Goal-Oriented Single-Letter Codes for Lossy Joint Source-Channel Coding

Photios A. Stavrou Marios Kountouris
Communication Systems Department
EURECOM, Sophia Antipolis, France
Email: {fotios.stavrou, marios.kountouris}@eurecom.fr

***Abstract*—A new variation of the classical point-to-point joint source-channel coding (JSCC) is studied here, which is relevant for identifying goal-oriented semantic aspects of a transmitted source message over a noisy channel. For this new formulation, coined *goal-oriented* JSCC, we first introduce optimality criteria followed by necessary and sufficient conditions for global optimality. The focus of our main theoretical results is on investigating the implications of a special subclass of block codes, namely, the single-letter codes, via a theorem in which we provide necessary and sufficient conditions for the probabilistic matching of a goal-oriented source message with a noisy channel. We corroborate our theoretical results with two examples, in which goal-oriented single-letter codes and uncoded transmission perform optimally. Our results further highlight the important role of multiple fidelity constraints in goal-oriented communications.**

I. Introduction

Goal-oriented semantic communication has recently received a growing interest from various research communities spanning from information and communication theory [1], [2], game theory [3], control theory [4] to networking [5]. The main motivation behind this in the vast majority of the literature, at least in the area of information and communication theory, can be found in Shannon's seminal work [6]. A key claim therein is that the fundamental problem of communication ought to deliberately disregard the semantic aspects of a message and its impact. Although this argument is valid and relevant in general, a goal-oriented semantic theory of information transmission with tangible applications to communication systems is still an open quest.

In this work, we consider a variant of the classical point-to-point JSCC model in which a remote semantic source is indirectly observed by a transmitter via an observable message and conveyed across a noisy channel whereas at the destination, the goal is for both the remote and observable messages to be received as accurate as possible depending on some fidelity constraints. Here we furnish the semantic and observations sources with single-letter distortion constraints responsible to penalize their transmission rates. Our setup possess goal-oriented attributes and intrinsic representation of information (e.g., features, structural and qualitative properties, embedding) because the presence of fidelity constraints, dictates whether it makes sense to recover the remote or observable messages or both.

The fundamental study of JSCC dates back to [6]. Although a general definite solution to this problem appears to be elusive, the convenient result of source-channel separation principle for point-to-point stationary and ergodic systems enables for modularity between the source and channel coding schemes without any loss of optimality [6]–[8]. Unfortunately, this result does not scale to networks without loss of optimality. In practice, a tandem source-channel coding necessitates non-causal knowledge of all the message signals hence entailing large coding delays and very complex coders. Surprisingly, the tandem scheme of Shannon is not unique for the fundamental communication problem. In fact, as two counterexamples suggest in [9], [10], single-letter codes and uncoded transmission, may provide *significant gains* to the end-to-end communication system in terms of *coding delays*, *complexity* and *robustness*. For a detailed analysis on necessary and sufficient conditions that need to be satisfied for any discrete-time memoryless point-to-point communication system to be optimal we refer to [11] (and the references therein).

Contributions. The goal-oriented lossy JSCC setup proposed herein is new and can be seen as a non-trivial generalization of the goal-oriented lossy source coding setup proposed in [12], [13]. Compared to that setup, communication takes place here over noisy instead of noiseless channels. We also define a variant of source-channel codes, which we refer to as *goal-oriented codes*. Additionally, we derive necessary and sufficient conditions for global optimality of the specific goal-oriented JSCC scheme (Lemma 2). More importantly, we study the criteria needed for the optimality of goal-oriented single-letter codes using a known lower bound on the so called semantic rate distortion function (SRDF) (Lemma 1). Provided that SRDF is a tight bound, we derive necessary and sufficient conditions for optimality of our JSCC scheme (Theorem 1, Corollary 1). Our theoretical findings are corroborated with two application examples, in which single-letter codes and uncoded transmission perform optimally (Section V). These examples are extensions to the celebrated ones introduced for binary sources and binary symmetric channels (BSC) [10] and Gaussian i.i.d sources and additive white Gaussian noise (AWGN) channels [9]. The key takeaway is that in goal-oriented communications, selecting the type of individual distortion measures according to the application/task requirements can significantly affect the choice of the recovered message over noisy channels.

II. Problem Statement

We consider a memoryless source described by the tuple $(\mathbf{x}, \mathbf{z})$ with probability distribution $p(x, z)$ in the product space $\mathcal{X} \times \mathcal{Z}$. The semantic or intrinsic information of the source is in $\mathbf{x}$, which can only be indirectly observed, whereas $\mathbf{z}$ is the noisy observation of the source at the encoder side. We also consider a memoryless noisy channel described by the conditional probability distribution $p(b|a)$ where $(\mathbf{a}, \mathbf{b}) \in \mathcal{A} \times \mathcal{B}$.

Formally, the generic source-channel communication model is illustrated in Fig. 1. According to this setup, we can model the *information source* as a sequence of n-length i.i.d random variables $(\mathbf{x}^n, \mathbf{z}^n)$ and we assume that we know $p(x)$ and the transition probability distribution $p(z|x)$. The *noisy channel* is assumed to be given, with known capacity and capacity achieving distribution $p(a^m)$. Since the source and channel have different cardinality to their respective alphabets, this means that the sequence of m-length random variables $(\mathbf{a}^m, \mathbf{b}^m)$ and the reproduction sequence of n-length random variables $(\widehat{\mathbf{x}}^n, \widehat{\mathbf{z}}^n)$ need not be i.i.d, in general.

Definition 1. *(Goal-oriented block source-channel codes) The source-channel encoder (E) and decoder (D), are modeled by the mappings*

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

where (g_o^D, g_s^D) denote the observations and the semantic information decoder, respectively. The coding rate κ is the number of source symbols that have to be transmitted per channel use and is given by $\kappa = \frac{n}{m}$.

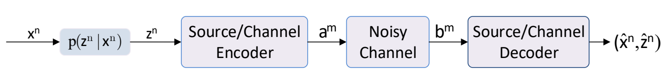

Fig. 1. The proposed goal-oriented source-channel communication model.

We consider two per-letter distortion measures responsible to penalize the semantic and observation information source 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, a cost to possibly penalize the channel given by $c : \mathcal{A} \mapsto [0, \infty)$, and their corresponding average per-letter distortions and cost by

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

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

$$c^m(a^m) = \frac{1}{m} \sum_{t=1}^{m} c(a_i). \tag{4}$$

The average distortion for the semantic information and the observations is defined as

$$\Delta_s \triangleq \mathbf{E}\left\{d_s^n(\mathbf{x}^n, \widehat{\mathbf{x}}^n)\right\}, \quad \Delta_o \triangleq \mathbf{E}\left\{d_o^n(\mathbf{z}^n, \widehat{\mathbf{z}}^n)\right\} \tag{5}$$

while the average cost for the noisy channel is defined as

$$\Gamma \triangleq \mathbf{E}\left\{c^m(\mathbf{a}^m)\right\}. \tag{6}$$

The main objecive of this work is to study necessary and sufficient conditions for the optimality of the goal-oriented JSCC setup of Fig. 1.

III. Optimality conditions of the setup in Fig. 1

To provide the main result of this section, we need to recall some results from [13] on goal-oriented compression, and on constrained/unconstrained channel capacity [7].

We start with the definition of the SRDF.

Definition 2. *(SRDF) [13, Lemma 1] For a given semantic/observation source $(p(x), p(z|x), d_s, d_o)$, the ergodic SRDF is characterized as follows*[1]

$$R(D_o, D_s) = \min_{\substack{q(\widehat{z}, \widehat{x}|z) \\ \mathbf{E}\left[\widehat{d}_s(\mathbf{z}, \widehat{\mathbf{x}})\right] \leq D_s \\ \mathbf{E}\left[d_o(\mathbf{z}, \widehat{\mathbf{z}})\right] \leq D_o}} I(\mathbf{z}; \widehat{\mathbf{z}}, \widehat{\mathbf{x}}) \tag{7}$$

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

We note that (7) has similar functional and topological properties as in classical rate distortion theory [14], e.g., monotonicity, continuity, convexity etc.

One interesting result that stems from (7) is the following lower bound.

Lemma 1. *(Lower Bound) [13, Lemma 2] The characterization in* (7) *admits the following lower bound:*

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

where $(R(D_s), R(D_o))$ represents the standard direct [14] and indirect RDFs [14], [15] subject to their individual distortion criteria, i.e.,

$$R(D_o) = \min_{\substack{q(\widehat{z}|z) \\ \mathbf{E}[d_o(\mathbf{z}, \widehat{\mathbf{z}})] \leq D_o}} I(\mathbf{z}; \widehat{\mathbf{z}}), \tag{9}$$

$$R(D_s) = \min_{\substack{q(\widehat{x}|z) \\ \mathbf{E}[\widehat{d}_s(\mathbf{z}, \widehat{\mathbf{x}})] \leq D_s}} I(\mathbf{z}; \widehat{\mathbf{x}}). \tag{10}$$

Moreover, $R^L(D_o, D_s)$ is tight if and only if (iff)

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

are concurrently satisfied.

Next, we define the ergodic constrained and unconstrained channel capacity.

Definition 3. *(Constrained/unconstrained capacity) [7]*

[1] The constrained set in (7) is compact (for both finite or abstract alphabets) and the mutual information in (7) is lower semi-continuous on $q(\widehat{z}, \widehat{x}|z)$. As a result, from Weierstrass' extreme value theorem, the infimum is attained by some $q^*(\widehat{z}, \widehat{x}|z)$ and we can formally replace it with minimum.

(i) For a given channel $(p(b|a),c)$, the ergodic constrained capacity is characterized as follows

$$C(P)=\max_{\substack{p(b|a)\\ \mathbf{E}[c(\mathbf{a})]\leq P}} I(\mathbf{a};\mathbf{b}) \tag{12}$$

where $P\in[P^{\min},P^{\max}]=[0,\infty)$.
(ii) The special case of (12) *at which the channel $p(b|a)$ disregards the input cost $c(\cdot)$ is characterized by the ergodic unconstrained capacity as follows*

$$C=\max_{p(b|a)} I(\mathbf{a};\mathbf{b}). \tag{13}$$

We note that (13) *is also obtained by* (12) *if for $c(a)<\infty$, $a\in\mathcal{A}$, $C=\lim_{P\longrightarrow\infty}C(P)$.*

In the sequel, we give two definitions, one for the achievable average distortion(s)-cost triplet $(\Delta_s,\Delta_o,\Gamma)$ and one for the optimality of the setup in Fig. 1.

Definition 4. *(Average distortion(s)-cost triplet) For a given $(p(x),p(z|x),d_o,d_s)$ and a fixed channel $(p(b|a),c)$, we write $(\Delta_s,\Delta_o,\Gamma)$ given by* (5)-(6) *to denote the average distortion(s)-cost triplet achieved by a fixed code (f^E,g_o^D,g_s^D).*

In view of Definition 4, we will denote by (D_s,D_o,P) a general distortion(s)-cost triplet not necessarily achievable.

Definition 5. *(Optimality criterion) The transmission of a semantic/observations source $(p(x),p(z|x),d_s,d_o)$, across a noisy channel $(p(b|a),c)$ results into a source-channel code (f^E,g_o^D,g_s^D) that is optimal if the following two conditions are concurrently satisfied:*
(i) *the distortion(s) (Δ_o,Δ_s) caused by the choice of (f^E,g_o^D,g_s^D), correspond to the minimum achievable distortions (hereinafter denoted by $D_o^{\min},D_s^{\min}$) operating at an input cost Γ with the best possible source-channel code of rate κ;*
(ii) *the cost Γ obtained by (f^E,g_o^D,g_s^D), corresponds to the minimum cost required to achieve the distortion(s) (Δ_o,Δ_s) with the best possible source-channel code of rate κ.*

Clearly, Definitions 4, 5 are generalizations to those considered in [11, Definitions 4, 5].

Next, we state a lemma that identifies optimality conditions for the goal-oriented JSCC model in Fig. 1.

Lemma 2. *(Optimality conditions) For a semantic-observations information source $(p(x),p(z|x),d_o,d_s)$ and a channel $(p(b|a),c)$, the transmission using a goal-oriented source-channel code (f^E,g_o^D,g_s^D) of rate κ is optimal iff the following conditions are concurrently satisfied:*
(i) *$\kappa R(\Delta_o,\Delta_s)=C(\Gamma)$;*
(ii) *the pair (Δ_o,Δ_s) cannot be decreased without increasing $R(\Delta_o,\Delta_s)$ and Γ cannot be decreased without decreasing $C(\Gamma)$.*

Proof: We sketch the proof because it is a modified version of the separation theorem [6, Theorem 21]. From the convexity and concavity of SRDF and the channel capacity, respectively, and the data processing inequality of the setup in Fig. 1, i.e., $\mathbf{x}^n-\mathbf{z}^n-\mathbf{a}^m-\mathbf{b}^m-(\widehat{\mathbf{x}}^n,\widehat{\mathbf{z}}^n)$, we obtain a converse part of the separation theorem, i.e., $\kappa R(\Delta_o,\Delta_s)\leq C(\Gamma)$. For the direct part, suppose that there exists a triplet (f^E,g_o^D,g_s^D) such that $\kappa R(\Delta_o,\Delta_s)=C(\Gamma)-\epsilon$, for some $\epsilon>0$. Then, an immediate application of the direct part of the separation principle implies that there exists a better source-channel code $(\tilde{f}^E,\tilde{g}_o^D,\tilde{g}_s^D)$ such that $\kappa R(\Delta_o,\Delta_s)=C(\Gamma)-\tilde{\epsilon}$ for some $0<\tilde{\epsilon}<\epsilon$. Hence, the optimal distortion(s)-cost triplet $(\Delta_0,\Delta_s,\Gamma)$ implies **(i)**. This result is only necessary as there may be cases where each or both (Δ_o,Δ_s) (resp. Γ) can be reduced without increasing $R(\Delta_o,\Delta_s)$ (resp. decreasing $(C\Gamma)$). Such cases can be prohibited by condition **(ii)**. ■

Remark 1. *(On Lemma 2)* **(i)** *In Lemma 2, the necessary condition is $\kappa R(\Delta_0,\Delta_s)=C(\Gamma)$ as the second condition that prohibits each or both (Δ_0,Δ_s) (resp. Γ) to reduce their value(s) may only happen when $R(\Delta_o,\Delta_s)=0$ (resp. $C(\Gamma)=C$).* **(ii)** *It should be noted that Lemma 2 includes the case at which the bound in Lemma 1 is tight.*

IV. Goal-oriented single-letter codes

In this section, we consider single-letter codes to study the optimal performance of the goal-oriented JSCC system in Fig. 1. Our results apply once the bound in Lemma 1 is tight.

We start by introducing a special class to block source-channel codes introduced in Definition 1.

Definition 6. *(Goal-oriented single-letter source-channel codes) For the system model in Fig. 1, a goal-oriented single-letter source-channel code (f^E,g_o^D,g_s^D) is specified by an encoding function $f^E:\mathcal{Z}\rightarrow\mathcal{A}$, the observable decoding function $g_o^D:\mathcal{B}\rightarrow\widehat{\mathcal{Z}}$ and the semantic decoding function $g_s^D:\mathcal{B}\rightarrow\widehat{\mathcal{X}}$.*

Clearly, from the definition of block source-channel codes, we obtain $\kappa=1$ if goal-oriented single-letter codes are considered. Since in Lemma 2, the crucial condition is $\kappa R(\Delta_o,\Delta_s)=C(\Gamma)$, we next state a theorem where we derive necessary and sufficient conditions for that condition to hold when $\kappa=1$.

Theorem 1. *(Conditions for $R^L(\Delta_o,\Delta_s)=C(\Gamma)$) $R^L(\Delta_o,\Delta_s)=C(\Gamma)$ holds iff the following conditions are concurrently satisfied:*
(i) *the distribution $p(a)$ of $a=f^E(z)$ is capacity achieving for the channel $(p(b|a),c)$ at an average input cost $\Gamma=\mathbf{E}[c(\mathbf{a})]$, that is, $I^*(\mathbf{a};\mathbf{b})=C(\Gamma)$;*
(ii) *the conditional independence $p(\widehat{x}|\widehat{z},z)=p(\widehat{x}|\widehat{z})$ holds with $\widehat{z}=g_o^D(b)$ given z be the achieving distribution of SRDF at $R(\Delta_o)$ for an average distortion $\Delta_o=\mathbf{E}[d(\mathbf{z},\widehat{\mathbf{z}})]$, that is, $I^*(\mathbf{z};\widehat{\mathbf{z}})=R(\Delta_o)$;*
(iii) *the conditional independence $p(\widehat{z}|\widehat{x},z)=p(\widehat{z}|\widehat{x})$ holds with $\widehat{x}=g_o^D(b)$ given z be the achieving distribution of SRDF at $R(\Delta_s)$ for an average distortion $\Delta_s=\mathbf{E}[\widehat{d}(\mathbf{z},\widehat{\mathbf{x}})]$, that is, $I^*(\mathbf{z};\widehat{\mathbf{x}})=R(\Delta_s)$;*

(iv) *the pair* $(f^E(\cdot), g_o^D(\cdot))$ *results into* $I(\mathbf{z};\widehat{\mathbf{z}}) = I(\mathbf{a};\mathbf{b})$*;*
(v) *the pair* $(f^E(\cdot), g_s^D(\cdot))$ *results into* $I(\mathbf{z};\widehat{\mathbf{x}}) = I(\mathbf{a};\mathbf{b})$.

Proof: Suppose that for some values of (Δ_o, Δ_s), $R^L(\Delta_o, \Delta_s) = R(\Delta_o)$ holds. Then, the system model in Fig. 1 admits the following series of inequalities

$$R^L(\Delta_o,\Delta_s) \stackrel{(\star)}{=} R(\Delta_o) = \min_{\substack{p(\widehat{z}|z)\\ \mathbf{E}[d(\mathbf{z},\widehat{\mathbf{z}})]\leq\Delta_o}} I(\mathbf{z};\widehat{\mathbf{z}}) \stackrel{(\star\star)}{\leq} I(\mathbf{z};\widehat{\mathbf{z}})$$
$$\stackrel{(\star\star\star)}{\leq} I(\mathbf{a};\mathbf{b}) \stackrel{(\star\star\star\star)}{\leq} \max_{\substack{p(a)\\ \mathbf{E}[c(\mathbf{a})]\leq\Gamma}} I(\mathbf{a};\mathbf{b}) = C(\Gamma).$$

where $(\star)$ holds iff $p(\widehat{z}|\widehat{x},z) = p(\widehat{z}|z)$ (see condition **(ii)**); $(\star\star)$ holds with equality iff $p(\widehat{z}|z)$ achieves $R(\Delta_o)$ (see condition **(ii)**); $(\star\star\star)$ follows by data processing inequality that corresponds to the system in Fig. 1 and holds with equality iff condition **(iv)** holds; $(\star\star\star\star)$ holds with equality iff $p(a)$ is capacity achieving distribution (see condition **(i)**). Similarly, one can show that when for some value of (Δ_o, Δ_s), $R^L(\Delta_o, \Delta_s) = R(\Delta_s)$, $R^L(\Delta_o, \Delta_s) = R(\Delta_s) = C(\Gamma)$ iff conditions **(i)**, **(iii)**, **(v)**, hold. Hence $R^L(\Delta_o, \Delta_s) = C(P)$ holds iff conditions **(i)**-**(v)** hold. ■

Theorem 1 (except from **(i)**) is a generalization of [11, Lemma 2]. One interesting question that stems from Theorem 1 is if one can always find, constructively, goal-oriented single-letter source-channel codes that allow $R^L(\Delta_o, \Delta_s)$ to be achieved despite the fact that this bound is not achievable, in general, via the separation based schemes.

Next, we state a corollary to ensure that Lemma 2, **(ii)**, is always true when $R(\Delta_o, \Delta_s) = R^L(\Delta_o, \Delta_s) = C(\Gamma)$.

Corollary 1. *Suppose that for the setup in Fig. 1, the tuple* $(p(x), p(z|x), d_o, d_s)$ *is transmitted across a channel* $(p(b|a), c)$ *using goal-oriented single-letter codes* (f^E, g_o^D, g_s^D) *such that* $R^L(\Delta_o, \Delta_s) = C(\Gamma)$*. Then, the following statements hold.*
(i) Γ *cannot be decreased without decreasing* $C(\Gamma)$ *iff one of the following two conditions is satisfied:*

- (A) $I(\mathbf{a};\mathbf{b}) < C$*;*
- (B) $I(\mathbf{a};\mathbf{b}) = C$ *and among the possible maximizers, the capacity achieving distribution* $p(a)$ *is the unique maximer that achieves the lowest* Γ.

(ii) *Suppose that* $R^L(\Delta_o, \Delta_s) = R(\Delta_o)$*. Then* Δ_o *cannot be decreased without increasing* $R(\Delta_o)$ *iff one of the following two conditions is satisfied:*

- (A) $I(\mathbf{z};\widehat{\mathbf{z}}) > 0$*;*
- (B) $I(\mathbf{z};\widehat{\mathbf{z}}) = 0$ *and among the possible achieving minimizers, the rate-distortion achieving distribution* $p(\widehat{z}|z)$ *is the unique minimizer that achieves the lowest* Δ_o.

(iii) *Suppose that* $R^L(\Delta_o, \Delta_s) = R(\Delta_s)$*. Then* Δ_s *cannot be decreased without decreasing* $R(\Delta_s)$ *iff one of the following two conditions is satisfied:*

- (A) $I(\mathbf{z};\widehat{\mathbf{x}}) > 0$*;*
- (b) $I(\mathbf{z};\widehat{\mathbf{x}}) = 0$ *and among the possible achieving minimizers, the rate-distortion achieving distribution* $p(\widehat{x}|z)$ *is the unique minizer that achieves the lowest* Δ_s.

Proof: The proof of this corollary follows using similar arguments to [11, Proposition 5] with some small changes. For completeness, we give the missing part of the proof in the Appendix A. ■

V. Examples

In this section, we construct two examples assuming transmission using goal-oriented single-letter codes with one channel use per source symbol, i.e., $\kappa = 1$.

A. Binary sources - BSC

First, we give an example for binary alphabet sources and channels. This example is a generalization of the celebrated result by Jelinek in [10, Section 11.8].

Example 1. *Consider* $\mathcal{X} = \mathcal{Z} = \mathcal{A} = \mathcal{B} = \widehat{\mathcal{X}} = \widehat{\mathcal{Z}} = \{0, 1\}$. *Moreover, the semantic source* $\mathbf{x}$ *is uniform and the transition matrix* $\mathbf{z}$ *given* $\mathbf{x} = x$ *is doubly stochastic both modeled by*

$$p(x) = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix},\ p(z|x) = \begin{bmatrix} 1-w_1 & w_1 \\ w_1 & 1-w_1 \end{bmatrix} \tag{14}$$

where $w_1 \in \left[0, \frac{1}{2}\right)$*. Moreover, the distortion functions* $d_o(z, \widehat{z})$ *and* $d_s(x, \widehat{x})$ *are given by*

$$d_o(z,\widehat{z}) = \begin{cases} 0, \textit{if } z = \widehat{z} \\ 1, \textit{otherwise} \end{cases},\ d_s(x,\widehat{x}) = \begin{cases} 0, \textit{if } x = \widehat{x} \\ 1, \textit{otherwise} \end{cases}.$$

The channel is assumed to be the BSC with transition probability matrix $p(b|a)$ *given by*

$$p(b|a) = \begin{bmatrix} 1-\epsilon_1 & \epsilon_1 \\ \epsilon_1 & 1-\epsilon_1 \end{bmatrix}, \epsilon_1 \in \left[0, \frac{1}{2}\right). \tag{15}$$

Using (14) *under Hamming distortions, the SRDF was obtained in closed form in [13, Example 1] as follows*

$$R(D_o, D_s) = R^L(D_o, D_s) = \max\{R(D_o), R(D_s)\} \tag{16}$$

where

$$R(D_o) = \begin{cases} 1 - H_b(D_o), & \textit{if } D_o \in (0, \frac{1}{2}) \\ 0, & \textit{if } D_o \in [\frac{1}{2}, \infty) \end{cases} \tag{17}$$

and

$$R(D_s) = \begin{cases} 1 - H_b(\frac{D_s - w_1}{1 - 2w_1}), & \textit{if } D_s \in (w_1, \frac{1}{2}) \\ 0, & \textit{if } D_s \in [\frac{1}{2}, \infty) \end{cases} \tag{18}$$

where $H_b(\cdot)$ *denotes the binary entropy function.*

On the other hand, the channel capacity of the BSC is given by $C = 1 - H_b(\epsilon_1)$ *[16].*

Now choose the triplet (f^E, g_o^D, g_s^D) *in Definition 6 such that* $f^E(z) = z$, $g_o^D(b) = b$, $g_s^D(b) = b$ *(i.e., identity maps), which further implies that* $a = z$, $b = \widehat{z}$ *or* $b = \widehat{x}$.

Following, the optimality conditions of Lemma 2, we first require $R^L(\Delta_o, \Delta_s) = C$*. For this criterion to be valid, we need to check the following two cases. Suppose that* $R^L(\Delta_o, \Delta_s) = R(\Delta_o)$*, then, from* (5)*, we obtain* $\Delta_o = \mathbf{E}[d_o(\mathbf{z}, \widehat{\mathbf{z}})] = \epsilon_1$*, which further means that* $R(\Delta_o) = 1 - H_b(\Delta_o) = 1 - H_b(\epsilon_1)$*, hence* $R(\Delta_o) = C$*. Suppose that* $R^L(\Delta_o, \Delta_s) = R(\Delta_s)$*; then, again from* (5)*, we obtain*

$\Delta_s = \mathbf{E}[d_s(\mathbf{x},\widehat{\mathbf{x}})] = \mathbf{E}[\widehat{d}_s(\mathbf{z},\widehat{\mathbf{x}})] = \epsilon_1(1-2w_1)+w_1$, *which further means that* $R(\Delta_s) = 1 - H_b(\frac{\Delta_s - w_1}{1-2w_1}) = 1 - H_b(\epsilon_1)$, *hence* $R(\Delta_s) = C$. *The previous analysis confirms Lemma 2,* **(i)**. *On top of that, one can easily see that the conditions of Theorem 1 are also satisfied. Lemma 2,* **(ii)**, *is clearly satisfied as an immediate consequence of Corollary 1 because the specific choices of* w_1 *and* ϵ_1 *guarantee that the maximizer* $p(a)$ *is unique and achieves* C *and similarly the minimizers* $p(\widehat{z}|z)$ *and* $p(\widehat{x}|z)$ *achieve* $R(\Delta_o)$ *and* $R(\Delta_s)$, *respectively. Hence the specific realization of this goal-oriented source-channel communication model in Fig. 1 using single-letter codes is optimal.*

B. i.i.d *Gaussian sources-AWGN channel*

In the next example, we consider a jointly Gaussian semantic source and noisy observations conveyed across an AWGN channel with the cost function being a power and the distortions the standard squared error. This example is a generalization of a result by Goblick in [9].

Example 2. *Consider a joint Gaussian* i.i.d *process* $(\mathbf{x}^n, \mathbf{z}^n)$, *where the semantic source is such that* $\mathbf{x} \sim \mathcal{N}(0;\sigma_{\mathbf{x}}^2)$, $\sigma_{\mathbf{x}}^2 > 0$, *and the conditionally Gaussian distribution* $p^G(z|x)$ *is modeled by* $\mathbf{z} = \mathbf{x} + \mathbf{n}$, *where* $\mathbf{n} \sim \mathcal{N}(0;\sigma_{\mathbf{n}}^2)$ *is an* i.i.d *sequence of random variables. Moreover,* $d_o(z,\widehat{z})$ *and* $d_s(x,\widehat{x})$ *are given as follows*

$$d_o(z,\widehat{z}) = (z-\widehat{z})^2, \quad d_s(x,\widehat{x}) = (x-\widehat{x})^2. \tag{19}$$

The source is conveyed across a standard AWGN channel with transition probability $p(b|a)$ *modeled as* $\mathbf{b} = \mathbf{a} + \mathbf{m}$, *for which* $\mathbf{m} \sim \mathcal{N}(0;\sigma_{\mathbf{m}}^2), \sigma_{\mathbf{m}}^2 > 0$, *with input power constraint* $\mathbf{E}[\mathbf{a}^2] \leq P$, $P > 0$.

For the specific class of Gaussian semantic sources and noisy observations under squared-error distortions, the SRDF is obtained in closed form in [12, Proposition 3] as follows

$$R(D_o, D_s) = R^L(D_o, D_s) = \max\{R(D_o), R(D_s)\} \tag{20}$$

where

$$R(D_o) = \begin{cases} \frac{1}{2}\log\frac{\sigma_{\mathbf{z}}^2}{D_o}, & \textit{if } D_o \in (0, \sigma_{\mathbf{z}}^2) \\ 0, & \textit{otherwise} \end{cases} \tag{21}$$

and

$$R(D_s) = \begin{cases} \frac{1}{2}\log\frac{\sigma_{\mathbf{x}}^4}{\sigma_{\mathbf{z}}^2(D_s - \sigma_{\mathbf{x}|\mathbf{z}}^2)}, & \textit{if } D_s \in \left(\sigma_{\mathbf{x}|\mathbf{z}}^2, \frac{\sigma_{\mathbf{x}}^4}{\sigma_{\mathbf{z}}^2}\right) \\ 0, & \textit{otherwise} \end{cases} \tag{22}$$

where $\sigma_{\mathbf{x}|\mathbf{z}}^2 = \frac{\sigma_{\mathbf{x}}^2\sigma_{\mathbf{n}}^2}{\sigma_{\mathbf{x}}^2+\sigma_{\mathbf{n}}^2}$ *is the conditional variance of the conditionally Gaussian distribution* $p^G(x|z)$.

The constrained channel capacity of AWGN is well known and is given by $C(P) = \frac{1}{2}\log\left(1+\frac{P}{\sigma_{\mathbf{m}}^2}\right)$ *[16].*

Following Lemma 2, **(i)**, *we first require* $R^L(\Delta_o, \Delta_s) = C(\Gamma)$. *Similarly to Example 1, we check if (A)* $R(\Delta_o) = C(\Gamma)$*; (B)* $R(\Delta_s) = C(\Gamma)$. *Since we have now the presence of the cost* $\Gamma > 0$, *the approach is slightly different. To ensure case* (A), *we require (similar to [9])* $\Delta_o = \frac{\sigma_{\mathbf{z}}^2\sigma_{\mathbf{m}}^2}{\Gamma+\sigma_{\mathbf{m}}^2}$, *which based on standard terminology we referred to as the minimum observable achievable distortion (*$D_o^{\min}$*). To achieve this operational point, we need to find a specific encoder and decoder operating at* $D_o^{\min}$. *Indeed, if we define the single-letter source-channel encoder and decoder pair as*

$$\mathbf{a} = \sqrt{\frac{\Gamma}{\sigma_{\mathbf{z}}^2}}\mathbf{z}, \quad \widehat{\mathbf{z}} = \sqrt{\frac{\sigma_{\mathbf{z}}^2}{\Gamma}}\frac{\Gamma}{\Gamma+\sigma_{\mathbf{m}}^2}\mathbf{b}, \tag{23}$$

then, the resulting $\Delta_o = \mathbf{E}[(\mathbf{z}-\widehat{\mathbf{z}})^2] = D_o^{\min}$ *and case (A) is verified. To ensure case* (B), *we require* $\Delta_s = \frac{\sigma_{\mathbf{x}}^4\sigma_{\mathbf{m}}^2}{\sigma_{\mathbf{z}}^2(\sigma_{\mathbf{m}}^2+\Gamma)} + \sigma_{\mathbf{x}|\mathbf{z}}^2$, *which is the minimum semantic achievable distortion* $D_s^{\min}$. *To achieve* $D_s^{\min}$, *we choose the following pair of single-letter encoder and decoder:*

$$\mathbf{a} = \sqrt{\frac{\Gamma}{\frac{\sigma_{\mathbf{x}}^4}{\sigma_{\mathbf{z}}^2}}}\mu_{\mathbf{x}|\mathbf{z}}, \quad \widehat{\mathbf{x}} = \sqrt{\frac{\frac{\sigma_{\mathbf{x}}^4}{\sigma_{\mathbf{z}}^2}}{\Gamma}}\frac{\Gamma}{\Gamma+\sigma_{\mathbf{m}}^2}\mathbf{b} \tag{24}$$

where $\mu_{\mathbf{x}|\mathbf{z}} \equiv \mathbf{E}[\mathbf{x}|\mathbf{z}] = \frac{\sigma_{\mathbf{xz}}}{\sigma_{\mathbf{z}}^2}\mathbf{z} = \frac{\mathbf{E}[\mathbf{xz}]}{\sigma_{\mathbf{z}}^2}\mathbf{z} = \frac{\sigma_{\mathbf{x}}^2}{\sigma_{\mathbf{z}}^2}\mathbf{z}$ *is the conditional mean of* $p^G(x|z)$. *The choice in* (24) *results into an operational distortion*

$$\begin{aligned}\Delta_s &= \mathbf{E}[(\mathbf{x}-\widehat{\mathbf{x}})^2] \\ &= \mathbf{E}[(\mathbf{x}-\mu_{\mathbf{x}|\mathbf{z}}+\mu_{\mathbf{x}|\mathbf{z}}-\widehat{\mathbf{x}})^2] \\ &\overset{(*)}{=} \mathbf{E}[(\mathbf{x}-\mu_{\mathbf{x}|\mathbf{z}})^2] + \mathbf{E}[(\mu_{\mathbf{x}|\mathbf{z}}-\widehat{\mathbf{x}})^2] \\ &\overset{(**)}{=} D_s^{\min}\end{aligned}$$

where $(*)$ *follows from the orthogonality principle;* $(**)$ *follows from the choice of* (24). *One can observe that the above methodology verifies the iff conditions of Theorem 1. Finally, Lemma 2,* **(ii)**, *is verified because of Corollary 1. Hence, the specific realization of the goal-oriented communication system in Fig. 1, using single-letter codes and uncoded transmission is optimal.*

The two application examples described above, demonstrate the vital role of the fidelity constraints in goal-oriented communications over noisy channels using single-letter codes.

VI. Conclusions and Ongoing Study

We studied a new variation of a JSCC problem that has natural goal-oriented features from an inference perspective. We defined a variation of block source-channel codes and gave optimality conditions for the system model to operate optimally. Our main focus was on single-letter codes for which we derived a new theorem that provides necessary and sufficient conditions for the probabilistic matching of the semantic source message with the noisy channel. Two application examples were presented to corroborate our theoretical results.

The optimality conditions of our JSCC setup can be further studied to take into account bandwidth constraints (i.e., $\kappa < 1$, $\kappa > 1$) as Lemma 2 suggests. This study will follow in the extended version of the paper. It would be interesting to use tools from the machine learning community for more complex analysis of our setup and its extensions.

References

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

[2] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journ. on Sel. Areas in Commun.*, vol. 41, no. 1, pp. 5–41, 2023.

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

[4] S. Wang, Q. Liu, P. U. Abara, J. S. Baras, and S. Hirche, "Value of information in networked control systems subject to delay," in *Proc. IEEE Conf. Decision Control*, 2021, pp. 1275–1280.

[5] 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.

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

[7] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., 1968.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[9] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. 11, no. 4, pp. 558–567, 1965.

[10] F. Jelinek, *Probabilistic Information Theory: discrete and memoryless models*. McGraw-Hill, NY, 1968.

[11] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, 2003.

[12] 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.

[13] P. A. Stavrou and M. Kountouris, "A rate distortion approach to goal-oriented communication," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2022, pp. 590–595.

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

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

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

Appendix A
Proof of Corollary 1

We only prove statement **(iii)** as statements **(i)**, **(ii)** are already known from [11, Proposition 5].

(iii) *Sufficiency of (A)*: Define $\Delta_s^{\max} \triangleq \min\{D_s : R(D_s) = 0\}$ and $\Delta_s^{\min} \triangleq \{D_s : R(D_s) = \{\text{maximum positive value}\}\}$. For $\Delta_s \geq \Delta_s^{\min} > 0$ and using the fact that $R^L(\Delta_o, \Delta_s) = R(\Delta_s) = C(\Gamma)$, we have that $0 < I(\mathbf{z}; \widehat{\mathbf{x}}) = R(\Delta_s)$ which further implies that $R(\Delta_s) > 0$. But when $R(\Delta_s) > 0$ we know that its curve is convex and strictly decreasing with respect to Δ_s (see e.g., [8, Chapter 8]), hence there is a unique Δ_s for each point on the curve $R(\Delta_s)$. This implies that Δ_s cannot be decreased without increasing $R(\Delta_s)$.

Necessity of (A): If $\Delta_s \geq \Delta_s^{\max}$, then, $I(\mathbf{z}; \widehat{\mathbf{x}}) = 0$, hence it is possible to have multiple Δ_s's without changing the value of $R(\Delta_s)$.

Sufficiency of (B): If $I(\mathbf{z}; \widehat{\mathbf{x}}) = 0$ and among the possible minimizers, the rate-distortion achieving distribution $p(\widehat{x}|z)$ is the one that achieves the lowest Δ_s this means that it is impossible to alter Δ_s without changing $R(\Delta_s)$ because we will need to change $p(\widehat{x}|z)$ (this argument is trivially obtained if $p(\widehat{x}|z)$ is unique).

Necessity of (B): If $I(\mathbf{a}; \mathbf{b}) = 0$, then, $\Delta_s \geq \Delta_s^{\max}$, hence there are multiple minimizers $p(\widehat{x}|z)$ that achieve different values of Δ_s's but not necessarily its minimum value which is $\Delta_s^{\max}$. (again this argument is trivially obtained as long as $p(\widehat{x}|z)$ is unique).