The Role of Fidelity in Goal-Oriented Semantic Communication: A Rate Distortion Approach

Photios A. Stavrou, *Senior Member, IEEE* and Marios Kountouris, *Senior Member, IEEE*

Abstract

We study a variant of a robust description source coding framework via its corresponding characterization, which is a relevant model for goal-oriented semantic information transmission. Considering two individual single-letter separable distortion constraints and input and output data acting as the intrinsic and extrinsic message, respectively, we first derive bounds on the optimal rates of the problem, as well as necessary and sufficient conditions for these bounds to be tight. Subsequently, we prove a general result that provides in parametric form the optimal solution of the characterization of this problem. Capitalizing on these results, we examine the structure of the solution for one case study of general binary alphabets under Hamming distortions and solve in closed form a special case. We also solve another general binary alphabet case where a Hamming and an erasure distortion coexist, as a means to highlight the importance of selecting the type of the distortion constraint in goal-oriented semantic communication. We also develop a semantic-aware Blahut-Arimoto (BA) algorithm, which can be used for the computation of any finite alphabet intrinsic or extrinsic message under individual distortion criteria. Finally, we revisit the problem for multidimensional independent and identically distributed (IID) jointly Gaussian processes with individual mean-square error (MSE) distortion constraints, providing new insights that have previously been overlooked. This work reveals the cardinal role of context-dependent fidelity criteria in goal-oriented semantic communication.

I. Introduction

Shannon, in his seminal work [2], has deliberately considered the semantic aspects and the effectiveness of transmitted messages as irrelevant to the communication problem. Setting aside

Part of the material of this paper was accepted for publication in IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, 2022 [1].

The authors are with the Communication Systems Department at EURECOM, Sophia-Antipolis, France, email: {fotios.stavrou, marios.kountouris}@eurecom.fr

an issue which is otherwise confusing, this dichotomy between information content and its significance has been instrumental in achieving reliability and efficiency in information transmission over noisy channels. Nevertheless, in [3], Shannon has indirectly provided a means to study semantic information sources because the coding aspect determined by the probabilistic model of the source is dictated by a distortion constraint imposed in the system. Various endeavors have been made to incorporate semantics into Shannon's communication theory. Letting aside epistemic and doxastic logic theories, the most important efforts include probabilistic logic approaches [4]–[7], complexity theory approaches [8], [9], and semantic coding and communication games [10], [11]. The effectiveness problem has been considered using the concepts of pragmatic information [12] and value of information [13], [14]. Nevertheless, Shannon's communication model has remained virtually unchallenged. None of the proposed extensions has ever been recognized as a general theory of semantic or pragmatic information. The aforementioned theories have remained at a conceptual level, failing to have any tangible practical applications to or impact on communication networks. The quest for a goal-oriented semantic communication theory has recently gained new impetus [15]–[18], fueled by the emergence of networks of autonomous agents with advanced sensing, learning, and decision-making capabilities.

In this work, we consider the problem of communicating a memoryless source, whose semantic, remote or intrinsic information is not directly observable, based on noisy observations. Our objective is to investigate the impact of context-dependent fidelity criteria and distortion measures on goal-oriented information transmission and semantic source reconstruction. For that, we revisit a lossy compression framework, recently introduced in [19], [20], considering both finite alphabets and Gaussian $\mathrm{i.i.d}$ random variables (RVs), and we study the effect of multiple individual distortion criteria in goal-oriented semantic information transmission. The objective of this work is twofold. First, we aim at complementing and extending the work in [20], which only considers continuous alphabet sources (i.e., $\mathrm{i.i.d}$ Gaussian sources) and mean-square error (MSE) distortion criteria, providing results and new insights that have been overlooked in prior work. Second, we aim at further emphasizing on the role of the context-dependent fidelity criteria in goal-oriented semantic communication by showing cases with new outcomes that do not appear through the analysis of [19], [20].

A. Related work

The problem considered here falls within the realm of remote source coding problem [21]–[23] and rate distortion with multiple distortion constraints [24], [25], [26, Problem 10.19]. The most closely related setup to our work is the rate distortion framework introduced in [19], [20], which provides characterizations and optimal closed form expressions for $\mathrm{i.i.d}$ scalar-valued Gaussian processes and sub-optimal characterizations assuming linear based state-observation models with Gaussian observations with ways to numerically compute the solution beyond scalar RVs. Some additional results in [19], include a case study for the classification of a binary uniform semantic source with an extrinsic observable scalar-valued Gaussian mixture model. It should be noted that the rate distortion framework considered here and in [19], [20] can be seen as a generalization of the robust description problem for two individual distortion criteria, which in turn is a special case of the two description coding problem [24]. Rate distortion with two individual distortion criteria has been studied in many papers under various contexts, see, e.g., [27]–[30]. Another relevant yet different setup is the recently introduced rate-distortion-perception representations, see, e.g., [31], [32] (and the references therein), in which perception quality, measured by some divergence between distributions, is included in addition to the classical distortion criterion. One major difference between rate-distortion-perception problems and the setup here is that in the former the characterizations are solved for various examples using separately each distortion constraint, whereas in the latter, one can study from an optimization standpoint the joint behavior of the two distortion penalties.

B. Contributions

In this paper, we consider a variation of the robust lossy source coding model, similarly to [19], [20], which captures goal-oriented semantic attributes and intrinsic representation of information (e.g., features, structural/qualitative properties, embedding). For this setup and its corresponding characterization (see Lemma 1, eq. (6)), we derive the following new results.

- We obtain bounds on the semantic rate distortion function (cf. (6)) and identify necessary and sufficient conditions based on which these bounds can be tight (see Lemma 2).
- We prove a general theorem, which gives parametrically the implicit solution of (6) for arbitrary finite alphabet sets with individual semantic and observable distortion criteria (see Theorem 1).

- We develop a *semantic-aware Blahut-Arimoto algorithm* (see Algorithm 1) that allows the computation of (6) for any finite alphabet set of intrinsic or extrinsic messages with arbitrary individual single-letter distortion criteria.
- We revisit the problem for multidimensional jointly Gaussian RVs with individual MSE distortion constraints, initially studied in [20, Section IV], providing a different angle that is also aligned to the proposed semantic information transmission setup and our findings. For this class of input data we also derive an example that demonstrates the non-triviality of our lower bound in Lemma 2 (see Example 2).

The aforementioned results are not the sole contributions of this paper. Lemma 2 and Theorem 1 are applied into two examples (see Problems 1, 2) using specific setups with general binary alphabets and two types of distortion measures, namely Hamming and erasure distortions. For Problem 1, we derive structural properties on the optimal minimizer (test channel) consistent with Lemma 2 and characterize its solution (see Theorem 2). We enhance this result by solving in closed form a special case to illustrate the rate distortion surface of the problem (see Example 1). For Problem 2, we characterize and solve in closed form the solution (see Theorem 3). An interesting observation that stems from Theorem 3 is that depending on the distortion constraint, we can make the system choose which source (i.e, semantic or observation) to transmit. Simply put, in goal-oriented semantic communication, selecting the type of individual distortion measures or context-dependent fidelity criteria according to the application/task requirements can significantly affect the remote reconstruction of the semantic source.

II. Problem statement and new bounds

We consider a memoryless source described by the tuple $(\mathbf{x}, \mathbf{z})$ with probability distribution $p(x, z)$ in the product alphabet space $\mathcal{X} \times \mathcal{Z}$. The semantic or intrinsic information of the source is in $\mathbf{x}$, which is not directly observable, whereas $\mathbf{z}$ is the noisy observation of the source at the encoder side. The goal is to study how the distortion penalties can affect goal-oriented information transmission and source reconstruction using lossy compression.

Formally, the system model (without the distortion penalties) is illustrated in Fig. 1 and can be interpreted as follows. An *information source* is a sequence of n-length i.i.d RVs $(\mathbf{x}^n, \mathbf{z}^n)$. In this setup, we assume we know $p(x)$ and the transition probability distribution $p(z|x)$. The *encoder (E)* and the *decoder (D)*, are modeled by the mappings

$$\begin{aligned} &f^E : \mathcal{Z}^n \rightarrow \mathcal{W} \\ &g_o^D : \mathcal{W} \rightarrow \widehat{\mathcal{Z}}^n,\; g_s^D : \mathcal{W} \rightarrow \widehat{\mathcal{X}}^n \end{aligned}, \tag{1}$$

where the index set $\mathcal{W} \in \{1, 2, \ldots, M\}$ and (g_o^D, g_s^D) denote the observations and semantic information decoder, respectively.

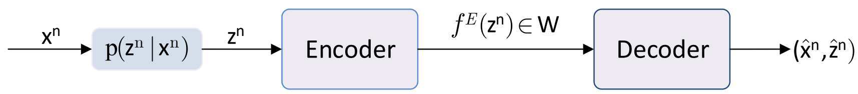

Fig. 1: System model.

We consider two per-letter distortion measures responsible to penalize the semantic and observations information source in Fig. 1, given by $d_s: \ \mathcal{X} \times \widehat{\mathcal{X}} \mapsto [0, \infty)$ and $d_o: \ \mathcal{Z} \times \widehat{\mathcal{Z}} \mapsto [0, \infty)$, respectively, and their corresponding average per-symbol distortions by

$$d_s^n(x^n, \widehat{x}^n) = \frac{1}{n} \sum_{t=1}^{n} d_s(x_i, \widehat{x}_i) \tag{2}$$

$$d_o^n(z^n, \widehat{z}^n) = \frac{1}{n} \sum_{t=1}^{n} d_o(z_i, \widehat{z}_i). \tag{3}$$

The encoding and decoding is done in blocks of length n and the fidelity criterion for the semantic and observable information is the pair of average distortions defined as

$$\Delta_s \triangleq \mathbf{E}\left\{d_s^n(\mathbf{x}^n, \widehat{\mathbf{x}}^n)\right\}, \quad \Delta_o \triangleq \mathbf{E}\left\{d_o^n(\mathbf{z}^n, \widehat{\mathbf{z}}^n)\right\}. \tag{4}$$

Next, we give the definitions of the achievable rates and the infimum of all achievable rates.

Definition 1. *For two distortion levels $D_o \geq 0$, $D_s \geq 0$, a number R is said to be $(D_o, D_s)-$achievable if for an arbitrary $\epsilon > 0$, there exists, for n large enough, a semantic-aware lossy source code $(n, M, \Delta_o, \Delta_s)$ with $M \leq 2^{n(R+\epsilon)}$ such that $\Delta_o \leq D_o + \epsilon$ and $\Delta_s \leq D_s + \epsilon$. Moreover, suppose that sequences of distortion functions $\{(d_o^n, d_s^n): \ n = 1, 2, \ldots\}$ are given. Then,*

$$R(D_o, D_s) = \inf\left\{R : (R, D_o, D_s) \text{ is achievable}\right\}. \tag{5}$$

The goal of the setup in Fig. 1 and of our results in the following sections is to further demonstrate the impact of the fidelity criterion in a remote source coding problem with individual distortion measures.

A. Characterization of the operational rates

The information theoretic characterization of (5) is given by the following lemma.

Lemma 1. *For a given $p(x)$ and $p(z|x)$, the semantic rate distortion function (SRDF) of the setup in Fig. 1 is characterized as follows*

$$R(D_s, D_o) = \inf_{\substack{q(\widehat{z},\widehat{x}|z)\\ \mathbf{E}\left[\widehat{d}_s(\mathbf{z},\widehat{\mathbf{x}})\right]\leq D_s\\ \mathbf{E}\left[d_o(\mathbf{z},\widehat{\mathbf{z}})\right]\leq D_o}} I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}}), \tag{6}$$

where $\widehat{d}_s(z,\widehat{x}) = \sum_{x\in\mathcal{X}} p(x|z)d_s(x,\widehat{x})$*,* $D_s \in [0,\infty]$*,* $D_o \in [0,\infty]$*,*

$$I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}}) \triangleq \mathbf{E}\left[\log\left(\frac{q(\widehat{\mathbf{z}},\widehat{\mathbf{x}}|\mathbf{z})}{\nu(\widehat{\mathbf{z}},\widehat{\mathbf{x}})}\right)\right] \equiv \mathbb{I}(\mathbf{p},\mathbf{q}), \tag{7}$$

with $\mathbb{I}(\mathbf{p},\mathbf{q})$ *demonstrating the functional dependence of the mutual information on* $\{p(z), q(\widehat{z},\widehat{x}|z)\}$.

A detailed proof of Lemma 1 is omitted because the achievability part follows from a special case of the achievability proof of the multiple description source coding problem called robust description [24, Theorem 2] in view of the fact that one can modify the indirect to a direct rate distortion function (RDF) formulation using an amended version for the semantic distortion constraint, i.e., $\mathbf{E}\left[d_s(\mathbf{x}^n,\widehat{\mathbf{x}}^n)\right] = \mathbf{E}\left[\widehat{d}_s(\mathbf{z}^n,\widehat{\mathbf{x}}^n)\right]$ (see e.g., [19, Theorem 1] and references therein). Although the converse part is not provided due to space limitations, it can be easily obtained following standard arguments, see, e.g., [26, p. 316].

We conclude this subsection with certain functional and topological properties of (6).

Remark 1. *The following functional properties of SRDF can be obtained using standard arguments that stem from classical rate distortion theory, see, e.g., [25].*

- **(i)** $R(D_o, D_s)$ *is a non-increasing function of* $D_s \in [0,\infty)$ *and* $D_o \in [0,\infty)$ *and (jointly) convex with respect to* (D_s, D_o).
- **(ii)** *In* (6)*,* $\mathbb{I}(\mathbf{p},\mathbf{q})$ *is a convex functional of* $q(\widehat{z},\widehat{x}|z)$ *for a fixed* $p(z)$.
- **(iii)** *If* $R(D_o, D_s) < \infty$*, then* $R(\cdot)$ *is continuous for* $D_o \in [0,\infty)$ *and* $D_s \in [0,\infty)$.

We conclude this remark by pointing out that the constrained set in (6) *is compact (for both finite or abstract alphabets) and the objective function in* (6) *is lower semi-continuous with respect to* $q(\widehat{z},\widehat{x}|z)$*. As a result, from Weierstrass extreme value theorem, we know that the infimum is attained by a* $q^*(\widehat{z},\widehat{x}|z)$ *and we can formally replace it with minimum in the sequel.*

B. Bounds and conditions for the tightness of these bounds

In what follows, we derive new bounds on (6), as well as information structures (i.e., conditional independence constraints) that allow for these bounds to be tight.

Lemma 2. **(1)** *The optimization problem in* (6) *admits the following bounds:*

$$\max\left\{R(D_s), R(D_o)\right\} = R^L(D_o, D_s) \leq (6) \leq R^U(D_o, D_s) = R(D_o) + R(D_s), \tag{8}$$

where $(R(D_s), R(D_o))$ *represent the standard direct and indirect RDFs obtained via their individual distortion criteria, i.e,,*

$$R(D_o) = \min_{\substack{q(\widehat{z}|z)\\ \mathbf{E}\{d_o(\mathbf{z},\widehat{\mathbf{z}})\}\leq D_o}} I(\mathbf{z};\widehat{\mathbf{z}}), \quad \textit{and} \quad R(D_s) = \min_{\substack{q(\widehat{x}|z)\\ \mathbf{E}\{\widehat{d}_s(\mathbf{z},\widehat{\mathbf{x}})\}\leq D_s}} I(\mathbf{z};\widehat{\mathbf{x}}). \tag{9}$$

(2) $R^L(D_o, D_s)$ *is tight if and only if*

$$\mathbf{z} - \widehat{\mathbf{z}} - \widehat{\mathbf{x}} \qquad \textit{and} \qquad \mathbf{z} - \widehat{\mathbf{x}} - \widehat{\mathbf{z}}, \tag{10}$$

are concurrently satisfied. On the other hand, $R^U(D_o, D_s)$ *is tight if and only if*

$$\widehat{\mathbf{z}} - \mathbf{z} - \widehat{\mathbf{x}}, \tag{11}$$

is satisfied.

Proof: **(1)** Clearly, $R^L(D_o, D_s)$ in (8) corresponds to the best possibly achievable rates because (6) cannot be lower than the best rate achieved in either less constrained problem (individual rate distortion problems in (9)). On the other hand, $R^U(D_o, D_s)$ in (8) is always allowed because we can always minimize assuming $\widehat{\mathbf{z}}$ independent of $\widehat{\mathbf{x}}$ (denoted herein after by $\widehat{\mathbf{z}} \perp \widehat{\mathbf{x}}$), i.e., $\nu(\widehat{z},\widehat{x}) = \nu(\widehat{z})\nu(\widehat{x})$. **(2)** Next, we derive conditional independent constraints (i.e., information structures) which allow the bounds on (6) to be tight. Recall that by the chain rule of mutual information (7) (see, e.g., [26]), we have

$$I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}}) = I(\mathbf{z};\widehat{\mathbf{x}}) + I(\mathbf{z};\widehat{\mathbf{z}}|\widehat{\mathbf{x}}) = I(\mathbf{z};\widehat{\mathbf{z}}) + I(\mathbf{z};\widehat{\mathbf{x}}|\widehat{\mathbf{z}}). \tag{12}$$

From (12), we obtain

$$I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}}) \stackrel{(a)}{\geq} I(\mathbf{z};\widehat{\mathbf{x}}), \quad I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}}) \stackrel{(b)}{\geq} I(\mathbf{z};\widehat{\mathbf{z}}), \tag{13}$$

where (a) follows from the fact that $I(\mathbf{z};\widehat{\mathbf{z}}|\widehat{\mathbf{x}}) \geq 0$ and (b) follows from the fact that $I(\mathbf{z};\widehat{\mathbf{x}}|\widehat{\mathbf{z}}) \geq 0$. Clearly the bounds in (13) are tight *if and only if* for inequality (a) $I^*(\mathbf{z};\widehat{\mathbf{z}}|\widehat{\mathbf{x}}) = 0$, i.e., the Markov chain $\mathbf{z} - \widehat{\mathbf{x}} - \widehat{\mathbf{z}}$ holds and for inequality (b) $I^*(\mathbf{z};\widehat{\mathbf{x}}|\widehat{\mathbf{z}}) = 0$, i.e., the Markov chain $\mathbf{z} - \widehat{\mathbf{z}} - \widehat{\mathbf{x}}$ holds. In view of the previous simple observation we arrive to the following bounds on (6).

Case 1: $R(D_o, D_s) \overset{(c)}{\geq} R(D_s)$ where (c) holds with equality if and only if the condition of inequality (13), (a) holds.

Case 2: $R(D_o, D_s) \overset{(d)}{\geq} R(D_o)$ where (d) holds with equality if and only if the condition of inequality (13), (b) holds.

If both *Case 1* and *Case 2* are concurrently true, i.e., (10) holds, then, from Lagrange duality theorem [33], we can write the individual unconstrained dual problems for $R(D_s)$ and $R(D_o)$ associated with their corresponding Lagrangian multipliers, say (s_1, s_2), and choose the Lagrangian multiplier that corresponds to the maximal rates between $R(D_s)$ and $R(D_o)$ which means precisely $R^L(D_s, D_o)$ in (8).

Now for $R^U(D_s, D_o)$ to be tight, we need to make sure that provided that if $\widehat{\mathbf{x}} \perp \widehat{\mathbf{z}}$, the set of minimizers in (6) coincides with the smaller set of the individual minimizers in both problems in (9), i.e., $q(\widehat{x}, \widehat{z}|z) = q(\widehat{x}|z)q(\widehat{z}|z)$, which is guaranteed *if and only if* (11) is satisfied. In other words, from (7) we obtain

$$\text{(7)} \overset{(e)}{=} \mathbf{E}\left[\log\left(\frac{q(\widehat{\mathbf{z}}|\mathbf{z})q(\widehat{\mathbf{x}}|\mathbf{z})}{\nu(\widehat{\mathbf{z}})\nu(\widehat{\mathbf{x}})}\right)\right] = I(\mathbf{z};\widehat{\mathbf{z}}) + I(\mathbf{z};\widehat{\mathbf{x}}) \tag{14}$$

where (e) follows if and only if (11) is true. The equality in (14) implies that the upper bound in (8) is tight. This completes the proof. ∎

III. Main Results

In this section, we provide the majority of our new results. Before giving our first result, we note that the constrained problem in Lemma 1 can be written as an unconstrained problem via the Lagrange duality theorem [33] as follows

$$R(D_o, D_s) = \max_{\substack{s_1 \leq 0 \\ s_2 \leq 0}} \min_{\substack{q(\widehat{z},\widehat{x}|z) \geq 0 \\ \sum_{\widehat{x},\widehat{z}} q(\widehat{z},\widehat{x}|z)=1}} \left\{ I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}}) - s_1\left(\mathbf{E}\left[\widehat{d}_s(\mathbf{z},\widehat{\mathbf{x}})\right] - D_s\right) - s_2\left(\mathbf{E}\left[d_o(\mathbf{z},\widehat{\mathbf{z}})\right] - D_o\right)\right\} \tag{15}$$

where $s_1 \leq 0$ and $s_2 \leq 0$ are the Lagrange multipliers.

In view of (15) we can prove the following general result.

Theorem 1. *Suppose that* $p(x)$ *and* $p(z|x)$ *are given. Then, the following parametric solutions for* (6) *may appear.*

(i) *If $s_1 < 0$ and $s_2 < 0$, the implicit optimal form of the minimizer that achieves the minimum in* (6) *is*

$$q^*(\widehat{z}, \widehat{x}|z) = \frac{e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z}, \widehat{x})}{\sum_{\widehat{z},\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z}, \widehat{x})} \tag{16}$$

where (s_1, s_2) are the Lagrange multipliers associated with the individual distortion penalties and $\nu^*(\widehat{z}, \widehat{x}) = \sum_z q^*(\widehat{z}, \widehat{x}|z)p(z)$ is the $\widehat{\mathcal{Z}} \times \widehat{\mathcal{X}}$-marginal of the output process $(\widehat{\mathbf{z}}^n, \widehat{\mathbf{x}}^n)$. Moreover, the optimal parametric solution of (6) *when $R(D_s^*, D_0^*) > 0$ is given by*

$$R(D_o^*, D_s^*) = s_1 D_s^* + s_2 D_o^* - \sum_z p(z) \log\left(\sum_{\widehat{z},\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z}, \widehat{x})\right), \tag{17}$$

where

$$D_s^* = \sum_{z,\widehat{x}} \widehat{d}_s(z, \widehat{x}) q^*(\widehat{x}|z) p(z), \tag{18}$$

$$D_o^* = \sum_{z,\widehat{z}} d_o(z, \widehat{z}) q^*(\widehat{z}|z) p(z), \tag{19}$$

and $q^*(\widehat{x}|z) = \sum_{\widehat{z}} q^*(\widehat{z}, \widehat{x}|z)$, $q^*(\widehat{z}|z) = \sum_{\widehat{x}} q^*(\widehat{z}, \widehat{x}|z)$.

(ii) *If $s_1 < 0$, $s_2 = 0$, and $R(D_s^*, D_0^*) > 0$, we obtain*

$$R(D_o^*, D_s^*) \equiv R(D_s^*) = s_1 D_s^* - \sum_z p(z) \log\left(\sum_{\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x})} \nu^*(\widehat{x})\right), \tag{20}$$

where D_s^* is given by (18) *and $q^*(\widehat{x}|z) = \frac{e^{s_1 \widehat{d}_s(z,\widehat{x})} \nu^*(\widehat{x})}{\sum_{\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x})} \nu^*(\widehat{x})}$.*

(iii) *If $s_1 = 0$, $s_2 < 0$, and $R(D_s^*, D_0^*) > 0$, we obtain*

$$R(D_o^*, D_s^*) \equiv R(D_o^*) = s_2 D_o^* - \sum_z p(z) \log\left(\sum_{\widehat{z}} e^{s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z})\right), \tag{21}$$

where D_s^* is given by (19) *and $q^*(\widehat{z}|z) = \frac{e^{s_1 d_o(z,\widehat{z})} \nu^*(\widehat{z})}{\sum_{\widehat{z}} e^{s_1 d_o(z,\widehat{z})} \nu^*(\widehat{z})}$.*

(iv) *If $s_1 = 0$ and $s_2 = 0$, then, $R(D_o^*, D_s^*) = 0$.*

Proof: See Appendix A. ∎

Theorem 1 is pivotal as it can be used in various ways including the derivation of analytical expressions of (6) or for the construction of generalizations of the BA algorithm [34], which can optimally find parametrically the solution of (6) for arbitrary finite alphabet sets and general bounded distortion functions. In the sequel we also study these directions.

A. Binary alphabets with individual Hamming distortions

In what follows, we utilize both Theorem 1 and Lemma 2 to study the case of binary alphabets, i.e., $\mathcal{X}=\mathcal{Z}=\widehat{\mathcal{X}}=\widehat{\mathcal{Z}}=\{0,1\}$ with individual probability of error distortion penalties.

Problem 1. *Suppose in the setup of Fig. 1, the remote source* $\mathbf{x}$ *and the noisy channel of* $\mathbf{z}$ *given* $\mathbf{x}$ *are modeled as follows*

$$\begin{aligned} p(x) &= \begin{pmatrix} p(x=0) \\ p(x=1) \end{pmatrix} = \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}, \\ p(z|x) &= \begin{pmatrix} p(z=0|x=0) & p(z=0|x=1) \\ p(z=1|x=0) & p(z=1|x=1) \end{pmatrix} = \begin{pmatrix} \beta & \gamma \\ 1-\beta & 1-\gamma \end{pmatrix} \end{aligned} \tag{22}$$

where $(\alpha,\beta,\gamma)\in[0,1]\times[0,1]\times[0,1]$, $\beta\neq\gamma$[1] *and*

$$d_s(x,\widehat{x}) = \begin{cases} 0 \text{ if } & x=\widehat{x} \\ 1 \text{ if } & x\neq\widehat{x} \end{cases}, \quad d_o(z,\widehat{z}) = \begin{cases} 0 \text{ if } & z=\widehat{z} \\ 1 \text{ if } & z\neq\widehat{z} \end{cases}.$$

Remark 2. *(Special case) If we assume that the model of the noisy channel in* (22) *becomes "deterministic", i.e.,* $\beta=1$ *and* $\gamma=0$*, then, it can be easily shown that* $\mathbf{z}=\mathbf{x}$ *and the problem recovers the well-known robust description setup for binary alphabets studied in [24, Section VII].*

We provide now a major result of this paper.

Theorem 2. *Consider the setup in Fig. 1 restricted to the given data of Problem 1. Then, the following hold:*

- **(i)** *the necessary and sufficient conditions in* (10) *hold;*
- **(ii)** $R(D_o^*,D_s^*)=R^L(D_o^*,D_s^*)$.

Proof: See Appendix B. ■

The general result of Theorem 2 shows that the rate-splitting bound in (8) is achievable for the specific class of input data under the probability of error distortions assumed in Problem 1.

In what follows, we study the structural solution derived in Theorem 2. Recall that since $R(D_o^*,D_s^*)=R^L(D_o^*,D_s^*)$ holds, then, we need to compute the rate distortion functions in (9).

[1]If $\beta=\gamma$ then we end up having $p(x,z)=p(x)p(z)$, which is not consistent with the setup we assume in Section II hence this scenario is not allowed.

For the direct rate distortion problem with binary source, it is relatively easy to see that the closed form solution will be a straightforward generalization of the analytical solution derived for instance in [26, Theorem 10.3.1], [25, Example 2.7.1], yielding

$$R(D_o^*) = \begin{cases} H_b(\bar{p}) - H_b(D_o), & \text{if } 0 \leq D_o \leq \min\{\bar{p}, 1-\bar{p}\} \\ 0, & \text{if } D_o > \min\{\bar{p}, 1-\bar{p}\} \end{cases} \tag{23}$$

where $\bar{p} = p(z=0)$ is computed in (52) and $H_b(\cdot)$ denotes the binary entropy function. We stress that an optimal closed form solution of the binary indirect RDF (9) is not known, in general, and only bounds exist in the literature, see, e.g., [35]. Nevertheless, one can always use straightforward generalizations of the classical BA iterative schemes to numerically compute the optimal solution. The semantic-aware BA algorithm that we develop in the sequel, includes as a special case this particular case.

Example 1. *In the particular case where the semantic remote source is* i.i.d *Bernoulli*$(\frac{1}{2})$*, i.e.,* $p(x=0) = \frac{1}{2}$*, and the binary channel in* (22) *is symmetric with* $p(z=1|x=1) = \beta = 1-c$ *and crossover probability* $p(z=0|x=1) = \gamma = c, c \in [0, \frac{1}{2})$[2]*, one can easily infer via* (23) *that* $H_b(\bar{p}) = 1$ *bit source/sample and*

$$R(D_o^*) = \begin{cases} 1 - H_b(D_o), & \textit{if } 0 \leq D_o \leq \frac{1}{2} \\ 0, & \textit{if } D_o > \frac{1}{2} \end{cases}. \tag{24}$$

Moreover, for the same input data, it can be shown, see e.g., [25, Exercise 3.8], that

$$R(D_s^*) = \begin{cases} 1 - H_b\left(\frac{D_s - c}{1-2c}\right), & \textit{if } \beta < D_s \leq \frac{1}{2} \\ 0, & \textit{if } D_s > \frac{1}{2} \end{cases}. \tag{25}$$

Substituting (24)*,* (25) *in Theorem 2 we obtain*

$$R(D_o^*, D_s^*) = \max\left\{[1 - H_b(D_o)]^+, \left[1 - H_b\left(\frac{D_s - c}{1-2c}\right)\right]^+\right\} \tag{26}$$

where $[\cdot]^+ = \max\{0, \cdot\}$*. The rate distortion surface for* $c = 0.15$ *is displayed in Fig. 2.*

Based on (26)*, we observe an interesting interplay between* (c, D_s, D_o) *regarding the choice of the maximum achievable rates. In particular, it appears that if* $D_o > \frac{D_s - c}{1-2c}$*, then the system benefits more by encoding subject to a Hamming distortion only the semantic information,*

[2]The result for $c \in [\frac{1}{2}, 1]$ can be treated similarly.

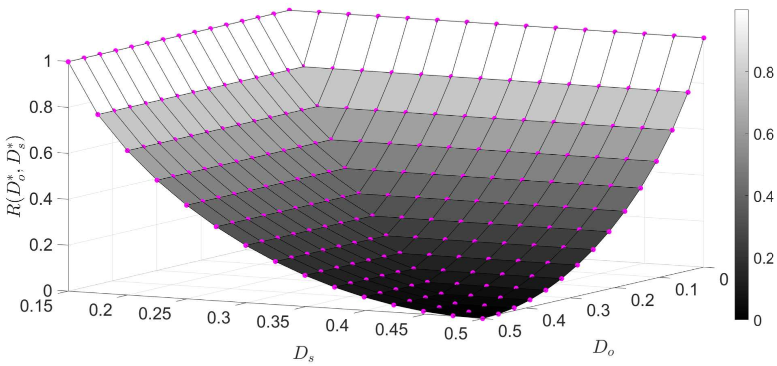

Fig. 2: $R(D_o^*, D_s^*)$ for binary alphabets with an equiprobable semantic source and binary symmetric channel with $c = 0.15$.

therefore the rate is $R(D_s^*)$*; whereas if* $D_o < \frac{D_s - c}{1-2c}$ *the system benefits more by encoding subject to its distortion the observable message of the source with rates* $R(D_o^*)$*. Clearly, if* $D_o = \frac{D_s - c}{1-2c}$*, then, by encoding either the semantic information or the observations does not offer any advantage for any value of the active distortion region.*

Next, we study an extreme scenario to highlight the importance of the distortion measure in the transmission or not of the semantic message. To do it, we consider two different individual distortion constraints (i.e., a standard erasure distortion [26, Exercise 10.7] and a Hamming distortion) to distinguish from Problem 1 where we have identical types of distortion constraints.

Problem 2. *Suppose that in Problem 1, the semantic distortion* $d_s(x, \widehat{x})$ *is replaced by the standard erasure distortion as follows*

$$d_s(x, \widehat{x}) = \begin{cases} 0 \textit{ if } \ x = \widehat{x} \\ 1 \textit{ if } \ x = e \\ \infty, \ x \neq \widehat{x} \end{cases}, \tag{27}$$

where $\widehat{\mathcal{X}} = \{0, e, 1\}$.

Based on the given data of Problem 2, we derive the following solution for the SRDF characterization.

Theorem 3. *Consider the setup in Fig. 1 restricted to Problem 2. Then, for the choice of the semantic distortion penalty in* (27)*, the characterization in* (6) *satisfies the Markov chain* $\mathbf{z}-\widehat{\mathbf{z}}-\widehat{\mathbf{x}}$ *and* $R(D_o^*, D_s^*) = R(D_o^*)$ *which can be explicitly computed via* (23).

Proof: See Appendix C. ∎

Interestingly, the choice of the erasure distortion measure in Theorem 3 demonstrates that the amended distortion of the semantic (remote) source allows only the erasures to be sent, which in turn results into the zero rate of the indirect rate distortion problem. This result comes as a rather extreme case of the general result of Theorem 2 and demonstrates the cardinal role of the distortion penalties into the solution.

B. Semantic-aware Blahut-Arimoto Algorithm

Next, we propose a generalization of the celebrated BA algorithm to treat the case of arbitrary finite alphabet sets with individual distortions.

First we re-state an equivalent way to arrive to the parametric solution of Theorem 1 using instead the alternating minimization approach [34] (see also [36]). This result and the subsequent corollaries form the basis of the semantic-aware BA algorithm that we aim to develop.

Lemma 3. *Let* $s_1 \leq 0$, $s_2 \leq 0$ *be given. Then the following statements hold.*

(1) *The optimal parametric solution of* (6) *can be expressed as follows:*

$$\begin{aligned} R(D_o^*, D_s^*) = s_1 D_s^* + s_2 D_o^* + \min_{\nu(\widehat{z},\widehat{x})} \min_{q(\widehat{z},\widehat{x}|z)} \Bigg[& \sum_{z} \sum_{\widehat{z},\widehat{x}} \log\left(\frac{q(\widehat{z},\widehat{x}|z)}{\nu(\widehat{z},\widehat{x})}\right) q(\widehat{z},\widehat{x}|z) p(z) \\ & - s_1 \sum_{z} \sum_{\widehat{z},\widehat{x}} \widehat{d}_s(z,\widehat{x}) q(\widehat{z},\widehat{x}|z) p(z) - s_2 \sum_{z} \sum_{\widehat{z},\widehat{x}} d_o(z,\widehat{z}) q(\widehat{z},\widehat{x}|z) p(z) \Bigg] \end{aligned} \tag{28}$$

where D_s^* *and* D_o^* *are given by* (18) *and* (19)*, respectively.*

(2) *For fixed* $q(\widehat{z},\widehat{x}|z)$*, the right hand side (RHS) in* (28) *is minimized by*

$$\nu^*(\widehat{z},\widehat{x}) = \sum_{z} q(\widehat{z},\widehat{x}|z) p(z).$$

(3) *For fixed* $\nu(\widehat{z},\widehat{x})$ *the RHS in* (28) *is minimized by* $q^*(\widehat{z},\widehat{x}|z)$ *given by* (16)*, i.e.,*

$$q^*(\widehat{z},\widehat{x}|z) = \frac{e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu(\widehat{z},\widehat{x})}{\sum_{\widehat{z},\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu(\widehat{z},\widehat{x})}.$$

Proof: **(1)** Due to the convexity and monotonicity of the optimization problem in (6) (see Remark 1), we can reformulate it as an unconstrained problem as follows

$$R(D_s,D_o)=\min_{q(\widehat{z},\widehat{x}|z)}\left[I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}})-s_1\left(\mathbf{E}\left[\widehat{d}_s(\mathbf{z},\widehat{\mathbf{x}})\right]-D_s\right)-s_2\left(\mathbf{E}\left[d_o(\mathbf{z},\widehat{\mathbf{z}})\right]-D_o\right)\right] \tag{29}$$

where the expectation operator $\mathbf{E}[\cdot]$ is taken with respect to the joint distribution $p(z,\widehat{z},\widehat{x})$. Using (29), the double minimization in (28) follows by applying [37, Theorem 5.2.6], namely, we can reformulate mutual information $I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}})$ as follows

$$I(\mathbf{z};\widehat{\mathbf{z}},\widehat{\mathbf{x}})=\min_{\bar{\nu}(\widehat{z},\widehat{x})}\sum_{z}\sum_{\widehat{z},\widehat{x}}\log\left(\frac{q(\widehat{z},\widehat{x}|z)}{\bar{p}(\widehat{z},\widehat{x})}\right), \tag{30}$$

where the minimization is over an arbitrary chosen output marginal distribution $\bar{\nu}(\widehat{z},\widehat{x})$ defined on $\widehat{\mathcal{Z}}\times\widehat{\mathcal{X}}$ and the minimization over $\nu(\widehat{z},\widehat{x})$ follows from the condition for equality in that theorem. **(2)**, **(3)** To optimize, we use Karush-Kuhn-Tucker (KKT) conditions similar to the ones utilized for the derivation of Theorem 1 hence we omit it. ■

In Lemma 3 we consider the Lagrange multipliers (s_1,s_2) to be non-positive. One can easily obtain all the special cases discussed in Theorem 1 by choosing to have only one or none of the distortion constraints active. Clearly, if we consider optimizing jointly with respect to $\{q(\widehat{z},\widehat{x}|z),\nu(\widehat{z},\widehat{x})\}$ in (28), then, the result of Lemma 3, will coincide to the general result of Theorem 1.

Next, we give two corollaries that are instrumental in the development of our algorithm because they give the implicit solution of $\{q(\widehat{z},\widehat{x}|z),\nu(\widehat{z},\widehat{x})\}$ parametrized by (s_1,s_2).

Corollary 1. *If $q(\widehat{z},\widehat{x}|z)$ achieves a point on $R(D_o,D_s)$ parametrized by (s_1,s_2), then $q(\widehat{z},\widehat{x}|z)$ is given by* (16) *and*

$$\nu(\widehat{z},\widehat{x})=\nu(\widehat{z},\widehat{x})\sum_{z}p(z)\frac{e^{s_1\widehat{d}_s(z,\widehat{x})+s_2d_o(z,\widehat{z})}}{\sum_{\widehat{z},\widehat{x}}e^{s_1\widehat{d}_s(z,\widehat{x})+s_2d_o(z,\widehat{z})}\nu(\widehat{z},\widehat{x})}. \tag{31}$$

Proof: The proof follows using the simultaneous satisfaction of Lemma 3, **(2)**, **(3)**. ■

Corollary 2. *In terms of (s_1,s_2), $R(D_o,D_s)$ can be written as follows:*

$$R(D_o^{s_1},D_s^{s_2})=s_1D_s^{s_1}+s_2D_o^{s_2}+\min_{\nu(\widehat{z},\widehat{x})}\left[-\sum_{z}p(z)\log\left(\sum_{\widehat{z},\widehat{x}}e^{s_1\widehat{d}_s(z,\widehat{x})+s_2d_o(z,\widehat{z})}\nu(\widehat{z},\widehat{x})\right)\right], \tag{32}$$

with

$$D_s^{s_1} = \sum_{z,\widehat{z},\widehat{x}} \widehat{d}_s(z,\widehat{x}) \frac{e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z},\widehat{x})}{\sum_{\widehat{z},\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z},\widehat{x})} p(z),$$

$$D_o^{s_2} = \sum_{z,\widehat{z},\widehat{x}} d_o(z,\widehat{z}) \frac{e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z},\widehat{x})}{\sum_{\widehat{z},\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z},\widehat{x})} p(z)$$

where $\nu^*(\widehat{z},\widehat{x})$ achieves $R(D_o^{s_1}, D_s^{s_2})$.

Proof: The proof follows by substituting $q^*(\widehat{z},\widehat{x}|z)$ of Lemma 3, **(3)** in Lemma 3, **(1)**. ■

We continue our analysis with a general theorem reminiscent of the one derived for the classical BA algorithm. This theorem demonstrated the convergence of the algorithm to the optimal limit point. In the theorem we denote $\mathbf{p} = p(z)$, $\boldsymbol{\nu} \triangleq \nu(\widehat{z},\widehat{x})$, and $\mathbf{q} = q(\widehat{z},\widehat{x}|z)$ the probability vector of those distributions, e.g., for $\mathcal{Z} = \{0, 1, \ldots, M_1\},\ M_1 \in \mathbb{Z}_+,\ \mathbf{p} = (p(z=0), \ldots, p(z=M_1))$.

Theorem 4. *Let the parameters $s_1 \leq 0$ and $s_2 \leq 0$ be given and denote $A(z,\widehat{z},\widehat{x}) = e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})}$ and $k \geq 1$ the number of iterations. Let any $\boldsymbol{\nu}^{(0)}$ be any initial marginal probability distribution on $\widehat{\mathcal{X}} \times \widehat{\mathcal{Z}}$ (given as a probability vector) with all the components to be nonzero. Let $\boldsymbol{\nu}^{(k+1)}$ be given in terms of $\boldsymbol{\nu}^{(k)}$ as follows*

$$\nu^{(k+1)}(\widehat{z},\widehat{x}) = \nu^{(k)}(\widehat{z},\widehat{x}) \sum_z \frac{p(z) A(z,\widehat{z},\widehat{x})}{\sum_{\widehat{z},\widehat{x}} A(z,\widehat{z},\widehat{x}) \nu^{(k)}(\widehat{z},\widehat{x})} \tag{33}$$

Then,

$$D_s(\mathbf{q}(\boldsymbol{\nu}^{(k)})) \longrightarrow D_s^{s_1}, \quad D_o(\mathbf{q}(\boldsymbol{\nu}^{(k)})) \longrightarrow D_o^{s_2}, \quad \mathbb{I}(\mathbf{p}; \mathbf{q}(\boldsymbol{\nu}^{(k)})) \longrightarrow R(D_s^{s_1}, D_o^{s_2}) \ \ \text{as } k \to \infty, \tag{34}$$

where

$$\mathbf{q}(\boldsymbol{\nu}^{(k)}) = \frac{A(z,\widehat{z},\widehat{x}) \nu^{(k)}(\widehat{z},\widehat{x})}{\sum_{\widehat{z},\widehat{x}} A(z,\widehat{z},\widehat{x}) \nu^{(k)}(\widehat{z},\widehat{x})}$$

and $(D_s^{s_1}, D_o^{s_2}, R(D_s^{s_1}, D_o^{s_2}))$ is a point on $R(D_o, D_s)$ parametrized by (s_1, s_2).

Proof: See Appendix D. ■

In order to develop an algorithm to compute (6) for arbitrary finite alphabets sets, we also need a termination criterion, for which we need a generalization of [37, Theorem 6.3.9] to our setup. This is stated next as a lemma.

Lemma 4. *An alternative definition of the RDF in* (6) *is the following*

$$R(D_o, D_s) = \max_{s_1 \leq 0, s_2 \leq 0, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}_{s_1,s_2}} \left[s_1 D_s + s_2 D_o + \sum_{z \in \mathcal{Z}\{0,1,\ldots,M_1\}} p(z) \log \lambda(z) \right], \tag{35}$$

where $\boldsymbol{\Lambda}_{s_1,s_2}$ *is the set of all vectors* $\boldsymbol{\lambda} = (\lambda(z=0), \ldots, \lambda(z=M_1))$ *with non-negative elements that satisfy inequality constraints* $\sum_{z \in \mathcal{Z}} p(z)\lambda(z) e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \leq 1$.

Proof: We sketch the proof because it can be easily understood. Let $s_1 \leq 0$, $s_2 \leq 0$, $\lambda \in \boldsymbol{\Lambda}_{s_1,s_2}$ and $\mathbf{q} \in \{q(\widehat{z}, \widehat{x}|z) : \mathbf{E}[\widehat{d}_s(\mathbf{z}, \widehat{\mathbf{x}})] \leq D_s, \mathbf{E}[d_o(\mathbf{z}, \widehat{\mathbf{z}})] \leq D_s\}$. Then, from these conditions, it can be proved that $\mathbb{I}(\mathbf{p}, \mathbf{q}) - s_1 D_s - s_2 D_o - \sum_{z \in \mathcal{Z}} p(z) \log \lambda(z) \geq 0$, which implies that $R(D_o, D_s) \geq s_1 D_s + s_2 D_o + \sum_{z \in \mathcal{Z}} p(z) \log \lambda(z)$. However, we know that for some (s_1, s_2, λ) this inequality holds with equality, i.e., $\lambda(z) = \left(\sum_{\widehat{z},\widehat{x}} e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z}, \widehat{x}) \right)^{-1}$ with (s_1, s_2) and $\nu^*(\cdot)$ achieving $R(D_o, D_s)$. ∎

The next theorem gives bounds on the RDF, which allow to estimate in the algorithm the residual error at each iteration. This theorem serves as a stopping criterion for our algorithm.

Theorem 5. *Let the parameters* $s_1 \leq 0$*,* $s_2 \leq 0$ *be given and let* $A(z, \widehat{z}, \widehat{x}) = e^{s_1 \widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})}$*. Suppose that* $\boldsymbol{\nu}$ *is any output probability vector and let*

$$c(\widehat{z}, \widehat{x}) = \sum_z p(z) \frac{A(z, \widehat{z}, \widehat{x})}{\sum_{\widehat{z},\widehat{x}} A(z, \widehat{z}, \widehat{x}) \nu(\widehat{z}, \widehat{x})}. \tag{36}$$

Then at the points

$$D_s = \sum_{z,\widehat{z},\widehat{x}} \widehat{d}_s(z, \widehat{x}) \frac{A(z, \widehat{z}, \widehat{x}) \nu(\widehat{z}, \widehat{x})}{\sum_{\widehat{z},\widehat{x}} A(z, \widehat{z}, \widehat{x}) \nu(\widehat{z}, \widehat{x})} p(z), \quad D_o = \sum_{z,\widehat{z},\widehat{x}} d_o(z, \widehat{z}) \frac{A(z, \widehat{z}, \widehat{x}) \nu(\widehat{z}, \widehat{x})}{\sum_{\widehat{z},\widehat{x}} A(z, \widehat{z}, \widehat{x}), \nu(\widehat{z}, \widehat{x})} p(z), \tag{37}$$

and by setting $\Gamma \triangleq s_1 D_s + s_2 D_o - \sum_z p(z) \log \left(\sum_{\widehat{z},\widehat{x}} A(z, \widehat{z}, \widehat{x}) \nu(\widehat{z}, \widehat{x}) \right)$*, we have the following bounds*

$$R(D_o, D_s) \leq \Gamma - \sum_{\widehat{z},\widehat{x}} \nu(\widehat{z}, \widehat{x}) c(\widehat{z}, \widehat{x}) \log c(\widehat{z}, \widehat{x}), \tag{38}$$

$$R(D_o, D_s) \geq \Gamma - \max_{\widehat{z},\widehat{x}} \log c(\widehat{z}, \widehat{x}). \tag{39}$$

Proof: See Appendix E. ∎

We are now ready to give the semantic-aware BA algorithm for the setup in Fig. 1. This is illustrated in Algorithm 1.

Algorithm 1 Semantic-aware Blahut-Arimoto algorithm

Initialize: Choose $\mathbf{p} = (p(z = j),\ j = 0, 1, \ldots, N_1)$, and $\mathbf{p}' \triangleq (p(z = j|x = i),\ i = 0, 1, \ldots, M_1,\ j = 0, 1, \ldots, N_1)$, $M_1 \in \mathbb{Z}_+$, $N_1 \in \mathbb{Z}_+$; $s_1 \leq 0$; $s_2 \leq 0$; error tolerance ϵ; choose the alphabet of $\widehat{\mathcal{Z}}, \widehat{\mathcal{X}}$; choose the distortion functions $d_s(x, \widehat{x})$ and $d_o(z, \widehat{z})$; initial output probability vector $\boldsymbol{\nu}^{(0)}$ (all the elements to be non-zero).

Compute $\widehat{d}_s(z, \widehat{x}) = \sum_{x \in \mathcal{X}} p(x|z) d_s(x, \widehat{x})$, $\mathcal{X} = \{0, 1, \ldots, R_1\}$, $R_1 \in \mathbb{Z}_+$.

Compute $A(z, \widehat{z}, \widehat{x}) = e^{s_1 \widehat{d}_s(z, \widehat{x}) + s_2 d_o(z, \widehat{z})}$, for all $\widehat{z} \in \widehat{\mathcal{Z}},\ \widehat{x} \in \widehat{\mathcal{X}}$.

while flag $= 0$ **do**

Step 1: Set $k = 1$.

Step 2: For all $\widehat{z} \in \widehat{\mathcal{Z}}, \widehat{x} \in \widehat{\mathcal{X}}$,

Compute $c^{(k)}(\widehat{z}, \widehat{x}) = \sum_{z \in \mathcal{Z}} p(z) \frac{A(z, \widehat{z}, \widehat{x})}{\sum_{\widehat{z}, \widehat{x}} A(z, \widehat{z}, \widehat{x}) \nu^{(k-1)}(\widehat{z}, \widehat{x})}$.

Compute $\nu^{(k)}(\widehat{z}, \widehat{x}) = \nu^{(k-1)}(\widehat{z}, \widehat{x}) c^{(k)}(\widehat{z}, \widehat{x})$.

Compute diff$=\max_{\widehat{z}, \widehat{x}} \log c^{(k)}(\widehat{z}, \widehat{x}) - \sum_{\widehat{z}, \widehat{x}} \nu^{(k)}(\widehat{z}, \widehat{x}) \log c^{(k)}(\widehat{z}, \widehat{x})$.

if diff$> \epsilon$ **then**

Set $k = k + 1$ and return to Step 2.

else

flag $\leftarrow 1$

end if

end while

Output: $q(\widehat{z}, \widehat{x}|z) = \frac{A(z, \widehat{z}, \widehat{x}) \nu^{(k)}(\widehat{z}, \widehat{x})}{\sum_{\widehat{z}, \widehat{x}} A(z, \widehat{z}, \widehat{x}) \nu^{(k)}(\widehat{z}, \widehat{x})}$; $D_s = \sum_{z, \widehat{z}, \widehat{x}} \widehat{d}_s(z, \widehat{x}) q(\widehat{z}, \widehat{x}|z) p(z)$; $D_o = \sum_{z, \widehat{z}, \widehat{x}} d_o(z, \widehat{z}) q(\widehat{z}, \widehat{x}|z) p(z)$; $R(D_s, D_o) = s_1 D_s + s_2 D_o - \sum_z p(z) \log \left(\sum_{\widehat{z}, \widehat{x}} A(z, \widehat{z}, \widehat{x}) \nu^{(k)}(\widehat{z}, \widehat{x}) \right) - \sum_{\widehat{z}, \widehat{x}} \nu^{(k)}(\widehat{z}, \widehat{x}) \log c^{(k)}(\widehat{z}, \widehat{x})$.

Clearly Algorithm 1 is a generalization of the well-known BA algorithm [34]. By choosing appropriately the value of the Lagrangians (s_1, s_2) one can recover the classical BA algorithm or its extension to the indirect rate distortion problem.

IV. Revisiting the problem for jointly Gaussian i.i.d processes

In this section, we revisit the problem for jointly Gaussian i.i.d processes, which have recently been studied in [20]. Therein, the authors consider $(\mathbf{x}, \mathbf{z})$ to be zero mean jointly Gaussian random vectors such that $\mathbf{x}\in\mathbb{R}^l$, $\mathbf{z}\in\mathbb{R}^m$, $m\neq l$, with covariance matrix

$$\Sigma_{(\mathbf{x},\mathbf{z})}\triangleq\begin{bmatrix}\Sigma_{\mathbf{x}} & \Sigma_{\mathbf{xz}}\\ \Sigma_{\mathbf{xz}}^{\mathrm{T}} & \Sigma_{\mathbf{z}}\end{bmatrix}, \tag{40}$$

with *given* $p(z)\sim\mathcal{N}(0;\Sigma_{\mathbf{z}}), \Sigma_{\mathbf{z}}\succeq 0$ and *given* $p(x|z)$ represented by a linear state-observations model, i.e., $\mathbf{x}=H\mathbf{z}+\mathbf{n}$, such that $p(x|z)\sim\mathcal{N}(H\mathbf{z};\Sigma_{\mathbf{n}})$, where $H=\Sigma_{\mathbf{xz}}\Sigma_{\mathbf{z}}^{-1}$ and $\Sigma_{\mathbf{n}}=\Sigma_{\mathbf{x}}-\Sigma_{\mathbf{xz}}\Sigma_{\mathbf{z}}^{-1}\Sigma_{\mathbf{xz}}^{\mathrm{T}}\succeq 0$. Then, assuming in (2), (3), that $d_o(z_i,\widehat{z}_i)=||z_i-\widehat{z}_i||_2^2$ and $d_s(x_i,\widehat{x}_i)=||x_i-\widehat{x}_i||_2^2$ (or the quadratic Euclidean norm), respectively, SRDF (6) is as follows

$$R(D_o,D_s)=\min_{\substack{\Sigma_{\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}}\succeq 0\\ \mathrm{trace}(H\Sigma_{\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}}H^{\mathrm{T}})\leq D_s-\mathrm{trace}(\Sigma_{\mathbf{n}})\\ \mathrm{trace}(\Sigma_{\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}})\leq D_o}}\frac{1}{2}\left[\log\left(\frac{\det(\Sigma_{\mathbf{z}})}{\det(\Sigma_{\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}})}\right)\right]^+, \tag{41}$$

where $\det(\cdot)$ is the determinant of a matrix, the conditional covariance defined by $\Sigma_{\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}}\triangleq\mathbf{E}\left[(\mathbf{z}-\mathbf{E}\left[\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}\right])(\mathbf{z}-\mathbf{E}\left[\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}\right])^{\mathrm{T}}\right]$ is a design variable in the optimization in (41) and $[\cdot]^+=\max\{0,\cdot\}$. Log-determinant convex problems of the form (41), can be solved optimally using standard convex programming solvers that exist in the CVX platform [38].

Next,we state some important technical remarks about the result of [20].

Remark 3. **(1)** *For scalar jointly Gaussian RVs with individual MSE distortion constraints, [19, Corollary 1] "indirectly" showed that $R^L(D_o^*,D_s^*)$ in Lemma 2 is tight and also the information structures that are necessary and sufficient for this bound are satisfied.* **(2)** *If in* (41) *we assume that H is an orthogonal matrix, i.e., $H^{\mathrm{T}}H=I_{m\times m}$ (m-dimensional identity matrix), then, as [20] recognized, both distortion constraints simplify to* $\mathrm{trace}(\Sigma_{\mathbf{z}|\widehat{\mathbf{x}},\widehat{\mathbf{z}}})\leq\min\{D_s-\mathrm{trace}(\Sigma_{\mathbf{n}}),D_o\}$ *(consequence of Lagrange duality theorem) which is nothing more than $R^L(D_o,D_s)$ of Lemma 2.* **(3)** *The problem setup described in [20, Section IV] is not the same with the problem statement of Section II because according to the latter, we are given $p(x)$ and $p(z|x)$ and we want to design $p(x|z)$ and $p(z)$, respectively. In other words, following Section II, the linear state-observations model described in [20, eq. (19)] should be designed, i.e., matrix $H\in\mathbb{R}^{l\times m}$ and the covariance matrix $\Sigma_{\mathbf{n}}$ need to be designed instead of been freely chosen.*

Under the technical observation of Remark 3, **(3)**, in what follows, we re-formulate the problem to be consistent with Section II. Suppose that $(\mathbf{x},\mathbf{z})$ are zero mean jointly Gaussian random vectors such that $\mathbf{x}\in\mathbb{R}^l$, $\mathbf{z}\in\mathbb{R}^m$, $m\leq l$, with covariance matrix given by (40). Moreover, assume that $\mathbf{x}\sim\mathcal{N}(0,\Sigma_{\mathbf{x}})$, $\Sigma\succ 0$ and $p(z|x)$ is modeled by the linear realization of the form

$$\mathbf{z}=A\mathbf{x}+\mathbf{s}, \tag{42}$$

where $A \in \mathbb{R}^{m\times l}$ assumed to be full row rank matrix, and $\mathbf{s} \sim \mathcal{N}(0;\Sigma_{\mathbf{s}})$, $\Sigma_{\mathbf{s}} \succeq 0$. Then, the conditional Gaussian distribution $p(x|z) \sim \mathcal{N}(\mathbf{E}[\mathbf{x}|\mathbf{z}];\Sigma_{\mathbf{x}|\mathbf{z}})$, where the conditional mean $\mathbf{E}[\mathbf{x}|\mathbf{z}] = \Sigma_{\mathbf{xz}}\Sigma_{\mathbf{z}}^{-1}\mathbf{z}$ and the conditional covariance $\Sigma_{\mathbf{x}|\mathbf{z}} = \Sigma_{\mathbf{x}} - \Sigma_{\mathbf{xz}}\Sigma_{\mathbf{z}}^{-1}\Sigma_{\mathbf{xz}}^{\mathrm{T}} \succeq 0$. However, since $\Sigma_{\mathbf{xz}} = \mathbf{E}[\mathbf{x}(A\mathbf{x}+\mathbf{s})^{\mathrm{T}}] = \Sigma_{\mathbf{x}}A^{\mathrm{T}}$ and $\Sigma_{\mathbf{z}} = A\Sigma_{\mathbf{x}}A^{\mathrm{T}} + \Sigma_{\mathbf{s}}$, we can further infer that $\mathbf{E}[\mathbf{x}|\mathbf{z}] = \Sigma_{\mathbf{x}}A^{\mathrm{T}}(A\Sigma_{\mathbf{x}}A^{\mathrm{T}}+\Sigma_{\mathbf{s}})^{-1}\mathbf{z}$ and similarly, one can apply the same values in $\Sigma_{\mathbf{x}|\mathbf{z}}$.

Remark 4. *If one wants to draw connections with the problem studied in [20, Section IV], it is clear that in our approach we obtain, by design, the matrices* $(H,\Sigma_{\mathbf{n}})$*, i.e.,*

$$H = \Sigma_{\mathbf{x}}A^{T}(A\Sigma_{\mathbf{x}}A^{T}+\Sigma_{\mathbf{s}})^{-1} \text{ and } \Sigma_{\mathbf{n}} = \Sigma_{\mathbf{x}|\mathbf{z}} = \Sigma_{\mathbf{x}}A^{T}(A\Sigma_{\mathbf{x}}A^{T}+\Sigma_{\mathbf{s}})^{-1}A\Sigma_{\mathbf{x}}, \tag{43}$$

instead of freely choosing them. Hence we are consistent with the problem setup introduced in Section II.

The rest of the analysis follows using similar arguments to [20] resulting into the optimization problem of (41) with $(H,\Sigma_{\mathbf{n}})$ designed as in (43).

We provide next an example where we first compute and illustrate the solution of (41) and then we compare with the solution of the bounds in (8).

Example 2. *Suppose that* $p(x) \sim \mathcal{N}(0;\Sigma_{\mathbf{x}})$ *and* $p(z|x) \sim \mathcal{N}(A\mathbf{x};\Sigma_{\mathbf{s}})$ *are randomly chosen such that*

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} 1.1504 & 1.2689 & 0.5826 & 1.0533 \\ 1.2689 & 1.6594 & 0.5395 & 1.0366 \\ 0.5826 & 0.5395 & 0.4066 & 0.4463 \\ 1.0533 & 1.0366 & 0.4463 & 1.3054 \end{bmatrix}, \quad A = \begin{bmatrix} 0.0240 & 0.2593 & 0.2835 & 0.4405 \\ 0.5589 & 0.4151 & 0.6931 & 0.1569 \end{bmatrix},$$

with $\Sigma_{\mathbf{s}} = \begin{bmatrix} 0.9055 & 0.3401 \\ 0.3401 & 0.1431 \end{bmatrix}$ *.Moreover, using* (42)*, we can compute* $\Sigma_{\mathbf{z}} = \begin{bmatrix} 1.7771 & 1.8141 \\ 1.8141 & 2.7830 \end{bmatrix}$, *and, via* (43)*, we can compute the design variables* $(H,\Sigma_{\mathbf{n}})$ *as follows*

$$H = \begin{bmatrix} -0.2483 & 0.7866 \\ -0.3209 & 0.9043 \\ -0.2049 & 0.4574 \\ -0.0064 & 0.5550 \end{bmatrix}, \quad \Sigma_{\mathbf{n}} = \begin{bmatrix} 0.0275 & 0.0130 & -0.0108 & 0.0946 \\ 0.0130 & 0.2535 & -0.1261 & -0.0302 \\ -0.0108 & -0.1261 & 0.0897 & -0.0510 \\ 0.0946 & -0.0302 & -0.0510 & 0.4608 \end{bmatrix}. \tag{44}$$

Solving (41) *for* $\Sigma_{\mathbf{z}|\widehat{\mathbf{z}},\widehat{\mathbf{x}}} \succ 0$ *using the designed values obtained in* (44)*, results into Fig. 3. For this simulation, we note that the optimization problem* (41) *is well-defined if* $\Sigma_{\mathbf{z}|\widehat{\mathbf{z}},\widehat{\mathbf{x}}} \succ 0$ *and* $D_s > D_s^{\min} = \mathrm{trace}(\Sigma_{\mathbf{n}}) = 0.8315$ *(for this example).*

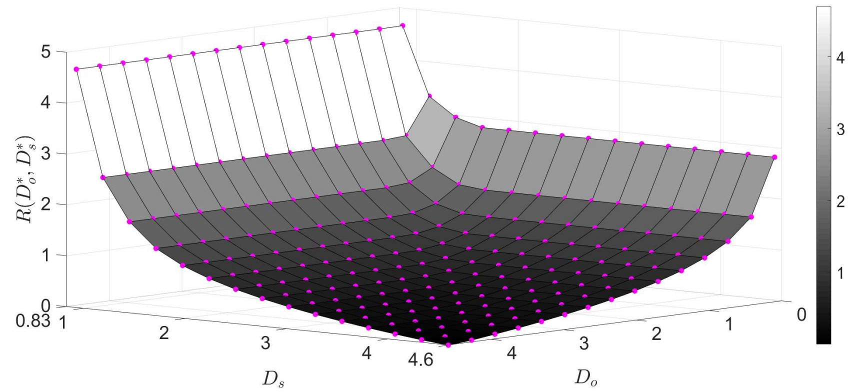

Fig. 3: $R(D_o^*, D_s^*)$ for jointly Gaussian RVs with individual MSE distortion constraints.

On the other hand, using the bounds of 8, we should first solve independently the direct and the indirect rate distortion problem as these are defined in (9) *and, then, for the lower bound on SRDF to choose the maximum between the two and for the upper bound to simply add the two rates. Since these two problems independently are already well studied in the literature, see, e.g., [21], [22], [25], [26], [39], we do not get into the details of their derivations but instead we only state the corresponding optimization problems for jointly Gaussian RVs and individual MSE distortion constraints. In particular,*

$$R^L(D_o, D_s) = \max\{R(D_s), R(D_o)\}, \quad \textit{and} \quad R^U(D_o, D_s) = R(D_s) + R(D_o), \tag{45}$$

with

$$R(D_s) = \min_{\substack{\Sigma_{\mathbf{z}|\widehat{\mathbf{x}}} \succ 0 \\ \mathrm{trace}(H\Sigma_{\mathbf{z}|\widehat{\mathbf{x}}}H^T) \leq D_s - \mathrm{trace}(\Sigma_{\mathbf{n}})}} \frac{1}{2}\log\left(\frac{\det(\Sigma_{\mathbf{z}})}{\det(\Sigma_{\mathbf{z}|\widehat{\mathbf{x}}})}\right), \tag{46}$$

$$R(D_o) = \min_{\substack{\Sigma_{\mathbf{z}|\widehat{\mathbf{z}}} \succ 0 \\ \mathrm{trace}(\Sigma_{\mathbf{z}|\widehat{\mathbf{z}}}) \leq D_o}} \frac{1}{2}\log\left(\frac{\det(\Sigma_{\mathbf{z}})}{\det(\Sigma_{\mathbf{z}|\widehat{\mathbf{z}}})}\right), \tag{47}$$

whereas the design variables for (46) *and* (47) *correspond to the conditional covariance matrices* $\Sigma_{\mathbf{z}|\widehat{\mathbf{x}}} \triangleq \mathbf{E}\left[(\mathbf{z} - \mathbf{E}\left[\mathbf{z}|\widehat{\mathbf{x}}\right])(\mathbf{z} - \mathbf{E}\left[\mathbf{z}|\widehat{\mathbf{x}}\right])^T\right]$ *and* $\Sigma_{\mathbf{z}|\widehat{\mathbf{z}}} \triangleq \mathbf{E}\left[(\mathbf{z} - \mathbf{E}\left[\mathbf{z}|\widehat{\mathbf{z}}\right])(\mathbf{z} - \mathbf{E}\left[\mathbf{z}|\widehat{\mathbf{z}}\right])^T\right]$*, respectively. In Table I, we compare for random values of the pair* (D_o, D_s)*, the solution of* (41) *with the values of the bounds obtained in* (45)*. From our numerical results we observe that for certain distortion pairs* (D_o, D_s)*, sometimes* $R^L(R_o^*, R_s^*) < R(D_o^*, D_s)$ *and sometimes* $R^L(R_o^*, R_s^*) = R(D_o^*, D_s^*)$ *whereas for this example* $R(D_o^*, D_s) < R^U(D_o^*, D_s^*)$*. These results verify numerically the bounds in Lemma 2.*

Type of Solution	Distortion Values			
	$(D_o, D_s) = (0.6, 0.9)$	$(D_o, D_s) = (1.1, 0.85)$	$(D_o, D_s) = (0.2, 1.9)$	$(D_o, D_s) = (1.5, 1.9)$
$R(D_o^*, D_s^*)$ via (41)	3.4518	5.1417	3.6853	1.0109
$R^L(D_o^*, D_s^*)$ via (45)	3.2677	5.1417	3.6853	0.9584
$R^U(D_o^*, D_s^*)$ via (45)	5.3680	6.4252	4.5936	1.8666

TABLE I: Comparison of the optimal values obtained via (41), (45). The simulation studies are performed on CVX platform.

V. Conclusions and Ongoing Research

A variant of a robust description source coding problem with two individual criteria, which is a relevant model for goal-oriented semantic communication, was studied here. We derived bounds on the semantic rate distortion function, as well as necessary and sufficient conditions for their tightness. We then proved a general theorem that provides the optimal solution of the characterization for this problem. Capitalizing on these results, we examined the structure of the solution for the case of general binary alphabets under Hamming and erasure distortion criteria. We also constructed a semantic-aware BA algorithm for the computation of any finite alphabet sets under individual distortion criteria. Finally, we provided new insights overlooked so far for multidimensional jointly Gaussian RVs with individual MSE distortion constraints. A key takeaway from our results is that the class of the fidelity criteria may significantly affect the system behavior irrespective of its task, hence it should be chosen appropriately.

Our ongoing research activity builds on two directions. First, constructing analytical examples where the presence of the additional distortion fidelity affects semantic communication. Second, we seek to generalize our goal-oriented communication setup beyond separable distortion criteria, since in real world applications one is more interested in relations between the input and output symbols that are highly non-linear.

APPENDIX A
PROOF OF THEOREM 1

The fully unconstrained problem of (6) using (15) is as follows

$$\mathcal{L}(\{s_i\}_{i=1}^2, \lambda(z), \mu(z,\widehat{z},\widehat{x})) = \sum_{x,\widehat{z},\widehat{z}} \log\left(\frac{q(\widehat{z},\widehat{x}|z)}{\nu(\widehat{z},\widehat{x})}\right) q(\widehat{z},\widehat{x}|z)p(z) - s_1\left(\mathbf{E}\left[\widehat{d}_s(\mathbf{z},\widehat{\mathbf{x}})\right] - D_s\right)$$

$$- s_2\left(\mathbf{E}\left[d_o(\mathbf{z},\widehat{\mathbf{z}})\right] - D_o\right) + \sum_z \lambda(z)\left(\sum_{\widehat{z},\widehat{x}} q(\widehat{z},\widehat{x}|z) - 1\right) - \sum_{z,\widehat{z},\widehat{x}} \mu(z,\widehat{z},\widehat{x}) q(\widehat{z},\widehat{x}|z), \tag{48}$$

where $s_1 \leq 0, s_2 \leq 0$ are the Lagrangian multipliers associated with the individual distortion constraints $\mathbf{E}\left[\widehat{d}_s(\mathbf{z},\widehat{\mathbf{x}})\right] \leq D_s$ and $\mathbf{E}\left[d_o(\mathbf{z},\widehat{\mathbf{z}})\right] \leq D_o$, respectively, whereas $\lambda(z) \geq 0$ is associated with the equality constraint $\sum_{\widehat{z},\widehat{x}} q(\widehat{z},\widehat{x}|z) = 1$, and $\mu(z,\widehat{z},\widehat{x}) \geq 0$ is responsible for the inequality constraint $q(\widehat{z},\widehat{x}|z) \geq 0$.

Due to the convexity of $\mathcal{L}(\cdot)$ with respect to $q(\cdot,\cdot|x)$, a necessary and sufficient condition for $q^*(\cdot,\cdot|z)$ to be the optimal minimizer is when $\frac{\partial \mathcal{L}(\{s_i\}_{i=1}^2, \lambda(x), \mu(x,\widehat{z},\widehat{x}))}{\partial q(\widehat{z},\widehat{x}|z))} = 0$ when $q^*(\cdot,\cdot|z) > 0$ and $\frac{\partial \mathcal{L}(\{s_i\}_{i=1}^2, \lambda(x), \mu(x,\widehat{z},\widehat{x}))}{\partial q(\widehat{z},\widehat{x}|z))} \leq 0$ when $q^*(\cdot,\cdot|z) = 0$, $\forall (\widehat{z},\widehat{x}) \in \widehat{\mathcal{Z}} \times \widehat{\mathcal{X}}$. Since there is nothing to prove for the latter case, we focus on the former case, in which the derivative of the fully unconstrained problem (48) is

$$\sum_z p(z)\left[\log\left(\frac{q^*(\widehat{z},\widehat{x}|z)}{\nu^*(\widehat{z},\widehat{x})}\right) - s_1\widehat{d}_s(z,\widehat{x}) - s_2 d_o(z,\widehat{z}) + \lambda^*(z)\right] = 0, \tag{49}$$

where we took $\mu(z,\widehat{z},\widehat{x}) = \mu^*(z,\widehat{z},\widehat{x}) = 0$ $\forall (z,\widehat{z},\widehat{x}) \in \mathcal{Z} \times \widehat{\mathcal{Z}} \times \widehat{\mathcal{X}}$. Moreover, in (49) we have that $\lambda(z) = \lambda^*(z) > 0$, $\forall z \in \mathcal{Z}$ because we require $\sum_{\widehat{z},\widehat{x}} q^*(\widehat{z},\widehat{x}|z) = 1$. Applying this result in (49) and solving with respect to $q^*(\cdot,\cdot|z)$ we obtain

$$q^*(\widehat{z},\widehat{x}|z) = e^{s_1\widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z}) - \lambda(z)} \nu^*(\widehat{z},\widehat{x}). \tag{50}$$

Since $\sum_{\widehat{z},\widehat{x}} q^*(\widehat{z},\widehat{x}|z) = 1$, we average both sides with respect to $(\widehat{z},\widehat{x}) \in \widehat{\mathcal{Z}} \times \widehat{\mathcal{X}}$ and solve to obtain $\lambda^*(z) > 0$, which is given by

$$\lambda^*(z) = \log\left(\sum_{\widehat{z},\widehat{x}} e^{s_1\widehat{d}_s(z,\widehat{x}) + s_2 d_o(z,\widehat{z})} \nu^*(\widehat{z},\widehat{x})\right). \tag{51}$$

Substituting (51) in (50), we obtain the implicit expression of (16) for $s_1 \leq 0, s_2 \leq 0$. Moreover, substituting (16) in (48) we obtain (17), provided that $R(D_o^*, D_s^*) > 0$, hence we obtain **(i)**. Clearly, the cases in **(ii)-(iv)** follow as special cases of case **(i)**. This completes the proof.

Appendix B
Proof of Theorem 2

Recall that the input data and the distortion functions are introduced in Problem 1. We first start with some preliminary calculations. In particular, using (22), we can obtain $p(z)$ as follows:

$$p(z) = \sum_{x\in\{0,1\}} p(z|x)p(x),$$

which gives

$$p(z) = \begin{pmatrix} p(z=0) \\ p(z=1) \end{pmatrix} = \begin{pmatrix} \alpha\beta + (1-\alpha)\gamma \\ \alpha(1-\beta) + (1-\alpha)(1-\gamma) \end{pmatrix}. \tag{52}$$

Using the fact that $p(z,x) = p(z|x)p(x)$ we obtain

$$p(z,x) = \begin{pmatrix} p(z=0,x=0) \\ p(z=0,x=1) \\ p(z=1,x=0) \\ p(z=1,x=1) \end{pmatrix} = \begin{pmatrix} \alpha\beta \\ \gamma(1-\alpha) \\ \alpha(1-\beta) \\ (1-\alpha)(1-\gamma) \end{pmatrix}. \tag{53}$$

Moreover, from (52), (53) and the fact that $\widehat{d}_s(z,\widehat{x}) = \sum_{x\in\mathcal{X}} p(x|z)d_s(x,\widehat{x})$ (from the characterization in Lemma 1), we can obtain $\widehat{d}_s(z,\widehat{x})$ as follows

$$\widehat{d}_s(z,\widehat{x}) = \begin{pmatrix} \frac{p(z=0,x=1)}{p(z=0)} & \frac{p(z=0,x=0)}{p(z=0)} \\ \frac{p(z=1,x=1)}{p(z=1)} & \frac{p(z=1,x=0)}{p(z=1)} \end{pmatrix}. \tag{54}$$

We can now proceed to prove **(i)**.

(i) We will prove this by finding the analytical solutions of $\{\nu^*(\widehat{z},\widehat{x}), (\widehat{z},\widehat{x}) \in \{0,1\}\times\{0,1\}\}$ and $\{q^*(\widehat{z},\widehat{x}|z), (z,\widehat{z},\widehat{x}) \in \{0,1\}\times\{0,1\}\times\{0,1\}\}$. We first recall, from Theorem 1, that

$$\nu^*(\widehat{z},\widehat{x}) = \sum_{z\in\{0,1\}} q^*(\widehat{z},\widehat{x}|z)p(z), \tag{55}$$

where $\{q^*(\widehat{z},\widehat{x}|z), (z,\widehat{z},\widehat{x}) \in \{0,1\}\times\{0,1\}\times\{0,1\}\}$ can be computed via (16).

Writing down (55) results into four third degree polynomial equations with four unknowns and one extra equation that ensures $\{\nu^*(\widehat{z},\widehat{x}), (\widehat{z},\widehat{x}) \in \{0,1\}\times\{0,1\}\}$ is a column vector that sums up to one. It can be shown that the system of non-linear equations has the following trivial solutions

$$\nu^*(\widehat{z},\widehat{x}) = \begin{pmatrix} p^*(\widehat{z}=0,\widehat{x}=0) \\ p^*(\widehat{z}=0,\widehat{x}=1) \\ p^*(\widehat{z}=1,\widehat{x}=0) \\ p^*(\widehat{z}=1,\widehat{x}=1) \end{pmatrix} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}, \tag{56}$$

which can be excluded from the general solution because they lead to zero rates, whereas the non-trivial solutions are as follows

$$\nu^*(\widehat{z},\widehat{x}) = \left\{ \begin{pmatrix} \frac{d_1(h_1-e_1)-p(z=0)(a_1h_1-d_1e_1)}{(a_1-d_1)(e_1-h_1)} \\ 0 \\ 0 \\ \frac{p(z=0)(a_1h_1-d_1e_1)-a_1(h_1-e_1)}{(a_1-d_1)(e_1-h_1)} \end{pmatrix}, \quad \begin{pmatrix} \frac{b_1(f_1-e_1)-p(z=0)(a_1f_1-b_1e_1)}{(a_1-b_1)(e_1-f_1)} \\ \frac{p(z=0)(a_1f_1-b_1e_1)-a_1(f_1-e_1)}{(a_1-b_1)(e_1-f_1)} \\ 0 \\ 0 \end{pmatrix}, \right. \tag{57}$$

$$\begin{pmatrix} \frac{c_1(g_1-e_1)-p(z=0)(a_1g_1-c_1e_1)}{(a_1-c_1)(e_1-g_1)} \\ 0 \\ \frac{p(z=0)(a_1g_1-c_1e_1)-a_1(g_1-e_1)}{(a_1-b_1)(e_1-f_1)} \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \frac{c_1(g_1-f_1)-p(z=0)(b_1g_1-c_1f_1)}{(b_1-c_1)(f_1-g_1)} \\ \frac{p(z=0)(b_1g_1-c_1f_1)-b_1(g_1-f_1)}{(b_1-c_1)(f_1-g_1)} \\ 0 \end{pmatrix}, \tag{58}$$

$$\left. \begin{pmatrix} 0 \\ 0 \\ \frac{d_1(h_1-g_1)-p(z=0)(c_1h_1-d_1g_1)}{(c_1-d_1)(g_1-h_1)} \\ \frac{p(z=0)(c_1h_1-d_1g_1)-c_1(h_1-g_1)}{(c_1-d_1)(g_1-h_1)} \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \frac{d_1(h_1-f_1)-p(z=0)(b_1h_1-d_1f_1)}{(b_1-d_1)(f_1-h_1)} \\ 0 \\ \frac{p(z=0)(b_1h_1-d_1f_1)-b_1(h_1-g_1)}{(b_1-d_1)(f_1-h_1)} \end{pmatrix} \right\}, \tag{59}$$

where $a_1 = e^{s_1\widehat{d}_s(0,0)}$, $b_1 = e^{s_1\widehat{d}_s(0,1)}$, $c_1 = e^{s_1\widehat{d}_S(0,0)+s_2}$, $d_1 = e^{s_1\widehat{d}_s(0,1)+s_2}$, $e_1 = e^{s_1\widehat{d}_s(1,0)+s_2}$, $f_1 = e^{s_1\widehat{d}_s(1,1)+s_2}$, $g_1 = e^{s_1\widehat{d}_s(1,0)}$, $h_1 = e^{s_1\widehat{d}_s(1,1)}$. Since we found analytically the marginal on the $\mathrm{i.i.d}$ output process in (57)-(59), we subsequently proceed to find the corresponding explicit expressions of the optimal minimizer for each of the explicit expressions in (57)-(59). Herein, we only give the explicit solution of the optimal minimizer $q^*(\widehat{z},\widehat{x}|z)$ that corresponds to the left hand side (LHS) solution of (57); the other cases, e.g., RHS matrix of (57), (58), (59), follow by simply substituting in (16). Substituting the LHS matrix of (57) in (16) we obtain

$$q^*(\widehat{z},\widehat{x}|z) = \begin{pmatrix} q(\widehat{z}=0,\widehat{x}=0|z=0) & q(\widehat{z}=0,\widehat{x}=0|z=1) \\ q(\widehat{z}=0,\widehat{x}=1|z=0) & q(\widehat{z}=0,\widehat{x}=1|z=1) \\ q(\widehat{z}=1,\widehat{x}=0|z=0) & q(\widehat{z}=1,\widehat{x}=0|z=1) \\ q(\widehat{z}=1,\widehat{x}=1|z=0) & q(\widehat{z}=1,\widehat{x}=1|z=1) \end{pmatrix}, \tag{60}$$

where

$$
\begin{aligned}
q^*(\widehat{z}=0,\widehat{x}=0|z=0) &= \frac{a_1d_1(h_1-e_1)+a_1p(z=0)(a_1h_1-d_1e_1)}{p(z=0)(a_1h_1-d_1e_1)(d_1-a_1)}\\
q^*(\widehat{z}=0,\widehat{x}=1|z=0) &= p(\widehat{z}=1,\widehat{x}=0|x=0)=0\\
q^*(\widehat{z}=1,\widehat{x}=1|z=0) &= \frac{a_1p(z=0)(a_1h_1-d_1e_1)-a_1d_1(h_1-e_1)}{p(z=0)(a_1h_1-d_1e_1)(d_1-a_1)}\\
q^*(\widehat{z}=1,\widehat{x}=0|z=0) &= p(\widehat{z}=1,\widehat{x}=0|x=1)=0\\
q^*(\widehat{z}=0,\widehat{x}=0|z=1) &= \frac{d_1e_1(h_1-e_1)+e_1p(z=0)(a_1h_1-d_1e_1)}{p(z=1)(a_1h_1-d_1e_1)(e_1-h_1)}\\
q^*(\widehat{z}=0,\widehat{x}=1|z=1) &= p(\widehat{z}=1,\widehat{x}=0|x=1)=0\\
q^*(\widehat{z}=1,\widehat{x}=1|z=1) &= \frac{h_1p(z=0)(a_1h_1-d_1e_1)-a_1h_1(h_1-e_1)}{p(z=1)(a_1h_1-d_1e_1)(e_1-h_1)}\\
q^*(\widehat{z}=1,\widehat{x}=0|z=1) &= p(\widehat{z}=1,\widehat{x}=1|x=1)=0.
\end{aligned}
$$

The general explicit solution of the optimal minimizer allows to find the information structure of $\{q^*(\widehat{z}|\widehat{x},z),(z,\widehat{z},\widehat{x})\in\{0,1\}\times\{0,1\}\times\{0,1\}\}$ and $\{q^*(\widehat{x}|\widehat{z},z),(z,\widehat{z},\widehat{x})\in\{0,1\}\times\{0,1\}\times\{0,1\}\}$, respectively, via the following expressions

$$
q^*(\widehat{z}|\widehat{x},z)=\frac{q^*(\widehat{z},\widehat{x}|z)p(z)}{\sum_{\widehat{z}\in\{0,1\}}q^*(\widehat{z},\widehat{x}|z)p(z)},\quad q^*(\widehat{x}|\widehat{z},z)=\frac{q^*(\widehat{z},\widehat{x}|z)p(z)}{\sum_{\widehat{x}\in\{0,1\}}q^*(\widehat{z},\widehat{x}|z)p(z)}. \tag{61}
$$

Surprisingly, the closed form expressions in (61), admit the following simple structure

$$
q^*(\widehat{z}|\widehat{x},z)=\begin{pmatrix}1&1&0&0\\0&0&1&1\end{pmatrix},\quad q^*(\widehat{x}|\widehat{z},z)=\begin{pmatrix}1&1&0&0\\0&0&1&1\end{pmatrix}, \tag{62}
$$

where for the left hand side (LHS) expression of (62) we have

$$
\begin{aligned}
q^*(\widehat{z}=0|\widehat{x}=0,z=0)&=q^*(\widehat{z}=0|\widehat{x}=0,z=1)=1\\
q^*(\widehat{z}=0|\widehat{x}=1,z=0)&=q^*(\widehat{z}=0|\widehat{x}=1,z=1)=0\\
q^*(\widehat{z}=1|\widehat{x}=0,z=0)&=q^*(\widehat{z}=1|\widehat{x}=0,z=1)=0\\
q^*(\widehat{z}=1|\widehat{x}=1,z=0)&=q^*(\widehat{z}=1|\widehat{x}=1,z=1)=1,
\end{aligned} \tag{63}
$$

and similarly, for the RHS expression in (62) each component represents the corresponding optimal minimizer given in (63) with $\widehat{z}$ and $\widehat{x}$, interchanged. The identical information structure in both expressions of (62), reveals the conditional independence $q^*(\widehat{z}|\widehat{x},z)=q^*(\widehat{z}|\widehat{x})$ and $q^*(\widehat{x}|\widehat{z},z)=q^*(\widehat{x}|\widehat{z})$ or, equivalently, the fact that $\widehat{\mathbf{z}}-\widehat{\mathbf{x}}-\mathbf{z}$ and $\widehat{\mathbf{x}}-\widehat{\mathbf{z}}-\mathbf{z}$. We note that the previous information structure of $\{q^*(\widehat{z}|\widehat{x},z),(z,\widehat{z},\widehat{x})\in\{0,1\}\times\{0,1\}\times\{0,1\}\}$ and

$\{q^*(\widehat{x}|\widehat{z},z),(z,\widehat{z},\widehat{x})\in\{0,1\}\times\{0,1\}\times\{0,1\}\}$ can be observed if we pick any $\{\nu^*(\widehat{z},\widehat{x}),(\widehat{z},\widehat{x})\in\{0,1\}\times\{0,1\}\}$ from (57)-(59) hence we omit the re-derivation. This proves **(i)**.

(ii) Since we proved **(i)**, then, from Lemma 2, **(2)**, the use of the Lagrange duality theorem [33] guarantees that the solution is the one for which the Lagrangian, i.e., s_1 or s_2 in our case, yields the greater rates hence $R(D_o^*,D_s^*)=\max\{R(D_o^*),R(D_s^*))\}$. This completes the proof.

APPENDIX C
PROOF OF THEOREM 3

We start the proof by noting that $\{p(z),\ p(z|x):\ (x,z)\in\{0,1\}\times\{0,1\}\}$ are given by (52) and (53), respectively. Moreover, from (27), we obtain

$$\widehat{d}_s(z,\widehat{x})=\begin{pmatrix}\widehat{d}_s(z=0,\widehat{x}=0) & \widehat{d}_s(z=0,\widehat{x}=e) & \widehat{d}_s(z=0,\widehat{x}=1)\\ \widehat{d}_s(z=1,\widehat{x}=0) & \widehat{d}_s(z=1,\widehat{x}=e) & \widehat{d}_s(z=1,\widehat{x}=1)\end{pmatrix}=\begin{pmatrix}\infty & 1 & \infty\\ \infty & 1 & \infty\end{pmatrix}. \tag{64}$$

Using (64) and the input data, we obtain that the optimal minimizer has the following structure

$$q^*(\widehat{z},\widehat{x}|z)=\begin{pmatrix}q^*(\widehat{z}=0,\widehat{x}=0|z=0) & q^*(\widehat{z}=0,\widehat{x}=0|z=1)\\ q^*(\widehat{z}=0,\widehat{x}=e|z=0) & q^*(\widehat{z}=0,\widehat{x}=e|z=1)\\ q^*(\widehat{z}=0,\widehat{x}=1|z=0) & q^*(\widehat{z}=0,\widehat{x}=1|z=1)\\ q^*(\widehat{z}=1,\widehat{x}=0|z=0) & q^*(\widehat{z}=1,\widehat{x}=0|z=1)\\ q^*(\widehat{z}=1,\widehat{x}=e|z=0) & q^*(\widehat{z}=1,\widehat{x}=e|z=1)\\ q^*(\widehat{z}=1,\widehat{x}=1|z=0) & q^*(\widehat{z}=1,\widehat{x}=1|z=1)\end{pmatrix}$$

$$=\begin{pmatrix}0 & 0\\ \frac{\nu^*(\widehat{z}=0,\widehat{x}=e)}{\nu^*(\widehat{z}=0,\widehat{x}=e)+e^{s_2}\nu^*(\widehat{z}=1,\widehat{x}=e)} & \frac{e^{s_2}\nu^*(\widehat{z}=0,\widehat{x}=e)}{e^{s_2}\nu^*(\widehat{z}=0,\widehat{x}=e)+\nu^*(\widehat{z}=1,\widehat{x}=e)}\\ 0 & 0\\ 0 & 0\\ \frac{e^{s_2}\nu^*(\widehat{z}=1,\widehat{x}=e)}{\nu^*(\widehat{z}=0,\widehat{x}=e)+e^{s_2}\nu^*(\widehat{z}=1,\widehat{x}=e)} & \frac{\nu^*(\widehat{z}=1,\widehat{x}=e)}{e^{s_2}\nu^*(\widehat{z}=0,\widehat{x}=e)+\nu^*(\widehat{z}=1,\widehat{x}=e)}\\ 0 & 0\end{pmatrix}, \tag{65}$$

where each of $\nu^*(\widehat{z}=0,\widehat{x}=e)$ and $\nu^*(\widehat{z}=1,\widehat{x}=e)$ can take the trivial solutions 0 or 1 and a non-trivial solution of the form

$$\nu^*(\widehat{z}=0,\widehat{x}=e)=\frac{p(z=0)(1+e^{s_2})-e^{s_2}}{1-e^{s_2}},\quad \nu^*(\widehat{z}=1,\widehat{x}=e)=\frac{1-p(z=0)(1+e^{s_2})}{1-e^{s_2}}. \tag{66}$$

The explicit structure of (65) reveals that the optimal solution to this problem is parametrized only by the Lagrangian multiplier $s_2<0$, which further means that this solution should be

$R(D_o^*, D_s^*) = R(D_o^*)$ (from Theorem 1). The latter implies that the Markov chain $\mathbf{z} - \widehat{\mathbf{z}} - \widehat{\mathbf{x}}$ holds. A way to compute analytically $R(D_o^*, D_s^*)$ is the following. Find $s_2^* < 0$ from (18) using the explicit expressions (65) and (66). This will give $s_2^* = \log\left(\frac{D_o^*}{1-D_o^*}\right)$. Then, by substituting all the pieces together in $R(D_o^*, D_s^*) = \mathbf{E}\left\{\frac{q^*(\widehat{\mathbf{z}},\widehat{\mathbf{x}}|\mathbf{z})}{p^*(\widehat{\mathbf{z}},\widehat{\mathbf{x}})}\right\}$ we obtain (23). This completes the proof.

Appendix D
Proof of Theorem 4

Let $V(\boldsymbol{\nu}) \triangleq \mathbb{I}(\mathbf{p}; \mathbf{q}(\boldsymbol{\nu}^{(k)})) - s_1 D_s(\boldsymbol{\nu}) - s_2 D_o(\boldsymbol{\nu})$ with $D_s(\boldsymbol{\nu}) = \sum_{z,\widehat{z},\widehat{x}} \widehat{d}_s(z,\widehat{x})\mathbf{q}(\boldsymbol{\nu})p(z)$ and $D_o(\boldsymbol{\nu}) = \sum_{z,\widehat{z},\widehat{x}} d_o(z,\widehat{z})\mathbf{q}(\boldsymbol{\nu})p(z)$. To show that $V(\boldsymbol{\nu})$ is nonincreasing, we can use Lemma 3, **(2)**, **(3)**. Minimizing over $\mathbf{q}$ gives a value $F(\boldsymbol{\nu}^{(k)})$ between $V(\boldsymbol{\nu}^{(k+1)})$ and $V(\boldsymbol{\nu}^{(k)})$, namely,

$$F(\boldsymbol{\nu}^{(k)}) = -\sum_{z} p(z)\left(\log \sum_{\widehat{z},\widehat{x}} \nu^{(k-1)}(\widehat{z},\widehat{x})A(z,\widehat{z},\widehat{x})\right), \tag{67}$$

and due to the successive minimizations, we have the nonincreasing sequence $\ldots \geq V(\boldsymbol{\nu}^{(k)}) \geq F(\boldsymbol{\nu}^{(k)}) \geq V(\boldsymbol{\nu}^{(k+1)}) \geq F(\boldsymbol{\nu}^{(k+1)}) \geq \ldots$. Combining the above two operations, we obtain a recursive definition of $\boldsymbol{\nu}^{(k)}$ of the theorem which by construction has positive elements and $V(\boldsymbol{\nu}^k)$ is nonincreasing. Since $V(\boldsymbol{\nu}^{(k)})$ is bounded, it must converge to some constant number V^{∞}. It remains to show that $V^{\infty} = R(D_o, D_s) - s_1 D_s - s_2 D_o$.

Let $\mathbf{q}^{(k+1)} = \mathbf{q}(\boldsymbol{\nu}^{(k)})$. Let ν^* be such that $V(\boldsymbol{\nu}^*) = R(D_o, D_s) - s_1 D_s - s_2 D_o$, and let $q^* = \mathbf{q}(\boldsymbol{\nu}^*)$. Now let

$$\begin{aligned}
&\sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\frac{q^{(k)}(\widehat{z},\widehat{x}|z)}{q^{(k+1)}(\widehat{z},\widehat{x}|z)}\right) \overset{(a)}{=} \sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\frac{q^{(k)}(\widehat{z},\widehat{x}|z)}{\nu^{(k)}(\widehat{z},\widehat{x})}\right) \\
&- \sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log A(z,\widehat{z},\widehat{x}) + \sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\sum_{\widehat{z},\widehat{x}} \nu^{(k)}(\widehat{z},\widehat{x})A(z,\widehat{z},\widehat{x})\right),
\end{aligned} \tag{68}$$

where (a) follows by the fact that we let $\mathbf{q}^{(k+1)} = \mathbf{q}(\boldsymbol{\nu}^{(k)})$. Next, we show that the first RHS in (68) is increased if we replace $\mathbf{q}^{(k)}$ and $\boldsymbol{\nu}^{(k)}$ are replaced by q^* and $\boldsymbol{\nu}^*$, respectively.

$$\begin{aligned}
&\sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\frac{q^{(k)}(\widehat{z},\widehat{x}|z)}{\nu^{(k)}(\widehat{z},\widehat{x})}\right) - \sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\frac{q^*(\widehat{z},\widehat{x}|z)}{\nu^*(\widehat{z},\widehat{x})}\right) \\
&= \sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\frac{q^{(k)}(\widehat{z},\widehat{x}|z)q^*(\widehat{z},\widehat{x}|z)}{\nu^{(k)}(\widehat{z},\widehat{x})\nu^*(\widehat{z},\widehat{x})}\right) \\
&\overset{(b)}{\leq} \sum_{z,\widehat{z},\widehat{x}} p(z)q^{(k)}(\widehat{z},\widehat{x}|z)\left(\frac{\nu^*(\widehat{z},\widehat{x})}{\nu^{(k)}(\widehat{z},\widehat{x})}\right) - \sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z) \overset{(c)}{=} 0,
\end{aligned} \tag{69}$$

where (b) follows from the inequality $\log(x) \leq x-1$ for $x>0$; (c) because both terms after the inequality average to 1. As a result, we can bound from above the LHS of (68) as follows

$$\begin{aligned}
&\sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\frac{q^{(k)}(\widehat{z},\widehat{x}|z)}{q^{(k+1)}(\widehat{z},\widehat{x}|z)}\right) \overset{(c)}{\leq} \sum_{z,\widehat{z},\widehat{x}} p(z)q^*(\widehat{z},\widehat{x}|z)\log\left(\frac{q^*(\widehat{z},\widehat{x}|z)}{\sum_z q^*(\widehat{z},\widehat{x}|z)p(z)}\right)\\
&- s_1\sum_{z,\widehat{z},\widehat{x}} \widehat{d}_s(z,\widehat{x})q^*(\widehat{z},\widehat{x}|z)p(z) - s_2\sum_{z,\widehat{z},\widehat{x}} d_o(z,\widehat{z})q^*(\widehat{z},\widehat{x}|z)p(z)\\
&+\sum_z p(z)\left(\log\sum_{\widehat{z},\widehat{x}}\nu^{(k)}(\widehat{z},\widehat{x})A(z,\widehat{z},\widehat{x})\right) \overset{(d)}{=} V(\boldsymbol{\nu}^*) - F(\boldsymbol{\nu}^{(k)}),
\end{aligned} \tag{70}$$

where (c) follows from (69) and the identify $q^*(\widehat{z},\widehat{x}) = \sum_z q^*(\widehat{z},\widehat{x}|z)p(z)$; (d) follows by definition of $V(\boldsymbol{\nu}^*)$ and (67).

Now by summing on k until some finite K, we obtain

$$\begin{aligned}
&\sum_{k=1}^K\left[V(\boldsymbol{\nu}^*) - F(\boldsymbol{\nu}^{(k)})\right] \overset{(e)}{\geq} \sum_{z,\widehat{z},\widehat{x}} q^*(\widehat{z},\widehat{x}|z)p(z)\sum_{k=1}^K\log\left(\frac{q^{(k)}(\widehat{z},\widehat{x}|z)}{q^{(k+1)}(\widehat{z},\widehat{x}|z)}\right)\\
&\overset{(f)}{=}\sum_{z,\widehat{z},\widehat{x}} q^*(\widehat{z},\widehat{x}|z)p(z)\log\left(\frac{q^{(1)}(\widehat{z},\widehat{x}|z)}{q^{(K+1)}(\widehat{z},\widehat{x}|z)}\right) \overset{(g)}{\geq} \sum_{z,\widehat{z},\widehat{x}} q^*(\widehat{z},\widehat{x}|z)p(z)\log\left(\frac{q^{(1)}(\widehat{z},\widehat{x}|z)}{q^*(\widehat{z},\widehat{x}|z)}\right)
\end{aligned}, \tag{71}$$

where (e) follows from (70); (f) follows from the telescopic sum of the logarithmic ratio; (g) follows from Kullback's discrimination inequality [40, Theorem 1], i.e., the information is a measurement that cannot be increased by subsequent processing.

Finally, with a change of sign in the LHS of (71) we obtain the inequality

$$\sum_{k=1}^K\left[F(\boldsymbol{\nu}^{(k)}) - V(\boldsymbol{\nu}^*)\right] \leq T(\boldsymbol{\nu}^*,\boldsymbol{\nu}^{(1)}). \tag{72}$$

The RHS in (72) is a constant independent of K and the LHS is smaller than this constant for any K. Since $F(\boldsymbol{\nu}^{(k)})$ is a larger value than $V(\boldsymbol{\nu}^*)$ and is nonincreasing, then, the term in the brackets is positive and must vanish as $K\to\infty$. This further means that $F(\boldsymbol{\nu}^{(k)}) \overset{K\to\infty}{\longrightarrow} V(\boldsymbol{\nu}^*)$, which completes the proof.

Appendix E
Proof of Theorem 5

We first prove (38). Recall that

$$q(\widehat{z},\widehat{x}|z) = \frac{A(z,\widehat{z},\widehat{x})\nu(\widehat{z},\widehat{x})}{\sum_{\widehat{z},\widehat{x}} A(z,\widehat{z},\widehat{x})\nu(\widehat{z},\widehat{x})}, \tag{73}$$

is a transition matrix that gives D_s and D_o. Then, the following inequality holds

$$\begin{aligned}R(D_o, D_s) &\overset{(a)}{\leq} \mathbb{I}(\mathbf{p}, \mathbf{q}) = \sum_{z,\widehat{z},\widehat{x}} \log\left(\frac{q(\widehat{z},\widehat{x}|z)}{\sum_z q(\widehat{z},\widehat{x}|z)p(z)}\right) q(\widehat{z},\widehat{x}|z)p(z)\\ &\overset{(b)}{=} \sum_{z,\widehat{z},\widehat{x}} \log\left(\frac{A(z,\widehat{z},\widehat{x})\nu(\widehat{z},\widehat{x})}{\left(\sum_{\widehat{z},\widehat{x}} A(z,\widehat{z},\widehat{x})\nu(\widehat{z},\widehat{x})\right)\left(\sum_z q(\widehat{z},\widehat{x}|z)p(z)\right)}\right) q(\widehat{z},\widehat{x}|z)p(z)\\ &\overset{(c)}{=} \text{RHS of (38)},\end{aligned} \tag{74}$$

where (a) follows by definition of (6); (b) follows from (73); (c) follows after some algebra and substituting (36), (37) and (73).

Next, we prove (39). Using Lemma 4, we have the bound $R(D_o, D_s) \geq s_1 D_s + s_2 D_o + \sum_{z\in\mathcal{Z}} p(z) \log \lambda(z)$ where $\boldsymbol{\lambda}$ is any vector such that $\sum_z p(z)\lambda(z)A(z,\widehat{z},\widehat{x}) \leq 1$. To prove the lower bound, we let

$$c_{\max}(\widehat{z},\widehat{x}) = \max_{\widehat{z},\widehat{x}} c(\widehat{z},\widehat{x}), \text{ and } \lambda(z) = \left(c_{\max}(\widehat{z},\widehat{x}) \sum_{\widehat{z},\widehat{x}} A(z,\widehat{z},\widehat{x})\nu(\widehat{z},\widehat{x})\right)^{-1}. \tag{75}$$

Applying the quantities of (75) in $\sum_z p(z)\lambda(z)A(z,\widehat{z},\widehat{x})$ we can observe that it is always less than one, therefore the lower bound obtain from Lemma 4 holds. Substituting the values of (75) in that bound we obtain the RHS of (39). This completes the proof.

REFERENCES

[1] P. A. Stavrou and M. Kountouris, "A rate distortion approach to goal-oriented communication," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2022, pp. 778–783.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Labs Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[3] C. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Conv. Rec.*, pp. 142–163, 1993.

[4] Y. Bar-Hillel and R. Carnap, "Semantic information," *British Journal for the Philosophy of Science*, vol. 4, no. 14, 1953.

[5] F. I. Dretske, *Knowledge and the flow of information*, 1st ed. MIT Press Cambridge, Mass., 1981.

[6] L. Floridi, "Semantic Conceptions of Information," in *The Stanford Encyclopedia of Philosophy*, Winter 2005.

[7] J. Bao *et al.*, "Towards a theory of semantic communication," in *IEEE Network Science Workshop*, 2011.

[8] O. Goldreich, B. Juba, and M. Sudan, "A theory of goal-oriented communication," *J. ACM*, vol. 59, no. 2, 2012.

[9] T. Okamoto, "A unified paradigm of organized complexity and semantic information theory," *arXiv preprint arXiv: 1608.00941*, 2016.

[10] F. M. Willems and T. Kalker, "Semantic coding: Partial transmission," in *Proc. IEEE Int. Symp. Inf. Theory*, 2008, pp. 1617–1621.

[11] B. Güler, A. Yener, and A. Swami, "The semantic communication game," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, 2018.

[12] B. Sharma, J.Mitter, and M.Mohan, "On measures of useful information," *Information and Control*, no. 39, 1978.

[13] R. L. Stratonovich, "On the value of information," *Izv. USSR Acad. Sci. Tech. Cybern.*, no. 5, 1965.

[14] R. A. Howard, "Information value theory," *IEEE Transactions on Systems Science and Cybernetics*, vol. 2, no. 1, 1966.

[15] P. Popovski, O. Simeone, F. Boccardi, D. Gündüz, and O. Sahin, "Semantic-effectiveness filtering and control for post-5G wireless connectivity," *J. Indian Inst. Sci.*, vol. 100, 2020.

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

[17] Z. Qin, X. Tao, J. Lu, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv: 2201.01389v2*, 2022.

[18] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, T. Soleymani, B. Soret, and K. H. Johansson, "Semantic communications in networked systems," *https://arxiv.org/abs/2103.05391*, 2021.

[19] J. Liu, W. Zhang, and H. V. Poor, "A rate-distortion framework for characterizing semantic information," in *Proc. IEEE Int. Symp. Inf. Theory*, 2021, pp. 2894–2899.

[20] J. Liu, S. Shao, W. Zhang, and H. V. Poor, "An indirect rate-distortion characterization for semantic sources: General model and the case of Gaussian observation," *https://arxiv.org/abs/2201.12477*, 2022.

[21] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Info. Theory*, vol. 8, no. 5, pp. 293–304, Sept. 1962.

[22] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, pp. 406–411, 1970.

[23] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 518–521, Sept. 1980.

[24] A. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, 1982.

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

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., NJ, 2006.

[27] R. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 480–489, 1973.

[28] J. Stjernvall, "Dominance-a relation between distortion measures," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 798–807, 1983.

[29] J. Xiao and Z.-Q. Luo, "Compression of correlated Gaussian sources under individual distortion criteria," in *Proc. Annual Allerton Conf. on Comm., Control, and Comput.*, Jan. 2005, pp. 438–447.

[30] E. Stylianou, C. D. Charalambous, and T. Charalambous, "Joint rate distortion function of a tuple of correlated multivariate Gaussian sources with individual fidelity criteria," in *Proc. IEEE Int. Symp. Inf. Theory*, 2021, pp. 2167–2172.

[31] Y. Blau and T. Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *Proceedings of the 36th International Conference on Machine Learning*, 2019, pp. 675–685.

[32] G. Zhang, J. Qian, J. Chen, and A. J. Khisti, "Universal rate-distortion-perception representations for lossy compression," in *35th Conference on Neural Information Processing Systems*, 2021.

[33] D. G. Luenberger, *Optimization by Vector Space Methods*, 1st ed. John Wiley & Sons, Inc., 1969.

[34] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[35] A. Kipnis, S. Rini, and A. J. Goldsmith, "The indirect rate-distortion function of a binary i.i.d source," in *Proc. IEEE Inf. Theory Workshop*, 2015, pp. 352–356.

[36] R. Yeung, *Information theory and network coding*. Springer, 2008.

[37] R. E. Blahut, *Principles and Practice of Information Theory*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[38] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014.

[39] S. Ihara, *Information theory - for Continuous Systems*. World Scientific, 1993.

[40] R. Blahut, "Information bounds of the Fano-Kullback type," *IEEE Trans. Inf. Theory*, vol. 22, no. 4, pp. 410–421, 1976.