $d \in L_1(\mu, BC(\mathcal{Y}))$, i.e., a bounded distortion metric, can be relaxed by assuming that, for each $x \in \mathcal{X}$, $d(x, \cdot)$ is continuous on $\mathcal{Y}$. It can be shown that, in this case, $\mathcal{L}_\mu(D)$ remains weak* closed, i.e., weak* compact [98, Lemma 2.4].

Given the general characterization of the properties of $\mathcal{L}_\mu(D)$ for any $\mu \in \mathcal{S}(\Gamma)$, the following lemma investigate the properties of the intersection of such sets.

Lemma 6.6.1. *Any family of intersections of $\mathcal{L}_\mu(D)$, for $\mu \in \mathcal{S}(\Gamma)$, is convex and weak* compact.*

Proof. Regarding the convexity of the intersection, see [103, Proposition 1.1.(4)]. Regarding the compactness, by [103, Proposition B.10] the arbitrary intersection of weak* closed sets is a weak* closed set. Since the intersection is a subset of $\mathcal{L}_{ad}$ - a weak* compact set - then also the intersection is weak* compact. □

We are now ready to prove the theorem. Let $\mathcal{A} = \mathcal{L}(D)$ and $\mathcal{B} = \mathcal{S}(\Gamma)$. By Lemma 6.6.1, $\mathcal{A}$ is a convex and weak* compact set, while $\mathcal{B}$ is a convex set - from the convexity of the KL divergence. We remark that $I(\mu, Q)$ is convex and lower semi continuous in $Q \in \mathcal{A}$, for every $\mu \in \mathcal{B}$, and concave in $\mu \in \mathcal{B}$, for every $Q \in \mathcal{A}$ [40]. Therefore, the conditions of the Minisup Theorem [97] are satisfied.

6.6.7 Proof of Theorem 6.4.2

For fixed (s, λ), $F_{s,\lambda}(\mu^*, q^*)$ can be expressed as

$$\begin{aligned} F_{s,\lambda}(\mu^*, q^*) &= \sup_{\mu \in \mathcal{M}_1(\mathcal{X})} F_{s,\lambda}(\mu, q^*) \\ &= -sD + \lambda\Gamma + \lambda \left(\sup_{\mu \in \mathcal{M}_1(\mathcal{X})} \int_{\mathcal{X}} \frac{g(x)}{\lambda} d\mu - D_{KL}(\mu \| \mu_0) \right) \end{aligned}$$

Since $s \geq 0$, we have that $g(\cdot) \geq 0$, implying that g is a bounded-below measurable function on $(\mathcal{X}, \Sigma_{\mathcal{X}})$. Leveraging the duality between relative entropy and free-energy [104, Proposition 2.3], the supremum is attained by μ^* characterized by (6.4.8), which, once substituted in (6.4.6), recovers (6.4.7). This concludes the proof.

6.6.8 Proof of Theorem 6.4.3

Following (6.4.1), we have $\mathcal{L}(D) \subset \cup_{\mu \in \mathcal{S}(\Gamma)} \mathcal{L}_\mu(D)$. Therefore, from classical results on the RDF of Bernoulli sources [55, Theorem 10.3.1], we have

$$R^-(D) \geq \max_{\beta \in \mathcal{B}} h_b(\beta) - h_b(D) \tag{6.6.7}$$

where $\mathcal{B} = \{\beta \in [0,1] : Bern(\beta) \in \mathcal{S}(\Gamma)\}$. We remark that the maximization problem over the source parameter β is convex. Therefore, introducing the Lagrangian multiplier $\lambda \geq 0$ and solving the Karush-Kuhn-Tucker conditions [56], we can characterize the sufficient optimality condition

$$\frac{1-\beta^*}{\beta^*} = \left(\frac{1-\beta_0}{\beta_0}\right)^{\frac{\lambda}{1+\lambda}}.$$

To show that (6.6.7) holds with equality, we just need to show that the conditional distribution Q_{β^*} achieving the RDF for $X \sim Bern(\beta^*)$, i.e,

$$\begin{aligned} Q_{\beta^*}(1|0) &= \frac{D}{1-\beta^*}\frac{\beta^*-D}{1-2D} \\ Q_{\beta^*}(0|1) &= \frac{D}{\beta^*}\frac{(1-\beta^*)-D}{1-2D} \end{aligned}$$

belongs to the set $\mathcal{L}(D)$, i.e., for all $\mu \in \mathcal{S}(\Gamma)$,

$$\mathbb{E}_{\mu Q_{\beta^*}}[d(X,Y)] = \beta Q_{\beta^*}(0|1) + (1-\beta)Q_{\beta^*}(1|0) \leq D. \tag{6.6.8}$$

Considering the limits $\lambda \to 0$ and $\lambda \to \infty$, we remark that $\beta^* \in [0, 0.5]$. Therefore, $Q_{\beta^*}(0|1) \geq Q_{\beta^*}(0|1)$.

Considering the case $\mathcal{B} \subset [0, 0.5)$, the monotonicity of $h_b(\cdot)$ implies that $\beta^* = \max \mathcal{B}$. Since $\mathbb{E}_{\mu Q_{\beta^*}}[d(X,Y)]$ is monotonically increasing in $[0, 0.5)$, (6.6.8) is satisfied. In all other cases, $\beta^* = 0.5$, implying that $Q_{\beta^*}(0|1) = Q_{\beta^*}(0|1) = D$, thus automatically satisfying (6.6.8). This concludes the proof.

Chapter 7

Conclusion and Future Work

The exponential growth of connected devices, multimedia streaming, and machine-to-machine communication has placed enormous pressure on the design of efficient data compression systems. Classical approaches, rooted in Shannon's rate-distortion theory, optimize for mathematical fidelity to the original source but do not account for the perceptual quality demanded by human end-users, nor for the semantic meaning that downstream tasks require. As communication systems increasingly mediate human experience, from remote sensing to generative media and autonomous systems, there is a pressing need for compression frameworks that align reconstruction quality with perception and semantic relevance, not merely with signal-level accuracy. In this thesis we addressed that need through a rigorous theoretical and algorithmic study of goal-oriented semantic communication, framed within an extended rate-distortion-perception theory. The contributions span discrete and continuous source models, novel algorithmic frameworks, and distributionally robust coding strategies.

In Part I, we addressed the computational challenge of evaluating the RDPF for discrete memoryless sources. The central contribution is the development of a family of iterative algorithmic schemes inspired by the classical Blahut-Arimoto algorithm for the rate-distortion function. The starting point is an Optimal Alternating Minimization (OAM) scheme, which provably converges to globally optimal solutions but cannot be directly implemented due to the presence of implicit equations defining the reconstruction distribution in its iteration structure. To overcome this limitation, we proposed two practical alternatives: a Newton-based Alternating Minimization scheme designed for smooth perception metrics, and a Relaxed Alternating Minimization scheme based on a tractable relaxation of the OAM iterates. The developed methods demonstrate how the additional distributional constraint substantially complicates the source coding problem, while also showing that alternating optimization methods can achieve global optimality solutions and can serve as the foundation for data-driven approaches, as shown in [105] where the OAM iterates provide a direct simplification of neural-based encoder architecture.

In Part II, we extended the algorithmic framework to continuous source models, progressing from finite-dimensional Gaussian vectors to infinite-dimensional Gaussian pro-

cesses and generic multivariate distributions. For Gaussian sources, we derived closed-form expressions for the RDPF under MSE distortion and a range of divergence-based perception metrics, developed a provably convergent alternating minimization algorithm for the vector case, and characterized a closed-form solution for the perfect realism regime. We then generalized these results to Gaussian processes, establishing that the optimal reconstruction shares the eigenvector structure of the source covariance operator, and derived a tractable analytical upper bound expressed in terms of the source's power spectral density. Finally, we introduced a copula-based estimation method for the perfect realism RDPF applicable to arbitrary multivariate continuous sources, establishing a unifying equivalence between the output-constrained rate-distortion function and the entropic optimal transport problem, and implemented a scalable stochastic gradient descent estimator based on approximate I-projections and validated on both scalar and vector sources.

In Part III, we relaxed the assumption of a perfectly known source distribution and investigated distributionally robust source coding with perceptual constraints. We developed two novel extensions of the Strong Functional Representation Lemma to settings where the source marginal belongs to an uncertainty set, covering both finite uncertainty sets and Kullback-Leibler balls around a nominal distribution. These extensions were used to derive robust coding schemes with operational guarantees in both the one-shot and asymptotic regimes. For KL-ball uncertainty sets, we then derived an implicit characterization of the set of distributions attaining the robust rate-distortion function and designed a novel iterative algorithm for its computation, with convergence properties analyzed via fixed-point theory. Numerical examples illustrated how robustness requirements translate into a quantifiable rate overhead relative to the nominal rate-distortion function.

Together, the three parts of this thesis form a continuous progression. Part I established the algorithmic foundations for computing perception-constrained rate-distortion tradeoffs in the discrete setting; Part II extended the algorithmic framework considering the richer mathematical structure of continuous and infinite-dimensional sources; and Part III confronted the practical reality that source statistics are rarely known with precision, extending the theory to distributionally robust settings. A unifying theme across all three parts is the role of divergence measures — as perception constraints, as uncertainty radii, and as the bridge between rate-distortion theory and optimal transport. The results collectively show that perception and robustness, far from being engineering afterthoughts, can be embedded directly into the information-theoretic foundations of source coding, with tractable algorithms that scale to real-world problems.

Future Work Several directions for future research emerged naturally from this work. First, while the copula-based estimator developed in Chapter 5 is applicable to generic multivariate sources, its convergence guarantees remain asymptotic and its sample complexity in high dimensions has not been fully characterized. A tighter analysis of the estimator's finite-sample behavior, and the development of variance-reduction techniques for the stochastic gradient updates, would substantially broaden its practical applica-

bility in both source coding and optimal transport. Second, the distributionally robust coding schemes derived in Part III focused on uncertainty sets defined by KL divergence balls around a nominal distribution. Extending this framework to other divergence families (such as Wasserstein balls or f-divergence neighborhoods) would yield robustness guarantees tailored to different notions of distributional shift. Such extensions are particularly relevant for neural image compression under domain mismatch, where the test distribution may differ significantly from the training distribution. Third, this thesis, like the broader RDP literature, treats perception constraint levels as fixed design parameters. A promising direction is to learn or adapt these constraints from user feedback or task-specific performance metrics. This would enable fully end-to-end goal-oriented communication systems in which the perception budget is co-optimized alongside the encoder and decoder. Such an approach could leverage techniques from continual learning and networked control systems to dynamically adjust perception constraints based on evolving requirements.

Bibliography

[1] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, "Generative Adversarial Networks for Extreme Learned Image Compression," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 221–231.

[2] "2025 Global Internet Phenomena Report," AppLogic Networks (formerly Sandvine), Technical Report, Mar. 2025, Global Analysis of Application-Level Internet Traffic Trends. [Online]. Available: https://www.applogicnetworks.com/blog/the-2025-global-internet-phenomena-report

[3] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[4] C. Zhang, H. Zou, S. Lasaulce, W. Saad, M. Kountouris, and M. Bennis, "Goal-Oriented Communications for the IoT and Application to Data Compression," *IEEE Internet of Things Magazine*, vol. 5, no. 4, pp. 58–63, 2022.

[5] P. Agheli, N. Pappas, and M. Kountouris, "Semantic Filtering and Source Coding in Distributed Wireless Monitoring Systems," *IEEE Transactions on Communications*, vol. 72, no. 6, pp. 3290–3304, 2024.

[6] Y.-C. Chao, Y. Chen, W. Wang, A. Wijesinghe, S. Wanninayaka, S. Zhang, and Z. Ding, "Task-Driven Semantic Quantization and Imitation Learning for Goal-Oriented Communications," 2025.

[7] A. K. Moorthy and A. C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Trans. Image Proc.*, vol. 20, no. 12, pp. 3350–3364, 2011.

[8] F. Mentzer, G. D. Toderici, M. Tschannen, and E. Agustsson, "High-Fidelity Generative Image Compression," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11 913–11 924, 2020.

[9] Y. Blau and T. Michaeli, "Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff," in *International Conference on Machine Learning*. PMLR, 2019, pp. 675–685.

[10] 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.

[11] ——, "Rate-Distortion-Perception Tradeoff of Variable-Length Source Coding for General Information Sources," *IEICE Comm. Express*, vol. 8, no. 2, pp. 38–42, 2019.

[12] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 105–114.

[13] T. R. Shaham and T. Michaeli, "Deformation Aware Image Compression," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[14] S. Kudo, S. Orihashi, R. Tanida, and A. Shimizu, "GAN-based Image Compression Using Mutual Information Maximizing Regularization," in *Picture Coding Symposium*, 2019, pp. 1–5.

[15] D. Minnen, J. Ballé, and G. D. Toderici, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[16] 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.

[17] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain," *IEEE Trans. Image Proc.*, vol. 21, no. 8, pp. 3339–3352, 2012.

[18] M. Kountouris and N. Pappas, "Semantics-Empowered Communication for Networked Intelligent Systems," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 96–102, 2021.

[19] S. Katakol, B. Elbarashy, L. Herranz, J. van de Weijer, and A. M. López, "Distributed Learning and Inference with Compressed Images," *IEEE Trans. Image Proc.*, vol. 30, pp. 3069–3083, 2021.

[20] 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.

[21] ——, "Alternating Minimization Schemes for Computing Rate-Distortion-Perception Functions With f-Divergence Perception Constraints," *IEEE Transactions on Information Theory*, vol. 71, no. 11, pp. 9100–9115, 2025.

[22] ——, "Computation of the Multivariate Gaussian Rate-Distortion-Perception Function," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024.

[23] ——, "On the Computation of the Gaussian Rate-Distortion-Perception Function," *IEEE Journal on Selected Areas in Information Theory*, vol. 5, pp. 314–330, 2024.

[24] ——, "On the Rate-Distortion-Perception Function for Gaussian Processes," in *2025 IEEE International Symposium on Information Theory (ISIT)*, 2025.

[25] ——, "Copula-Based Estimation of Continuous Sources for a Class of Constrained Rate-Distortion Functions," in *Proc. IEEE Int. Symp. Inf. Theory*, 2024, pp. 1089–1094.

[26] ——, "On Distributionally Robust Lossy Source Coding," in *2025 IEEE Information Theory Workshop (ITW)*, 2025.

[27] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*. Springer Verlag, 1998.

[28] R. Bhatia, *Matrix Analysis*. Springer Science & Business Media, 2013, vol. 169.

[29] R. Horn and C. Johnson, *Matrix Analysis*, ser. Matrix Analysis. Cambridge University Press, 2013.

[30] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, "On the Lambert W Function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 01 1996.

[31] L. Theis and A. B. Wagner, "A Coding Theorem for the Rate-Distortion-Perception Function," in *International Conference of Learning Representations (ICLR)*, 2021, pp. 1–5.

[32] C. T. Li and A. E. Gamal, "Strong Functional Representation Lemma and Applications to Coding Theorems," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

[33] N. Saldi, T. Linder, and S. Yüksel, "Output Constrained Lossy Source Coding With Limited Common Randomness," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4984–4998, 2015.

[34] 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.

[35] 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.

[36] J. Qian, S. Salehkalaibar, J. Chen, A. Khisti, W. Yu, W. Shi, Y. Ge, and W. Tong, "Rate-Distortion-Perception Tradeoff for Gaussian Vector Sources," *IEEE Journal on Selected Areas in Information Theory*, vol. 6, pp. 1–17, 2025.

[37] O. Kirmemis and A. M. Tekalp, "A Practical Approach for Rate-Distortion-Perception Analysis in Learned Image Compression," in *2021 Picture Coding Symposium (PCS)*. IEEE, 2021, pp. 1–5.

[38] S. Salehkalaibar, T. B. Phan, J. Chen, W. Yu, and A. Khisti, "On the Choice of Perception Loss Function for Learned Video Compression," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[39] 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," in *Proc. IEEE Int. Symp. Inf. Theory*, 2023, pp. 1074–1079.

[40] I. Csiszár, "On an extremum problem of information theory," *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, pp. 57–61, 1974.

[41] A. Rényi, "On Measures of Entropy and Information," *Proc. of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, vol. 4, pp. 547–461, 1961.

[42] I. Csiszár and P. C. Shields, "Information Theory and Statistics: A Tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 4, pp. 417–528, 2004.

[43] I. Sason, "On f-Divergences: Integral Representations, Local Behavior, and Inequalities," *Entropy*, vol. 20, no. 5, 2018.

[44] F. Nielsen, "On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means," *Entropy*, vol. 21, no. 5, 2019.

[45] C. Villani, *Topics in Optimal Transportation*. American Mathematical Soc., 2021, vol. 58.

[46] M. Gelbrich, "On a Formula for the L2 Wasserstein Metric between Measures on Euclidean and Hilbert Spaces," *Mathematische Nachrichten*, vol. 147, no. 1, pp. 185–203, 1990.

[47] R. E. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[48] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical Analysis*. Cengage learning, 2015.

[49] I. Csiszar and G. Túsnady, "Information Geometry and Alternating Minimization Procedures," *Statistics and Decisions, Dedewicz*, vol. 1, pp. 205–237, 1984.

[50] L. Grippo and M. Sciandrone, "On the Convergence of the Block Nonlinear Gauss–Seidel Method Under Convex Constraints," *Operations Research Letters*, vol. 26, no. 3, pp. 127–136, 2000.

[51] S. Arimoto, "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, 1972.

[52] R. Yeung, *Information Theory and Network Coding*. Springer, 2008.

[53] R. E. Blahut, *Principles and Practice of Information Theory*. Addison-Wesley, 1987.

[54] K. Nakagawa, Y. Takei, S.-i. Hara, and K. Watabe, "Analysis of the Convergence Speed of the Arimoto-Blahut Algorithm by the Second-Order Recurrence Formula," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6810–6831, 2021.

[55] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., NJ, 2006.

[56] R. T. Rockafellar, *Convex Analysis*. Princeton university press, 1970, vol. 18.

[57] J. Qian, "On the Rate-Distortion-Perception Tradeoff for Lossy Compression," Ph.D. dissertation, McMaster University, October 2023, http://hdl.handle.net/11375/28976.

[58] S. Fang and Q. Zhu, "Independent Gaussian Distributions Minimize the Kullback-Leibler (KL) Divergence from Independent Gaussian Distributions," *arXiv preprint arXiv:2011.02560*, 2020.

[59] P. A. Stavrou and M. Skoglund, "Asymptotic Reverse Waterfilling Algorithm of NRDF for Certain Classes of Vector Gauss–Markov Processes," *IEEE Trans. Autom. Control*, vol. 67, no. 6, pp. 3196–3203, 2022.

[60] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995.

[61] A. B. Wagner, "The Rate-Distortion-Perception Tradeoff: The Role of Common Randomness," 2022. [Online]. Available: https://arxiv.org/abs/2202.04147

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

[63] A. Beck and L. Tetruashvili, "On the Convergence of Block Coordinate Descent Type Methods," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2037–2060, 2013.

[64] J. Bernal and J. Lawrence, "Characterization and Computation of Matrices of Maximal Trace Over Rotations," *Journal of Geometry and Symmetry in Physics*, vol. 53, no. none, pp. 21 – 53, 2019.

[65] R. G. Gallager, *Information Theory and Reliable Communication*. Springer, 1968, vol. 588.

[66] D. Sakrison, "The Rate Distortion Function of a Gaussian Process with a Weighted Square Error Criterion (Corresp.)," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 506–508, 1968.

[67] I. Gohberg and S. Goldberg, *Basic operator theory*. Birkhäuser, 2013.

[68] I. Steinwart and C. Scovel, "Mercer's Theorem on General Domains: On the Interaction between Measures, Kernels, and RKHSs," *Constructive Approximation*, vol. 35, no. 3, pp. 363–417, Jun 2012.

[69] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

[70] A. Mallasto and A. Feragen, "Learning from Uncertain Curves: The 2-Wasserstein Metric for Gaussian Processes," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[71] J. A. Cuesta-Albertos, C. Matrán-Bea, and A. Tuero-Diaz, "On Lower Bounds for the L2-Wasserstein Metric in a Hilbert space," *Journal of Theoretical Probability*, vol. 9, no. 2, pp. 263–283, Apr 1996.

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

[73] M. Li, J. Klejsa, and W. B. Kleijn, "On Distribution Preserving Quantization," 2011. [Online]. Available: https://arxiv.org/abs/1108.3728

[74] 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.

[75] 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.

[76] 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.

[77] 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.

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

[79] J. Ma and Z. Sun, "Mutual Information is Copula Entropy," *Tsinghua Science and Technology*, vol. 16, no. 1, pp. 51–54, 2011.

[80] 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.

[81] C. P. Robert, G. Casella, C. P. Robert, and G. Casella, "Monte Carlo Integration," *Monte Carlo Statistical Methods*, pp. 71–138, 1999.

[82] G. Garrigos and R. M. Gower, "Handbook of Convergence Theorems for (Stochastic) Gradient Methods," 2023. [Online]. Available: https://arxiv.org/abs/2301.11235v2

[83] A. Klenke, *Probability Theory: A Comprehensive Course*. Springer Science & Business Media, 2013.

[84] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*. American Mathematical Society (RI), 1950, vol. 1.

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

[86] S. Bandyopadhyay, A. Datta, A. Pal, and S. R. R. Gadepally, "Intelligent Continuous Monitoring to Handle Data Distributional Changes for IoT Systems," in *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '22. New York, NY, USA: Association for Computing Machinery, 2023, p. 1189–1195.

[87] T. Chen and Z. Ma, "Toward Robust Neural Image Compression: Adversarial Attack and Model Finetuning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 12, pp. 7842–7856, 2023.

[88] J. Ziv and A. Lempel, "A Universal Algorithm for Sequential Data Compression," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 337–343, 1977.

[89] ——, "Compression of Individual Sequences via Variable-Rate Coding," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp. 530–536, 1978.

[90] E. Yang and J. Kieffer, "Simple Universal Lossy Data Compression Schemes Derived from the Lempel-Ziv Algorithm," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 239–245, 1996.

[91] E. Yang, Z. Zhang, and T. Berger, "Fixed-slope Universal Lossy Data Compression," *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1465–1476, 1997.

[92] A. R. Washburn *et al.*, "Two-Person Zero-Sum Games," 2014.

[93] D. J. Sakrison, "The Rate Distortion Function for a Class of Sources," *Information and Control*, vol. 15, no. 2, pp. 165–195, 1969.

[94] D. Sakrison, "The Rate of a Class of Random Processes," *IEEE Transactions on Information Theory*, vol. 16, no. 1, pp. 10–16, 1970.

[95] H. Poor, "The Rate-Distortion Function on Classes of Sources Determined by Spectral Capacities," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 19–26, 1982.

[96] V. Malik, T. Kargin, V. Kostina, and B. Hassibi, "A Distributionally Robust Approach to Shannon Limits using the Wasserstein Distance," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 861–866.

[97] M. Sion, "On General Minimax Theorems." *Pacific Journal of Mathematics*, vol. 8, no. 1, pp. 171 – 176, 1958.

[98] F. Rezaei, N. Ahmed, and C. D. Charalambous, "Rate Distortion Theory for General Sources with Potential Application to Image Compression," *International Journal of Applied Mathematical Sciences*, vol. 3, no. 2, pp. 141–165, 2006.

[99] L. B. Wolfe and C.-I. Chang, "A Simple Method for Calculating the Rate Distortion Function of a Source with an Unknown Parameter," *Signal processing*, vol. 33, no. 2, pp. 209–221, 1993.

[100] N. Dunford and J. T. Schwartz, *Linear Operators, Part 1: General Theory.* John Wiley & Sons, 1988.

[101] J. Diestel and J. Uhl, J., "Vector Measures," 1977.

[102] A. I. Tulcea and C. I. Tulcea, *Topics in the Theory of Lifting.* Springer Science & Business Media, 2012, vol. 48.

[103] R. E. Megginson, *An Introduction to Banach Space Theory.* Springer Science & Business Media, 2012, vol. 183.

[104] P. Dai Pra, L. Meneghini, and W. Runggaldier, "Connections Between Stochastic Control and Dynamic Games," *Mathematics of Control, Signals and Systems*, vol. 9, pp. 303–326, 01 1996.

[105] E. Lei, H. Hassani, and S. S. Bidokhti, "Optimal Neural Compressors for the Rate-Distortion-Perception Tradeoff," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.