On the Rate-Distortion-Perception Function for Gaussian Processes

Giuseppe Serra†, Photios A. Stavrou†, Marios Kountouris†‡
†Communication Systems Department, EURECOM, Sophia-Antipolis, France
‡DaSCI, Department of Computer Science and AI, University of Granada, Spain
Email: {giuseppe.serra, fotios.stavrou, marios.kountouris}@eurecom.fr

***Abstract*—In this paper, we investigate the rate-distortion-perception function (RDPF) of a source modeled as a Gaussian Process (GP) over a measure space Ω, under mean squared error (MSE) distortion and squared Wasserstein-2 perception metrics. 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, akin to the classical rate-distortion function (RDF), allows us to reformulate the RDPF problem in terms of the Karhunen–Loève (KL) transform coefficients of the involved GPs. Leveraging the similarities with the finite-dimensional Gaussian RDPF, we derive a tight analytical upper bound on the RDPF for GPs, which recovers the optimal solution in the *"perfect realism"* regime. 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 and $T \to \infty$ as a function of the spectral density of the source process. We complement our theoretical findings with relevant simulation studies.**

I. Introduction

The rate-distortion-perception (RDP) trade-off, formulated simultaneously by Blau and Michaeli in [1] and Matsumoto in [2], [3], proposes a generalization of the classical rate-distortion (RD) theory [4] introducing the concept of perceptual quality, that is, the property of a sample to appear pleasing from a human perspective. This is enacted by extending the classical single-letter RD formulation, incorporating a divergence constraint between the source distribution and its estimation at the destination. The divergence constraint acts as a proxy for human perception, quantifying the satisfaction experienced when using data, as shown by its correlation with the human opinion scores in [5], [6]. Moreover, this divergence constraint may have multiple interpretations, such as a semantic quality metric that measures the relevance of the reconstructed source from the observer's perspective [7].

Multiple coding theorems have been developed for the RDP framework. Under the assumption of infinite common randomness between the encoder and the decoder, Theis and Wagner in [8] prove a coding theorem for stochastic variable-length codes in both the one-shot and asymptotic regimes. Originally in the context of the output-constrained RDF, but also valid for the "perfect realism" RDP case, Saldi *et. al.* [9] provide coding theorems for when only finite common randomness between encoder and decoder is available.

Similarly to the classical RD theory, the mathematical embodiment of the RDP framework is represented by the rate-distortion-perception function (RDPF), which, as its classical counterpart, does not enjoy a general analytical solution. The absence of a general closed-form solution has prompted research into computational methods for RDPF estimation. Data-driven solutions have garnered significant attention in the field, predominantly involving architectures based on generative adversarial schemes that minimize a linear combination of distortion and perception metrics, see, e.g., [1], [10], [11]. However, while these approaches can directly optimize image or video codecs using only source samples, they often require substantial computational and data resources and may exhibit limited generalization capabilities. On the other hand, algorithmic results for the RDPF computation are proposed by both Serra *et. al.* in [12] and Chen *et. al.* in [13] considering the case of discrete sources, while [14] proposes a general algorithm for continuous sources for the *"perfect realism"* regime, i.e., when the reconstructed process and source process are constrained to have the same statistics. However, despite the general complexity, certain analytical expressions have been developed for specific categories of sources. For instance, binary sources subject to Hamming distortion and total variation distance have closed-form expressions, as discussed in [1]. Similarly, [15] provides closed-form expressions for the case of scalar Gaussian sources under mean squared error (MSE) distortion and various perceptual metrics.
Most theoretical results in the literature primarily focus on scalar or finite-dimensional sources, using the derived insights to guide specific design choices for data-driven architectures. Although this approach may be suitable for image data, sources like audio signals may require special attention because of their inherent suitability for modeling as stochastic processes.

A. Our Approach and Contributions

The aim of this work is to characterize the RDPF for a source modeled as a Gaussian process (GP-RDPF), defined over a measure space Ω under 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 [16], [17]. Noticing the similarities with the finite-dimensional Gaussian RDPF studied in our earlier work [15], 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.

II. Preliminaries

A. Rate-Distortion-Perception Tradeoff

We begin by providing the definition and some properties of the RDPF for general alphabets.

Definition 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 $\Delta:\mathcal{X}^2\to\mathbb{R}_0^+$ and divergence function $d:\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{X})\to\mathbb{R}_0^+$ is defined as follows:

$$\begin{aligned} R(D,P)\triangleq\ &\min_{P_{Y|X}} I(X,Y)\\ \text{s.t.}\quad &\mathbb{E}\left[\Delta(X,Y)\right]\le D\\ &d(P_X||P_Y)\le P \end{aligned}\tag{1}$$

where the minimization is among all conditional distributions $P_{Y|X}:\mathcal{X}\to\mathcal{P}(\mathcal{X})$.
We point out the following remark on Definition 1.

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

B. 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 $m(t)=\mathbb{E}\left[X(t)\right]$ and $k(t,s)=\mathbb{E}\left[(X(t)-m(t))(X(s)-m(s))\right]$, 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., [16, Chapter 8].

C. 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 $\mathcal{H}$ [18]. 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 (iff) $\mathrm{Tr}[(T^*T)^{\frac{1}{2}}]<\infty$, 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 paper, we assume that the underlying measure space $\mathcal{X}$ is such that $(L^2(\Omega),\langle\cdot,\cdot\rangle)$ is separable.

D. 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 [19]. 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

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

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 $\mathbb{E}\left[X_i\right]=0$ and $\mathbb{E}\left[X_iX_j\right]=\lambda_i\delta_{i,j}$. 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

[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.

represented by a countable set $\{X_i\}_{i=1}^{\infty}$ of independent Gaussians. This representation is commonly referred to as the KL transform [20] and we refer to $\{X_i\}_{i=1}^{\infty}$ as the KL coefficients of the process X.

E. Squared Wasserstein-2 distance for GP

Squared Wasserstein-2 distance was originally introduced in [21] as a specific instance of the optimal transport problem (see, e.g., [22, Chapter 7]). In particular, squared Wasserstein-2 distance is defined as follows

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

where $\Pi(P_X, P_Y)$ is the set of all joint distributions $P_{X,Y}$ with marginals P_X and P_Y. Following [23, 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

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

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.

III. Main Results

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

Theorem 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

$$\begin{aligned} R_{GP}(D,P) = &\min_{\{P_{Y_i|X_i}\}_{i=1}^{\infty}} \sum_{i=1}^{\infty} I(X_i;Y_i) \\ &\text{s.t.} \quad \sum_{i=1}^{\infty} \mathbb{E}\left[||X_i - Y_i||^2\right] \leq D \\ &\qquad\ \sum_{i=1}^{\infty} \mathrm{W}_2^2(X_i, Y_i) \leq P \end{aligned} \tag{3}$$

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: 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

$$[K_Y \psi] = \sum_{i=1}^{\infty} \nu_i \langle \psi, \eta_i \rangle \eta_i \qquad Y = \sum_{i=1}^{\infty} Y_i \eta_i, \tag{4}$$

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 $I\left(\{X_i\}_{i=1}^{\infty}; \{Y_i\}_{i=1}^{\infty}\right)$. 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, [24, 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 [17]. Consequently, the MSE of the two processes can be expressed as

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

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. ■

Theorem 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 [15]. The following corollary further simplifies (3) leveraging knowledge of the analytic solution of the scalar Gaussian RDPF under MSE distortion and $\mathrm{W}_2^2(\cdot,\cdot)$ perception.

Corollary 1: The optimization problem in (3) can be cast as follows

$$\begin{aligned} R_{GP}(D,P) = &\min_{\{(D_i,P_i)\}_{i=1}^{\infty}} \sum_{i=1}^{\infty} R_{X_i}(D_i, P_i) \\ &\text{s.t.} \quad \sum_{i=1}^{\infty} D_i \leq D \quad \sum_{i=1}^{\infty} P_i \leq P \end{aligned} \tag{5}$$

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: 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 (3) 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. ■

Serra *et. al.* [15] analyze the finite-dimensional version of (5), designing a solving algorithm based on alternating minimization for the computation of the optimal allocations $\{(D_i, P_i)\}_{i=1}^{\infty}$. Their solution leverages 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 (5) 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 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 [15, 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 [15] evaluated at $(\tilde{D}_i, P_i)$. This concludes the proof. ■

We stress the following technical remark for Theorem 2.

Remark 2: Fixing the perceptual levels $\{P_i\}_{i=1}^{\infty}$, the parametric solutions recovered in Theorem 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 [17]. 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 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 [15, 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.,

$$\tilde{P}_i = \alpha\lambda_i \quad \Longrightarrow \quad \alpha\sum_{i=1}^{\infty}\lambda_i = P \tag{6}$$

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.

A. 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 [16], [17], 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 2 to the case where $T \to \infty$, as shown in the following theorem.

Theorem 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

$$\tilde{R} = \frac{1}{2}\int_{\mathcal{S}_{RDP}} \log\left(\frac{2S_X^2(f)(1-\sqrt{\alpha})^2}{\gamma(\sqrt{\gamma^2 + 4S_X^2(f)(1-\sqrt{\alpha})^2} - \gamma)}\right) df + \frac{1}{2}\int_{\mathcal{S}_{RD}} \log^+\left(\frac{S_X(f)}{\gamma}\right) df \tag{7}$$

$$\tilde{D} = \int_{\mathcal{S}_{RDP}} \hat{D}(S_X(f))df + \int_{\mathcal{S}_{RD}} \min\{\gamma, S_X(f)\}df \qquad (8)$$

$$\hat{D}(x) = \gamma + x(1 + (1-\sqrt{\alpha})^2) - \sqrt{4x^2(1-\sqrt{\alpha})^2 + \gamma^2}$$

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 (6), the proof follows as a direct application of [16, Lemma 8.5.3] to the results of Theorem 2. ■

IV. Numerical Example

In this section, we provide a numerical example to better illustrate the results of Theorem 3. Let the source X be modeled as a stationary GP with power spectral density S_X as reported in Fig. 1.

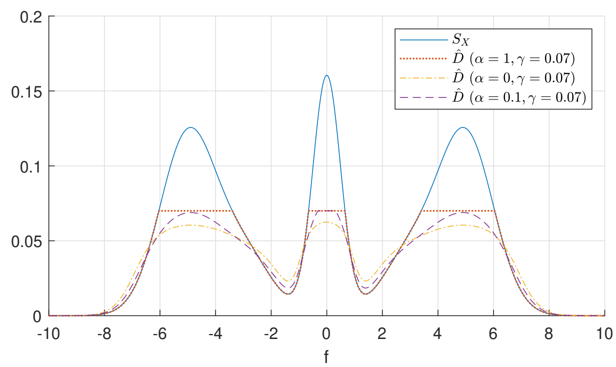

Fig. 1. Source Power Spectrum $S_X(f)$ vs. per-frequency distortion $\hat{D}(S_X(f))$ for $\gamma = 0.7$ and varying α.

We first investigate the per-frequency distortion $\hat{D}(S_X(f))$, defined as the integral argument of (8), 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 [16]. 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 2.

Fig. 2 shows the numerically estimated curves of the rate $\tilde{R}$ (Fig. 2a) and distortion $\tilde{D}$ (Fig. 2b) defined in (7) and (8), 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.

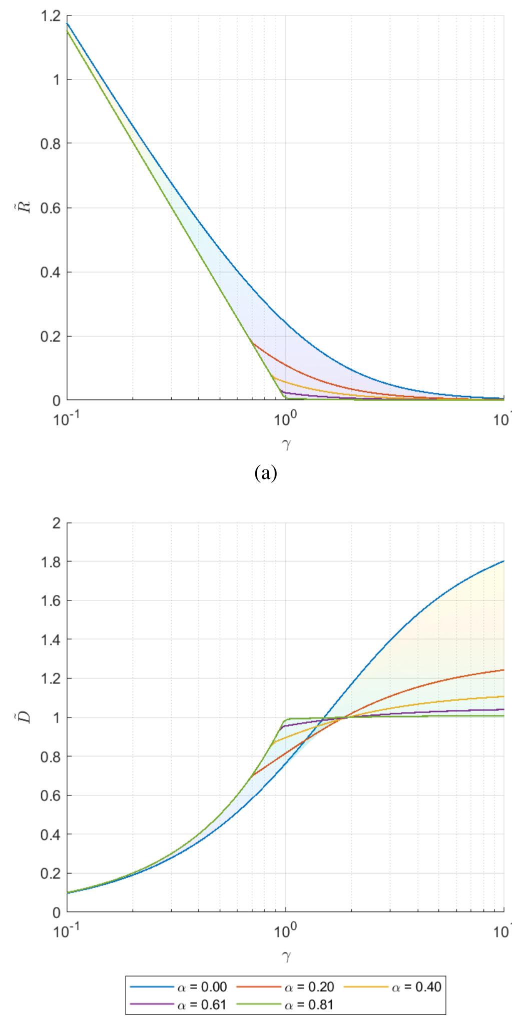

Fig. 2. (a) Rate $\tilde{R}$ and (b) distortion $\tilde{D}$ curves parameterized by (α, γ).

V. Conclusion

In this paper, we characterized the RDPF of a source modeled as a GP, under MSE distortion and squared Wasserstein-2 perceptual metrics. We first provided a general characterization for non-stationary sources using their Karhunen–Loève representation, leveraging the structure of both the distortion and perceptual constraints. We then derived an analytical upper bound that exactly characterizes the RDPF in the “perfect realism" regime. Finally, we extended this closed-form result to the case of stationary GPs on the real line, expressing the bound as a function of the source's spectral power density. As future work, we aim to use our findings to guide the design of neural compression schemes tailored to audio sources and to validate our theoretical results through empirical evaluation in more realistic scenarios.

Acknowledgment The work of G. Serra and M. Kountouris is supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement No. 101003431). The work of P. A. Stavrou is supported in part by the SNS JU project 6G-GOALS [25] under the EU's Horizon programme (Grant Agreement No. 101139232) and by the Huawei France-EURECOM Chair on Future Wireless Networks.

REFERENCES

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

[2] R. Matsumoto, “Introducing the perception-distortion tradeoff into the rate-distortion theory of general information sources,” *IEICE Comm. Express*, vol. 7, no. 11, pp. 427–431, 2018.

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

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

[5] 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.

[6] M. A. Saad, A. C. Bovik, and C. Charrier, “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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 *IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 1074–1079.

[14] G. Serra, P. A. Stavrou, and M. Kountouris, “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.

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

[16] R. G. Gallager, *Information theory and reliable communication*. Springer, 1968, vol. 588.

[17] 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.

[18] I. Gohberg and S. Goldberg, *Basic operator theory*. Birkhäuser, 2013.

[19] 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.

[20] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

[21] 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.

[22] C. Villani, *Topics in Optimal Transportation*. American Mathematical Soc., 2021, vol. 58.

[23] 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.

[24] 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.

[25] E. C. Strinati et al., “Goal-oriented and semantic communication in 6G AI-native networks: the 6G-GOALS approach,” in *European Conference on Networks and Communications & 6G Summit*, June 2024, pp. 1–6.