Indirect Rate Distortion Functions with f-Separable Distortion Criterion

Photios A. Stavrou*, Yanina Shkel†, Marios Kountouris*
* Communication Systems Department, EURECOM, Sophia-Antipolis, France
† School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland,
{fotios.stavrou,marios.kountouris}@eurecom.fr, yanina.shkel@epfl.ch

***Abstract*—We consider a remote source coding problem subject to a distortion fidelity. Contrary to the use of the classical separable distortion criterion, herein we consider the more general, f-separable distortion measure and study its implications on the characterization of the minimum achievable rates (also called f-separable indirect rate distortion function (iRDF)) under both excess and average distortion constraints. First, we provide a single-letter characterization of the optimal rates subject to an excess distortion using properties of the f-separable distortion. Our main result is a single-letter characterization of the f-separable iRDF subject to an average distortion constraint. As a consequence of the previous results, we also show a series of equalities that hold using either indirect or classical RDF under f-separable excess or average distortions. We corroborate our results with two application examples in which new closed-form solutions are derived, and based on these, we also recover known special cases.**

I. Introduction

The mathematical analysis of the lossy source coding under a distortion fidelity, called rate distortion theory [1], was developed under the assertion that an encoder observes an information source $\mathbf{x}$ with distribution $p(x)$ defined on the alphabet space $\mathcal{X}$, and the aim is for the decoder to reconstruct in a minimal end-to-end rate-constrained manner, its representation $\widehat{\mathbf{x}}$ defined on an alphabet $\widehat{\mathcal{X}}$ within a distortion measure $d:\mathcal{X}\times\widehat{\mathcal{X}}\mapsto[0,\infty)$. When the information source generates a sequence of n realizations, the source sequence induces the distribution $p(x^n)$ on the Cartesian product alphabet space $\mathcal{X}^n$, with its reconstruction alphabet being $\widehat{\mathcal{X}}^n$. For the latter case, Shannon in [1] extended the single-letter expression of the distortion measure to the n-letter expression $d^n:\mathcal{X}^n\times\widehat{\mathcal{X}}^n\mapsto[0,\infty)$ by taking the arithmetic mean of single-letter distortions, i.e.,

$$d^n(x^n,\widehat{x}^n)=\frac{1}{n}\sum_{i=0}^{n}d(x_i,\widehat{x}_i), \qquad (1)$$

which is often encountered as *separable*, *additive* or *per-letter* distortion measure.

A natural extension of the lossy source coding problem, called *indirect* or *remote* lossy source coding, was proposed almost fifteen year later in [2]. Therein the authors considered the case where the encoder observes a noisy version of the source $\mathbf{x}$, say $\mathbf{z}$, and the goal is to reconstruct $\widehat{\mathbf{x}}$ with minimal rates subject to an average distortion $d:\mathcal{X}\times\widehat{\mathcal{X}}\mapsto[0,\infty)$. A major result in [2] is that for stationary memoryless sources, the fundamental limit in the asymptotic regime corresponds to the classical lossy source coding problem with an amended average distortion constraint. Subsequently, this problem and some of its variants, e.g., non-asymptotic analysis, excess distortion measures, multi-terminal systems, were revisited by many researchers, see e.g., [3]–[12] and references therein.

All the aforementioned efforts in [3]–[12], consider separable distortion penalties. On one hand, the separability assumption is natural and quite appealing when it comes to the derivation of tractable characterizations of the fundamental trade-offs between the coding (or compressed) rate and its corresponding distortion. On the other hand, the separability assumption is very restrictive because it only models distortion penalties that are *linear functions* between the single-letter distortion and the reconstruction signal. However, in real-world applications, distortion measures may be highly *non-linear*. To address this issue and inspired by [13], here we consider a much broader class of distortion measures, namely, f-separable distortion measures (for details on this class of fidelity constraints see Appendix A).

In this work, we derive the following new results: (i) a single-letter characterization of the minimal rates subject to an excess distortion using properties of the f-separable distortion (see Lemma 1); (ii) a single-letter characterization of the f-separable iRDF (obtained for finite alphabets) subject to an average distortion constraint that is obtained under relatively mild regularity conditions and by making use of a strong converse theorem [8] (see Theorem 1); (iii) new series of equalities under f-separable excess or average distortion constraints using indirect or classical RDFs (see Corollary 1 and Theorem 1); (iv) two application examples in which new analytical solutions are derived for various types of f-separable average distortions; we also explain how these analytical expressions recover known results as special cases (see Examples 1, 2). It is worth mentioning that from (ii), we also derive the implicit solution of the optimal minimizer that achieves the characterization of the f-separable iRDF (see Corollary 2). This result can be readily used to derive new Blahut-Arimoto type of algorithms [14], [15] for a much richer class of distortion penalties.

II. Problem Formulation

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 remote information of the source is in $\mathbf{x}$ whereas $\mathbf{z}$ is the noisy observation at the encoder side. The goal is to study the remote source coding problem [2], [5], [6] under an f-separable distortion measure.

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 independent and identically distributed (i.i.d) RVs $(\mathbf{x}^n, \mathbf{z}^n)$. The *encoder (E)* and the *decoder (D)*, are modeled by the mappings

$$f^E : \mathcal{Z}^n \to \mathcal{W}, \quad g^D : \mathcal{W} \to \widehat{\mathcal{X}}^n \tag{2}$$

where the index set $\mathcal{W} \in \{1, 2, \ldots, M\}$.

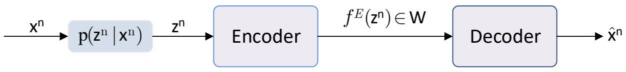

Fig. 1: System model.

We consider a per-letter distortion measure responsible to penalize the remote information source in Fig. 1 given by $d : \mathcal{X} \times \widehat{\mathcal{X}} \mapsto [0, \infty)$ and their corresponding n-letter expressions given by $d^n : \mathcal{X}^n \times \widehat{\mathcal{X}}^n \mapsto [0, \infty)$. This setting has recently gained attention in the context of goal-oriented communication [16], [17], where $\mathbf{x}$ can represent the semantic or intrinsic information of the source, which is not directly observable, whereas $\mathbf{z}$ is the noisy observation of the source at the encoder side.

Next, we define the precise terminology of the noisy lossy source codes for the single-letter and the multi-letter case (without restricting to i.i.d processes at this stage).

Definition 1. *(Noisy lossy source codes) Consider constants $\epsilon \in [0, 1)$, $D \geq 0$, and an integer M.*

(1) *We say that a noisy lossy source-code (f^E, g^D) is an (M, D)-noisy lossy source code on $(\mathcal{X}, \mathcal{Z}, \widehat{\mathcal{X}}, d)$ such that $\mathbf{x} - \mathbf{z} - \widehat{\mathbf{x}}$, if $\mathbf{E}[d(\mathbf{x}, \widehat{\mathbf{x}})] \leq D$, where $\widehat{\mathbf{x}} = g^D(f^E(\mathbf{z}))$.*
(2) *We say that a noisy lossy source-code (f^E, g^D) is an (M, D, ϵ)-noisy lossy source code on $(\mathcal{X}, \mathcal{Z}, \widehat{\mathcal{X}}, d)$ such that $\mathbf{x} - \mathbf{z} - \widehat{\mathbf{x}}$, if $\mathbf{P}[d(\mathbf{x}, \widehat{\mathbf{x}}) > D] \leq \epsilon$ where $\widehat{\mathbf{x}} = g^D(f^E(\mathbf{z}))$.*
(3) *If (f^E, g^D) is an (M, D)-noisy lossy source code on $(\mathcal{X}^n, \mathcal{Z}^n, \widehat{\mathcal{X}}^n, d^n)$ such that $\mathbf{x}^n - \mathbf{z}^n - \widehat{\mathbf{x}}^n$, we say that (f^E, g^D) is an (n, M, D)-noisy lossy source code.*
(4) *If (f^E, g^D) is an (M, D, ϵ)-noisy lossy source code on $(\mathcal{X}^n, \mathcal{Z}^n, \widehat{\mathcal{X}}^n, d^n)$ such that $\mathbf{x}^n - \mathbf{z}^n - \widehat{\mathbf{x}}^n$, we say that (f^E, g^D) is an (n, M, D, ϵ)-noisy lossy source code.*

We remark the following special case of Definition 1.

Remark 1. *(On Definition 1) In our analysis, we will also consider as a special case the classical (noiseless) lossy source codes subject to similar single-letter and multi-letter distortion measures as in the case of noisy lossy source coding. This means that we will use special cases of Definition 1. For example, for a noiseless lossy source code, Definition 1, (1), will be modified as follows*

- *we say that a lossy source-code (f^E, g^D) is an (M, D)-lossy source code on $(\mathcal{X}, \widehat{\mathcal{X}}, d)$ if $\mathbf{E}[d(\mathbf{x}, \widehat{\mathbf{x}})] \leq D$, where $\widehat{\mathbf{x}} = g^D(f^E(\mathbf{x}))$ (because $\mathbf{x} = \mathbf{z}$).*

Definition 1, (2)-(4), are modified accordingly.

Using [13, Definition 1], we consider an *f-separable distortion measure* associated with the remote information source of the setup in Fig. 1 defined as follows

$$d_f^n(x^n, \hat{x}^n) \triangleq f^{-1}\left(\frac{1}{n}\sum_{t=1}^{n} f(d(x_i, \hat{x}_i))\right) \tag{3}$$

where $f(\cdot)$ is a continuous, increasing function on $[0, \infty)$.

In the sequel, we give the definitions of indirect and direct (or classical) RDFs under f-separable distortion measures. To do it, we need the following definition of achievability.

Definition 2. *(Achievability) Suppose that a sequence of distortion measures $\{d^n : n = 1, 2, \ldots\}$ on $(\mathcal{X}^n, \widehat{\mathcal{X}}^n)$ is given, such that $\mathbf{x}^n - \mathbf{z}^n - \widehat{\mathbf{x}}^n$. Then, we define the following statements.*

(1) *The rate distortion tuple (R, D) is indirectly achievable if there exists a sequence (n, M_n, D^n)-noisy lossy source codes such that $\limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R$, $\limsup_{n\to\infty} D^n \leq D$.*
(2) *The rate distortion tuple (R, D) is indirectly and excess distortion achievable if for any $\gamma > 0$ there exists a sequence $(n, M_n, D + \gamma, \epsilon_n)$-noisy lossy source codes such that $\limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R$, $\limsup_{n\to\infty} \epsilon_n = 0$, where ϵ_n denotes the decoding error probability, i.e., $\epsilon_n = \mathbf{P}\big[\mathbf{x}^n \neq g^D(f^E(\mathbf{z}^n))\big]$.*

If we assume sequences of noiseless lossy source codes, we say that a rate distortion tuple (R, D) is directly (and excess distortion) achievable in analogous way to Definition 2, with $\mathcal{X}^n = \mathcal{Z}^n$. This means that the sequence of distortion measures $\{d^n : n = 1, 2, \ldots\}$ can be defined either on $(\mathcal{Z}^n, \widehat{\mathcal{X}}^n)$ or on $(\mathcal{X}^n, \widehat{\mathcal{X}}^n)$.

Definition 3. *(iRDF) Given a single-letter distortion measure $d \colon \mathcal{X} \times \widehat{\mathcal{X}} \to [0, \infty)$ and a continuous, increasing function f on $[0, \infty)$, let $\{d_f^n : n = 1, 2, \ldots\}$ be a sequence of f-separable distortion measures. Then,*

$$\mathcal{I}_{f,d}(D) = \inf\{R : (R, D) \text{ is indirectly achievable}\} \tag{4}$$

and

$$\widehat{\mathcal{I}}_{f,d}(D) = \inf\big\{R : (R, D) \text{ is indirectly and excess distortion achievable}\big\}.$$

If f is the identity function, then we have a sequence of separable distortion measures; in this case we omit the subscript f and write $\mathcal{I}_d(D)$ and $\widehat{\mathcal{I}}_d(D)$.

Definition 4. *(Direct RDF) Given a single-letter distortion measure $d \colon \mathcal{Z} \times \widehat{\mathcal{X}} \to [0, \infty)$ and a continuous, increasing function f on $[0, \infty)$, let $\{d_f^n : n = 1, 2, \ldots\}$ be a sequence of f-separable distortion measures. Then,*

$$\mathcal{R}_{f,d}(D) = \inf\{R : (R, D) \text{ is directly achievable}\} \tag{5}$$

and

$$\widehat{\mathcal{R}}_{f,d}(D) = \inf\{R : (R, D) \text{ is directly and excess distortion achievable}\}.$$

If f is the identity function, we omit the subscript f and write $\mathcal{R}_d(D)$ and $\widehat{\mathcal{R}}_d(D)$.

We give the following remark for the previous two Definitions.

Remark 2. *(On Definitions 3, 4) In this work our goal is to characterize the f-separable iRDFs $\mathcal{I}_{d,f}(D)$ and $\widehat{\mathcal{I}}_{d,f}(D)$ for a given distortion measure $d(\cdot,\cdot)$ and a function $f(\cdot)$. In addition to the f-separable iRDFs, we consider the following three special cases: (1) separable RDF $\mathcal{R}_d(D)$ and $\widehat{\mathcal{R}}_d(D)$, (2) separable iRDFs $\mathcal{I}_d(D)$ and $\widehat{\mathcal{I}}_d(D)$, and (3) f-separable RDFs $\mathcal{R}_{d,f}(D)$ and $\widehat{\mathcal{R}}_{d,f}(D)$. To state our results, we compare these different classes of RDFs to each other. While the iRDFs is defined over some space $(\mathcal{X}, \mathcal{Z}, \widehat{\mathcal{X}}, d)$, it is possible to generate modified direct RDFs from iRDFs in which case these are definite over the space $(\mathcal{Z}, \widehat{\mathcal{X}}, \tilde{d})$, where $\tilde{d}\colon \mathcal{Z} \times \widehat{\mathcal{X}} \to [0,\infty)$ is an amended distortion measure. In general, the underlying space for the direct RDFs should be clear from context. For example, $\mathcal{R}_d(D)$ refers to an RDF on $(\mathcal{X}, \widehat{\mathcal{X}}, d)$, while $\mathcal{R}_{\tilde{d}}(D)$ refers to an RDF on $(\mathcal{Z}, \widehat{\mathcal{X}}, \tilde{d})$.*

III. Prior Work

Next, we discuss more extensively some prior results that will be used in our main results.

A. RDF under Average and Excess Constraints

For i.i.d sources with finite alphabets $(\mathcal{X}, \widehat{\mathcal{X}})$ and bounded distortion measure d, the RDF is given by

$$\mathcal{R}_d(D) = \inf_{q(\widehat{x}|x)\colon\, \mathbf{E}[d(\mathbf{x},\widehat{\mathbf{x}})] \le D} I(\mathbf{x};\widehat{\mathbf{x}}).$$

See e.g., [18, Theorem 10.2.1] and [19, Theorem 5.2.1]. Moreover, we know that for stationary ergodic sources with a bounded distortion measure,

$$\mathcal{R}_d(D) = \widehat{\mathcal{R}}_d(D). \tag{6}$$

That is, the RDF is the same under average and excess distortion constraints [19, Theorem 5.9.1]. We also know that for stationary ergodic sources $\widehat{\mathcal{R}}_d(D)$ satisfies the so-called *strong converse* [8], [20], see Appendix D. Finally, the second order asymptotic expansion of $\widehat{\mathcal{R}}_d(D)$ is given as well, see e.g., [21], [22], but this type of analysis is beyond the scope of the present paper.

B. iRDF

For i.i.d sources with finite alphabets $(\mathcal{X}, \mathcal{Z}, \widehat{\mathcal{X}})$ and bounded distortion measure d, the iRDF is given by

$$\begin{aligned}\mathcal{I}_d(D) &= \inf_{\substack{q(\widehat{x}|z)\colon \\ \mathbf{E}[d(\mathbf{x},\widehat{\mathbf{x}})]\le D}} I(\mathbf{z};\widehat{\mathbf{x}}) \\ &\overset{(a)}{=} \inf_{\substack{q(\widehat{x}|z)\colon \\ \mathbf{E}[\tilde{d}(\mathbf{z},\widehat{\mathbf{x}})]\le D}} I(\mathbf{z};\widehat{\mathbf{x}}) \equiv \mathcal{R}_{\tilde{d}}(D)\end{aligned} \tag{7}$$

where (a) follows from [2] (see also Remark 2) and $\mathcal{R}_{\tilde{d}}(D)$ is the direct RDF for $(\mathcal{X}, \mathcal{Z})$ with the amended distortion given by $\tilde{d}(z,\widehat{x}) = \sum_{\mathcal{X}} p(x|z) d(x,\widehat{x})$. In other words, the indirect rate distortion problem reduces to a direct rate distortion problem with a modified per-letter distortion measure [2], [5], [6], [8]. Moreover, for i.i.d sources, the iRDF is the same under average and excess distortion constraints

$$\mathcal{I}_d(D) = \widehat{\mathcal{I}}_d(D) \tag{8}$$

and the strong converse also holds [8]. Finally, for this problem, the second-order asymptotic analysis has been addressed in [8] where it was shown that the equivalence between direct and indirect problems no longer holds in the second-order (dispersion) sense.

C. f-Separable RDF

Similar equivalence results hold for f-separable RDFs. Specifically, for i.i.d sources

$$\mathcal{R}_{f,d}(D) = \mathcal{R}_{\bar{d}}(f(D)) = \inf_{\substack{q(\widehat{x}|x) \\ \mathbf{E}[\bar{d}(\mathbf{x},\widehat{\mathbf{x}})]\le f(D)}} I(\mathbf{x};\widehat{\mathbf{x}}) \tag{9}$$

where $\mathcal{R}_{\bar{d}}(\cdot)$ is the separable RDF for $(\mathcal{X}, \widehat{\mathcal{X}})$ with the amended distortion given by $\bar{d}(x,\widehat{x}) = f(d(x,\widehat{x}))$, see [13]. More generally, it is shown in [13] that for the f-separable rate distortion problem

$$\widehat{\mathcal{R}}_{f,d}(D) = \widehat{\mathcal{R}}_{\bar{d}}(f(D)). \tag{10}$$

That is, under excess distortion criterion, the f-separable RDF reduces to the classical separable case without any assumption on the underlying source. In fact for stationary ergodic sources, this result extends to both average and excess distortion criteria under some regularity assumptions (see [13, Theorem 1]), namely,

$$\mathcal{R}_{f,d}(D) = \widehat{\mathcal{R}}_{f,d}(D). \tag{11}$$

We remark that the generalizations of the classical rate-distortion problem to indirect and f-separable rate distortion problems have intriguing parallels. Both generalizations could be expressed in terms of a classical amended rate distortion problem. The same insight holds when we apply both generalizations simultaneously. As we will see next, the resulting rate-distortion function could be expressed in terms of the classical amended rate distortion problem.

IV. Single-letter characterization of the operational rates for i.i.d sources

In this section, we characterize the f-separable iRDFs for the setup in Fig. 1 for i.i.d sources. Specifically, our main result states that for i.i.d sources over finite alphabets (under mild regularity assumptions) we have that

$$\mathcal{I}_{f,d}(D) = \mathcal{R}_{\tilde{d}}(f(D)) \tag{12}$$

where $\mathcal{R}_{\tilde{d}}(D)$ is the RDF for $(\mathcal{X}, \mathcal{Z}, \tilde{d})$ with the amended distortion given by $\tilde{d}(z,\widehat{x}) = \sum_{\mathcal{X}} p(x|z) f(d(x,\widehat{x}))$.

First, we give a lemma in which we characterize the f-separable iRDF under the excess distortion criterion.

Lemma 1. *(f-separable iRDF under excess distortion) Given a single-letter distortion measure $d\colon \mathcal{X}\times\widehat{\mathcal{X}}\to[0,\infty)$ and a continuous, increasing function f on $[0,\infty)$,*

$$\widehat{\mathcal{I}}_{f,d}(D)=\widehat{\mathcal{I}}_{\bar{d}}(f(D)) \tag{13}$$

where $\widehat{\mathcal{I}}_{\bar{d}}(f(D))$ is computed subject to the single-letter separable distortion measure $\bar{d}(x,\widehat{x})=f(d(x,\widehat{x}))$.

Proof: The proof is a generalization of the proof in [13, Lemma 1] and is given in Appendix B. ■

Next, we make assumptions that will be used to derive the single-letter information theoretic characterization to our problem. These assumptions are a counterpart of the assumptions utilized in [13, Theorem 1]; however, due to the difficulty of the indirect rate distortion problem, these assumptions are more restrictive, e.g., we only consider finite alphabets.

Assumptions. *Suppose that the following statements are true.*

(A1) *The joint process $\{(\mathbf{x}^n,\mathbf{z}^n):\ n=1,2,\ldots\}$ is* i.i.d *sequence of random variables, namely, $p(x^n,z^n)=p(x)p(z|x)\times\ldots\times p(x)p(z|x)=p(x|z)p(z)\times\ldots\times p(x|z)p(z)$, for any n;*

(A2) *The single-letter distortion $d(\cdot,\cdot)$ and is such that*

$$\max_{(x,\widehat{x})\in\mathcal{X}\times\widehat{\mathcal{X}}} d(x,\widehat{x})<\infty; \tag{14}$$

(A3) *The alphabets $(\mathcal{X},\mathcal{Z},\widehat{\mathcal{X}})$ are finite.*

In particular, assumption (A2) rules out pathological rate-distortion function for which finite distortion is only possible at full rate.

Corollary 1. *(Consequence of Lemma 1) Under Assumptions* **(A1)-(A3)***, a consequence of Lemma 1 is the following series of equalities*

$$\widehat{\mathcal{I}}_{f,d}(D)=\widehat{\mathcal{I}}_{\bar{d}}(f(D))=\widehat{\mathcal{R}}_{f,\hat{d}}(D)=\widehat{\mathcal{R}}_{\tilde{d}}(f(D)) \tag{15}$$

where

$$\bar{d}(x,\widehat{x})=f(d(x,\widehat{x})) \tag{16}$$

$$\hat{d}(z,\widehat{x})=f^{-1}\left(\sum_x p(x|z)f(d(x,\widehat{x}))\right) \tag{17}$$

$$\tilde{d}(z,\widehat{x})=\sum_x p(x|z)f(d(x,\widehat{x})). \tag{18}$$

Proof: The first equality, $\widehat{\mathcal{I}}_{f,d}(D)=\widehat{\mathcal{I}}_{\bar{d}}(f(D))$, is shown in Lemma 1. We have that $\widehat{\mathcal{I}}_{\bar{d}}(f(D))=\widehat{\mathcal{R}}_{\tilde{d}}(f(D))$ from (6), (7) and (8). Finally, $\widehat{\mathcal{R}}_{f,\hat{d}}(D)=\widehat{\mathcal{R}}_{\tilde{d}}(f(D))$ follows from (10). This completes the proof. ■

Next, we show the same result for the average rate-distortion functions.

Theorem 1. *(f-separable iRDF under average distortion) Under Assumptions* **(A1)-(A3)***, the f-separable iRDF under an average distortion constraint satisfies the following equality*

$$\mathcal{I}_{f,d}(D)=\mathcal{I}_{\bar{d}}(f(D)) \tag{19}$$

where $\bar{d}(x,\widehat{x})$ is given in (16)*. In particular, this implies that under Assumptions* **(A1)-(A3)***,*

$$\mathcal{I}_{f,d}(D)=\widehat{\mathcal{I}}_{f,d}(D)=\mathcal{R}_{f,\hat{d}}(D)=\mathcal{R}_{\tilde{d}}(f(D)) \tag{20}$$

and

$$\mathcal{I}_{f,d}(D)=\inf_{\substack{q(\widehat{x}|z)\\ \mathbf{E}\left[\tilde{d}(\mathbf{z},\widehat{\mathbf{x}})\right]\le f(D)}} I(\mathbf{z};\widehat{\mathbf{x}}) \tag{21}$$

where $\hat{d}(z,\widehat{x})$ and $\tilde{d}(z,\widehat{x})$ are given by (17) *and* (18)*, respectively.*

Proof: Equations (20) and (21) follow from (19) and the results in Section III. Namely, we have that $\mathcal{I}_{\bar{d}}(f(D))=\mathcal{R}_{\tilde{d}}(f(D))$ from (7); $\mathcal{R}_{f,\hat{d}}(D)=\mathcal{R}_{\tilde{d}}(f(D))$ from (10) and (11), and $\mathcal{I}_{\bar{d}}(f(D))=\widehat{\mathcal{I}}_{\bar{d}}(f(D))=\widehat{\mathcal{I}}_{f,d}(D)$ from (8) and Lemma 1. Likewise, (21) is a consequence of (19) and (7). It remains to show (19). First note that f-separable iRDF can be upper bounded as follows:

$$\mathcal{I}_{f,d}(D)\overset{(a)}{\le}\widehat{\mathcal{I}}_{f,d}(D)\overset{(b)}{=}\widehat{\mathcal{I}}_{\bar{d}}(f(D))\overset{(c)}{=}\mathcal{I}_{\bar{d}}(f(D)) \tag{22}$$

where (a) is a consequence of Assumption **(A2)** and Lemma 2 in Appendix C; (b) follows from Lemma 1; (c) follows from the equivalence between excess and average iRDF, see (8). The other direction,

$$\mathcal{I}_{f,d}(D)\ge\mathcal{I}_{\bar{d}}(f(D)) \tag{23}$$

is a consequence of the strong converse by [8] and is shown in Lemma 3. This completes the proof. ■

One pleasing consequence of Theorem 1 is the following corollary.

Corollary 2. *(Implicit solution of $\mathcal{I}_{\bar{d}}(f(D))$) The characterization in* (21) *via* (19) *admits the following implicit solution to its minimizer*

$$p^*(\widehat{x}|z)=\frac{e^{s\tilde{d}(z,\widehat{x})}p^*(\widehat{x})}{\sum_{\widehat{x}}e^{s\tilde{d}(z,\widehat{x})}p^*(\widehat{x})}, \tag{24}$$

where $s<0$ is the Lagrange multiplier associated with the amended distortion penalty $\mathbf{E}[\tilde{d}(z,\widehat{x})]\le f(D)$ and $p^*(\widehat{x})=\sum_z q^*(\widehat{x}|z)p(z)$ is the $\widehat{\mathcal{X}}$-marginal of the output i.i.d *process $\widehat{\mathbf{x}}^n$. Moreover, the optimal parametric solution of* (21) *via* (20) *when $\mathcal{I}_{f,d}(D)>0$ is given by*

$$\mathcal{I}_{f,d}(D^*)=sf(D^*)-\sum_z p(z)\log\left(\sum_{\widehat{x}}e^{s\tilde{d}(z,\widehat{x})}p^*(\widehat{x})\right). \tag{25}$$

Proof: See Appendix E. ■

By taking $p(z|x)$ to be a noiseless channel, Corollary 2 gives us an implicit solution for $\mathcal{R}_{f,d}(D)$ which was suggested in [13].

V. EXAMPLES

In what follows, we give two examples to demonstrate the impact of f-separable distortion measures to a popular class of finite alphabet sources.

Example 1. *(Binary memoryless sources) Let the joint process* $(\mathbf{x}^n, \mathbf{z}^n)$ *form an* i.i.d *sequence of RVs such that* $\mathcal{X} = \mathcal{Z} = \widehat{\mathcal{X}} = \{0, 1\}$ *furnished with the classical single-letter Hamming distortion, i.e.,*

$$d(x, \widehat{x}) = \begin{cases} 0, & \text{if } x = \widehat{x} \\ 1 & \text{if } x \neq \widehat{x}. \end{cases} \tag{26}$$

Moreover, let $\mathbf{x}_i \sim Bernoulli(\frac{1}{2})$ *and a binary memoryless channel that induces a transition probability of the form*

$$p(z|x) = \begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}, \quad \beta \in \left[0, \frac{1}{2}\right). \tag{27}$$

Using the above input data, we obtain the following result.

Theorem 2. *(Closed-form solution) For the previous inputs and for any continuous, increasing function* $f(\cdot)$*, we obtain*

$$\mathcal{I}_{f,d}(D) = \mathcal{I}_{\tilde{d}}(f(D)) =$$
$$\left[1 - h_b\left(\frac{f(D) - (1-\beta)f(0) - \beta f(1)}{(1-\beta)f(1) + \beta f(0) - (1-\beta)f(0) - \beta f(1)}\right)\right]^+ \tag{28}$$

where $[\cdot]^+ = \max\{0, \cdot\}$, $f(D) \in \left[(1-\beta)f(0) + \beta f(1), \frac{f(0)+f(1)}{2}\right]$ *and* $h_b(\cdot)$ *denotes the binary entropy function.*

Proof: See Appendix F. ∎

In Fig. 2 we illustrate some plots of (28) for various functions $f(\cdot)$ and different distortion levels D. It should be noted that due to the nature of the indirect rate distortion problem compared to the classical rate distortion problem, there are different minimum distortion thresholds for which the curves are well-defined. In particular, when the function f is exponential, with $\beta = 0.01$ and $\rho = 9.2$, Fig. 2 demonstrates that the f-separable iRDF curve is non-convex, monotonic and well-defined for $D \in (D^{\exp}_{\min}, D^{\exp}_{\max}] = \left(\frac{1}{\rho}\log(1 - \beta + \beta\exp(\rho)), \frac{1}{\rho}\log\left(\frac{1+\exp(\rho)}{2}\right)\right]$. Similarly, if the function f is third order polynomial with $\beta = 0.15$ and $\alpha = 0.4$ or quadratic with $\beta = 0.001$, then, from Fig. 2 we observe that $\mathcal{I}_{f,d}(D)$ is again non-convex, monotonic and well-defined for $D \in \left(D^{pol}_{\min}, D^{pol}_{\max}\right] = \left(\sqrt[3]{(1-a)^3\beta - a^3(1-b)} + a, \sqrt[3]{\frac{(1-a)^3 - a^3}{2}} + a\right]$ and for $D \in (D^{qua}_{\min}, D^{qua}_{\max}] = \left(\sqrt{\beta}, \sqrt{\frac{1}{2}}\right]$, respectively. Clearly, if in Fig. 2 we consider the function f to be the identity map, then, as Fig. 2 demonstrates, we obtain $\mathcal{I}_{f,d}(D) = \mathcal{I}_{\tilde{d}}(f(D)) = \mathcal{R}_{\tilde{d}}(D)$ and the closed-form solution of (28) recovers the solution of [3, Exercise 3.8] i.e.,

$$\mathcal{I}_{f,d}(D) = \left[1 - h_b\left(\frac{D-\beta}{1-2\beta}\right)\right]^+ \text{if } D \in [\beta, \tfrac{1}{2}]. \tag{29}$$

This example aims at further emphasizing on the impact of the f-separable (non-linear) distortion constraint on the indirect rate distortion curve as opposed to the classical separable (linear) distortions for which the indirect rate-distortion curve is always convex.

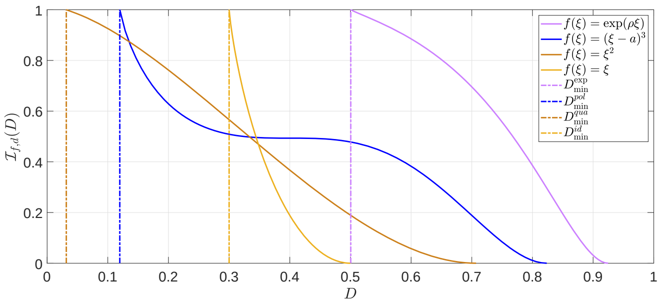

Fig. 2: Computation of $\mathcal{I}_{f,d}(D)$ for various functions $f(\cdot)$ and single-letter Hamming distance.

Special case: If in Example 1 we assume that in (27) we have $\beta = 0$, then our problem recovers the solution of [13, eq. (44)] for $\mathbf{x} \sim Bernoulli(\frac{1}{2})$.

Example 2. *Let the joint process* $(\mathbf{x}^n, \mathbf{z}^n)$ *form an* i.i.d *sequence of RVs such that* $\mathcal{X} = \widehat{\mathcal{X}} = \{0, 1\}$*,* $\mathcal{Z} = \{0, e, 1\}$ *furnished with the Hamming distortion in* (26)*. Moreover, let* $\mathbf{x}_i \sim Bernoulli(\frac{1}{2})$ *and a binary memoryless erasure channel that induces a transition probability of the form*

$$p(z|x) = \begin{bmatrix} 1-\delta & 0 \\ \delta & \delta \\ 0 & 1-\delta \end{bmatrix}, \quad \delta \in [0, 1]. \tag{30}$$

Using the above input data, we obtain the following result.

Theorem 3. *(Closed-form solution) For the previous input data, and for any continuous, increasing function* $f(\cdot)$ *we obtain*

$$\mathcal{I}_{f,d}(D) = \mathcal{I}_{\tilde{d}}(f(D)) =$$
$$\left[(1-\delta)\left(\log(2) - h_b\left(\frac{f(D) - \frac{\delta}{2}f(1) - f(0)(1 - \frac{\delta}{2})}{(1-\delta)(f(1) - f(0))}\right)\right)\right]^+ \tag{31}$$

where $f(D) \in \left[(1 - \frac{\delta}{2})f(0) + \frac{\delta}{2}f(1), \frac{f(1)+f(0)}{2}\right]$.

Proof: See Appendix G. ∎

Special case: If the chosen f-separable distortion measure is additive (function f corresponds to the identity map), then the closed-form solution of (31) recovers the solution of [8, Eq. (76)], which in turn admits the closed-form solution

$$\mathcal{I}_{f,d}(D) = \left[(1-\delta)\left(\log(2) - h_b\left(\frac{D - \frac{\delta}{2}}{1-\delta}\right)\right)\right]^+ \tag{32}$$

where $D \in \left[\frac{\delta}{2}, \frac{1}{2}\right]$.

References

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

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

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

[4] D. Sakrison, "Source encoding in the presence of random disturbance (corresp.)," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 165–167, 1968.

[5] 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.

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

[7] 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.

[8] V. Kostina and S. Verdú, "Nonasymptotic noisy lossy source coding," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6111–6123, 2016.

[9] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *IEICE TRANSACTIONS*, vol. 63, no. 10, pp. 700–706, 1980.

[10] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, 1996.

[11] Y. Oohama, "Distributed source coding of correlated Gaussian remote sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5059–5085, 2012.

[12] K. Eswaran and M. Gastpar, "Remote source coding under Gaussian noise: Dueling roles of power and entropy power," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4486–4498, 2019.

[13] Y. Shkel and S. Verdú, "A coding theorem for f-separable distortion measures," *Entropy*, vol. 20, no. 2, 2018.

[14] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, 1972.

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

[16] O. Goldreich, B. Juba, and M. Sudan, "A theory of goal-oriented communication," *Journal of ACM*, vol. 59, no. 2, pp. 1–65, May 2012.

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

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

[19] T. S. Hun, *Information spectrum methods in information theory*. Springer: Berlin, Germany, 2003.

[20] J. Kieffer, "Strong converses in source coding relative to a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 257–262, 1991.

[21] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Allerton Conference*, 2011, arXiv:1109.6310.

[22] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.

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

[24] M. Kuczma, *An Introduction to the Theory of Functional Equations and Inequalities: Cauchy's Equation and Jensen's Inequality*, 2nd ed., A. Gilányi, Ed. Birkh auser Basel, 2009.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[26] A. N. Tikhomirov, "On the notion of mean," in *Selected works of A. N. Kolmogorov*, A. N. Tikhomirov, Ed. Kluwer academic publisher, 1991, vol. 25, pp. 144–146.

APPENDIX A
ON f-SEPARABLE DISTORTIONS

We first recall the definition of f-separable distortion measure as this was introduced in [13].

Definition 5. *(f-separable distortion) [13, Definition 1] Let $f(\xi_i)$ be a continuous, increasing function on $[0,\infty)$ for any i. An n-letter distortion measure $d(x^n,\widehat{x}^n)$ is f-separable with respect to a single-letter distortion $d(x_i,\widehat{x}_i)$ if it can be written as*

$$d^n(x^n,\widehat{x}^n)=f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}f(d(x_i,\widehat{x}_i))\right). \tag{33}$$

Clearly, if in (33) we have $f(\xi_i)=\xi_i$ for any i, then we recover the classical separable distortion formulation. It should be noted that Definition 5 is inspired by the definition and properties of the so-called *quasi-arithmetic mean* (for details see Appendix H).

In what follows, we give a numerical example where we demonstrate various plots of the f-separable distortion measures computed based on the Hamming single-letter distortion versus (vs) the number of reconstruction errors. This example is similar but more detailed compared to the one given in [13, Fig. 1].

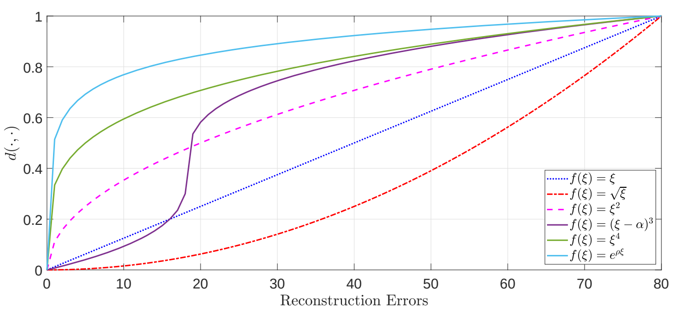

Fig. 3: f-separable distortion computed based on the Hamming single-letter distortion vs. the reconstruction errors.

Example 3. *Suppose that in (33), $d(x_i,\widehat{x}_i)$ is the classical Hamming single-letter distortion $\forall i$, i.e., $z_i=d(x_i,\widehat{x}_i)$ $\forall i$. The goal is to compute the reconstruction errors of an information source of 80 bits vs. the distortion penalty computed by f-separable distortion measure having as a benchmark the Hamming single-letter distortion that is $f(\xi_i)=\xi_i$, $\forall i$. In Fig. 3 we illustrate plots for the following functions, $f(\xi)=\xi$, $f(\xi)=\sqrt{\xi}$, $f(\xi)=\xi^2$, $f(\xi)=(\xi-\alpha)^3$, $f(\xi)=\xi^4$, and $f(\xi)=e^{\rho\xi}$ where α,ρ are some constants, which for this example are taken to be $\alpha=0.4$, $\rho=9$. We note that these plots can be computed by finding the quasi-arithmetic mean of $f(\xi)=\xi$, the quasi-arithmetic mean for functions of the form $f(\xi)=\xi^p$ and the one for functions of the form $f(\xi)=e^{\xi}$.*

A. Sub-additive distortion measures

A distinct property of the separable distortion measures discussed previously, is the sub-additivity of the distortion measures[1]. Sub-additivity property in distortion measures is crucial as it allows for the convexity of the rate-distortion function (RDF) characterization [23, Theorem 9.8.1]. Using a trivial extension of the definition of sub-additive function, see e.g., [24, Chapter 16], a distortion measure is called sub-additive if [19, Ch. 5.9]

$$d^n(x^n,\widehat{x}^n)\leq\frac{1}{n}\sum_{i=1}^{n}d(x_i,\widehat{x}_i). \tag{34}$$

A rather simple implication of Jensen's inequality, see e.g., [18, Theorem 2.6.2] reveals that the f-separable distortion measure (33) is sub-additive if $f(\cdot)$ is concave. In particular,

$$d^n(x^n,\widehat{x}^n)=f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}f(d(x_i,\widehat{x}_i))\right)\stackrel{(\star)}{\leq}\frac{1}{n}\sum_{i=1}^{n}d(x_i,\widehat{x}_i) \tag{35}$$

where $(\star)$ follows from Jensen's inequality.

APPENDIX B
PROOF OF LEMMA 1

Let $\{d^n_f(\cdot,\cdot):n=1,2,\ldots\}$ be sequences of f-separable distortions subject to the single-letter distortion $d(\cdot,\cdot)$ and let $\{\bar{d}^n(\cdot,\cdot):n=1,2,\ldots\}$ be a sequence of modified separable distortion measures such that $\bar{d}(\cdot,\cdot)=f(d(\cdot,\cdot))$. Due to the fact that $f(\cdot)$ is continuous, increasing functions at D, we obtain that for any $\gamma>0$ there exists a $\delta>0$ such that

$$f(D+\gamma)-f(D)=\delta. \tag{36}$$

The reverse follows due to the continuity of $f(\cdot)$, i.e., for any $\delta>0$ there exists a γ such that (36) is true.

Any source code (f^E_n,g^D_n) is an $(n,M_n,D+\gamma,\epsilon_n)$-lossless code under the f-separable distortion $d^n_f(\cdot,\cdot)$ if and only if (f^E_n,g^D_n) is additionally an $(n,M_n,f(D)+\delta,\epsilon_n)$-lossless code under the modified separable distortions $\bar{d}^n(\cdot,\cdot)$. In particular,

$$\begin{aligned}\epsilon_n&\geq\mathbf{P}\big[d^n_f(\mathbf{x}^n,g^D_n(f^E_n(\mathbf{z}^n)))\geq D+\gamma\big]\\&\stackrel{(a)}{=}\mathbf{P}\left[f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}f(d(\mathbf{x}_i,\widehat{\mathbf{x}}_i))\right)\geq D+\gamma\right]\\&\stackrel{(b)}{=}\mathbf{P}\left[\frac{1}{n}\sum_{i=1}^{n}f(d(\mathbf{x}_i,\widehat{\mathbf{x}}_i))\geq f(D+\gamma)\right]\\&\stackrel{(c)}{=}\mathbf{P}\big[\bar{d}^n(\mathbf{x}^n,g^D_n(f^E_n(\mathbf{z}^n)))\geq f(D)+\delta\big],\end{aligned} \tag{37}$$

where (a) follows by definition; (b) follows from the properties of function $f(\cdot)$ that hold for the inverse function $f^{-1}(\cdot)$ which always exists; (c) follows from (36) and $\widehat{\mathbf{x}}=g^D_n(f^E_n(\mathbf{z}^n))$ under the definition

$$\bar{d}^n(x^n,\widehat{x}^n)=\frac{1}{n}\sum_{i=1}^{n}f(d(x_i,\widehat{x}_i)). \tag{38}$$

Based on the above, it follows that (R,D) is excess distortion achievable with respect to the sequences of f-separable distortion measures $\{d^n_f:n=1,2,\ldots\}$ if and only if $(R,f(D))$

[1] Additive distortion measures are both sub-additive and super-additive.

is excess distortion achievable with respect to the sequences $\{\bar{d}^n : n = 1, 2, \ldots\}$. Using this observation and Definition 3 the result follows.
This completes the proof.

Appendix C
On the connection of Excess distortion and Average Distortion

In this section, we restate with a slight modification the result that exists in [13, Theorem A1]. In particular, this allows us to show that under the given condition

$$\mathcal{I}_{f,d}(D) \leq \widehat{\mathcal{I}}_{f,d}(D).$$

Lemma 2. *Suppose that the remote source $(\mathbf{x}^n, \mathbf{z}^n)$ and the sequence of distortion measures $\{d^n\}_{n=1}^{\infty}$ are such that*

$$\limsup_{n\to\infty} \sup_{(x^n, \widehat{x}^n)} d^n(x^n, \widehat{x}^n) \leq \Delta < \infty. \tag{39}$$

Then, if the rate-distortion pair (R, D) is excess distortion achievable, it is achievable under the average distortion.

Proof. Choose $\gamma > 0$. Suppose that there exists a code (f_n^E, g_n^D) with M codewords that achieves

$$\lim_{n\longrightarrow\infty} \mathbf{P}\big[d^n(\mathbf{x}^n, g_n^D(f_n^E(\mathbf{z}^n))) > D + \gamma\big] = 0. \tag{40}$$

Let us define

$$\mathbf{v}_n = d^n(\mathbf{x}^n, g_n^D(f_n^E(\mathbf{z}^n))).$$

Then, the following holds

$$\begin{aligned}\mathbf{E}\{\mathbf{v}_n\} &\leq \mathbf{E}[\mathbf{v}_n \mathbf{1}\{\mathbf{v}_n \leq D + \gamma\}] + \mathbf{E}[\mathbf{v}_n \mathbf{1}\{D + \gamma < \mathbf{v}_n\}] \\ &\leq_n (D + \gamma) + (\Delta + \gamma)\mathbf{P}[D + \gamma < \mathbf{v}_n] \\ &\leq_n (D + \gamma) + \gamma = D + 2\gamma,\end{aligned}$$

where the first inequality follow from (39) and the second inequality follows from (40). This completes the proof. □

Appendix D
Strong Converse from Dispersion Analysis

In this section we derive the strong converse for the remote rate-distortion function as a corollary of [8, Theorem 5]. We assume that the joint process (x^n, z^n) satisfies the assumptions in Section IV.

Corollary 3 (Strong Converse)**.** *Consider an arbitrary sequence of (n, M_n, D, ϵ_n)-lossy source codes for $(\mathbf{x}^n, \mathbf{z}^n)$. If*

$$\lim_{n\to\infty} \frac{1}{n} \log M_n < \mathcal{I}_d(D)$$

then

$$\lim_{n\to\infty} \epsilon_n = 1.$$

Proof. Suppose that $\lim_{n\to\infty} \frac{1}{n} \log M_n < \mathcal{I}_d(D)$ and $\lim_{n\to\infty} \epsilon_n < 1$. In particular, this implies that for n sufficiently large, $\epsilon_n < \bar{\epsilon} < 1$ for some $\bar{\epsilon}$.

Let $M^\star(n, D, \epsilon)$ be the smallest number of representation points compatible with excess distortion constraints given by (D, ϵ). Then, for n sufficiently large

$$\frac{1}{n} \log M^\star(n, D, \epsilon) \leq \frac{1}{n} \log M_n.$$

This implies that

$$\lim_{n\to\infty} \frac{1}{n} \log M^\star(n, D, \epsilon) \leq \lim_{n\to\infty} \frac{1}{n} \log M_n < \mathcal{I}_d(D).$$

However, this contradicts [8, Theorem 5] which says that

$$\lim_{n\to\infty} \frac{1}{n} M^\star(n, D, \epsilon) = \mathcal{I}_d(D).$$

This completes the proof. □

If we would like to drop or modify some of the assumptions, we may also attempt to prove the strong converse direction from [8, Corollary 1].

Under the stated assumption, the converse part of our main result should follow along the lines of [13].

Lemma 3. *Let $(\mathbf{x}, \mathbf{z})$ be a source that satisfies Assumptions* **(A1)-(A3)** *and let D be such that $D \in (D_{\min}, D_{\max}]$. Then,*

$$\mathcal{I}_{f,d}(D) \geq \mathcal{I}_{\bar{d}}(f(D)).$$

Proof. The proof follows along the lines of [13, Lemma A1]. If $\mathcal{I}_{\bar{d}}(f(D)) = 0$, there is nothing to prove. Suppose that $\mathcal{I}_{\bar{d}}(f(D)) > 0$. Assume there exists a sequence of (n, M_n, D_n)-noisy lossy codes (under f-separable distortion) with

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n < \mathcal{I}_{\bar{d}}(f(D)) \quad \text{and} \quad \limsup_{n\to\infty} D_n \leq D.$$

Since $\mathcal{I}_{\bar{d}}(f(D))$ is continuous and decreasing (for $D \in (D_{\min}, D_{\max}]$), there exists some $\gamma > 0$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n < \mathcal{I}_{\bar{d}}(f(D + \gamma)) < \mathcal{I}_{\bar{d}}(f(D)).$$

For every n, (n, M_n, D_n)-noisy lossy code (f^E, g^D) is also an $(n, M_n, D + \gamma, \epsilon_n)$-noisy lossy source code for some $\epsilon_n \in [0, 1]$ and f-separable distortion $d^n(x^n, \hat{x}^n)$. It is also an $(n, M_n, f(D + \gamma), \epsilon_n)$-noisy lossy source code with respect to a separable distortion $f(d(x, \hat{x}))$. We can therefore apply Corollary 3 to obtain

$$\lim_{n\to\infty} \epsilon_n = 1.$$

Thus,

$$\begin{aligned}D_n &\geq \mathbb{E}\big[d^n\big(\mathbf{x}^n, g^D(f^E(\mathbf{z}^n))\big)\big] \geq \epsilon_n(D + \gamma) \\ &> D + \frac{\gamma}{2}\end{aligned}$$

where the last line holds for sufficiently large n. The result follows since we obtain a contradiction.
This completes the proof. □

APPENDIX E
PROOF OF COROLLARY 2

Using KKT conditions [3], [25], we can write the constrained optimization problem in (21) via (19) as an unconstrained problem as follows

$$\mathcal{L}(s,\lambda(z),\mu(z,\widehat{x})) = \sum_{z,\widehat{x}} \log\left(\frac{p(\widehat{x}|z)}{p(\widehat{x})}\right)p(\widehat{x}|z)p(z) - s\Big(\mathbf{E}\Big[\tilde{d}(\mathbf{z},\widehat{\mathbf{x}})\Big] - f(D)\Big) - \sum_{z,\widehat{x}} \mu(z,\widehat{x})p(\widehat{x}|z) + \sum_{z}\lambda(z)\left(\sum_{\widehat{x}} p(\widehat{x}|z) - 1\right) \tag{41}$$

where $s \leq 0$ is the Lagrangian multiplier associated with the amended distortion constraint $\mathbf{E}\Big[\tilde{d}(\mathbf{z},\widehat{\mathbf{x}})\Big] \leq f(D)$, $\lambda(z) \geq 0$ is associated with the equality constraint $\sum_{\widehat{x}} p(\widehat{x}|z) = 1$, and $\mu(z,\widehat{x}) \geq 0$ is responsible for the inequality constraint $p(\widehat{x}|z) \geq 0$.
Due to the convexity of $\mathcal{L}(\cdot)$ with respect to $p(\cdot,|x)$, a necessary and sufficient condition for $p^*(\cdot,|x)$ to be the optimal minimizer is when $\frac{\partial\mathcal{L}(s,\lambda(z),\mu(z,\widehat{x}))}{\partial p(\widehat{x}|z))} = 0$ when $p^*(\cdot|z) > 0$ and $\frac{\partial\mathcal{L}(s,\lambda(z),\mu(z,\widehat{x}))}{\partial p(\widehat{x}|z))} \leq 0$ when $p^*(\cdot|z) = 0$, $\forall \widehat{x} \in \widehat{\mathcal{X}}$. Since there is nothing to prove for the latter case, we focus on the former case, for which the derivative after some algebraic manipulations on (41) gives

$$\sum_{z} p(z)\left[\log\left(\frac{p^*(\widehat{x}|z)}{p^*(\widehat{x})}\right) - s\tilde{d}(z,\widehat{x}) + \lambda^*(z)\right] = 0. \tag{42}$$

To obtain (42), we consider $\mu(z,\widehat{x}) = \mu^*(z,\widehat{x}) = 0\ \forall (z,\widehat{x}) \in \mathcal{Z}\times\widehat{\mathcal{X}}$. Moreover, in (42) we have that $\lambda(z) = \lambda^*(z) > 0$, $\forall z \in \mathcal{Z}$ because we require $\sum_{\widehat{z},\widehat{x}} p^*(\widehat{x}|z) = 1$. Applying this result in (42) and solving with respect to $p^*(\cdot|z)$ we obtain

$$p^*(\widehat{x}|z) = e^{s\tilde{d}(z,\widehat{x})-\lambda(z)}p^*(\widehat{x}). \tag{43}$$

Leveraging the fact that $\sum_{\widehat{x}} p^*(\widehat{x}|z) = 1$, we average both sides of (43) with respect to $\widehat{x} \in \widehat{\mathcal{X}}$ and solve to obtain $\lambda^*(z) > 0$, which is given by

$$\lambda^*(z) = \log\left(\sum_{\widehat{x}} e^{s\tilde{d}(z,\widehat{x})}p^*(\widehat{x})\right). \tag{44}$$

By substituting (44) in (43), we obtain the implicit expression of (24) for $s < 0$. Moreover, substituting (24) in (41) we obtain (25) provided that $\mathcal{I}_{f,d}(D) > 0$.
This completes the proof.

APPENDIX F
PROOF OF EXAMPLE 1

First note that for a given $p(x)$ and $p(z|x)$, we can compute $p(x,z) = p(z|x)p(x)$ and $p(z) = \sum_{x\in\{0,1\}} p(x,z)$. These two probability masses can lead to computing

$$p(x|z) = \frac{p(x,z)}{\sum_{x\in\{0,1\}} p(x,z)} = \begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}. \tag{45}$$

Moreover, the modified distortion $\tilde{d}(z,\widehat{x}) = \sum_{x\in\{0,1\}} p(x|z)f(d(x,\widehat{x}))$ yields

$$\tilde{d}(z,\widehat{x}) = \begin{bmatrix} (1-\beta)f(0)+\beta f(1) & (1-\beta)f(1)+\beta f(0) \\ (1-\beta)f(1)+\beta f(0) & (1-\beta)f(0)+\beta f(1) \end{bmatrix}. \tag{46}$$

Now observe that the following series of equalities hold,

$$\begin{aligned}
\mathcal{I}_{f,d}(D) &= \mathcal{I}_{\bar{d}}(f(D)) = \inf_{q(\widehat{x}|z):\ \mathbf{E}\left[\tilde{d}(\mathbf{z},\widehat{\mathbf{x}})\right]\leq f(D)} I(\mathbf{z};\widehat{\mathbf{x}}) \\
&= \inf_{\substack{q(\widehat{x}|z): \\ \frac{\mathbf{E}\left[\tilde{d}(\mathbf{z},\widehat{\mathbf{x}})\right]-(1-\beta)f(0)-\beta f(1)}{(1-\beta)f(1)+\beta f(0)-(1-\beta)f(0)-\beta f(1)} \\ \leq \frac{f(D)-(1-\beta)f(0)-\beta f(1)}{(1-\beta)f(1)+\beta f(0)-(1-\beta)f(0)-\beta f(1)}}} I(\mathbf{z};\widehat{\mathbf{x}}) \\
&= \inf_{\substack{q(\widehat{x}|z): \\ \mathbf{E}\left[\frac{\tilde{d}(\mathbf{z},\widehat{\mathbf{x}})-(1-\beta)f(0)-\beta f(1)}{(1-\beta)f(1)+\beta f(0)-(1-\beta)f(0)-\beta f(1)}\right] \\ \leq \frac{f(D)-(1-\beta)f(0)-\beta f(1)}{(1-\beta)f(1)+\beta f(0)-(1-\beta)f(0)-\beta f(1)}}} I(\mathbf{z};\widehat{\mathbf{x}}) \\
&= \inf_{\substack{q(\widehat{x}|z): \\ \mathbf{E}\left[\tilde{d}'(\mathbf{z},\widehat{\mathbf{x}})\right]\leq \frac{f(D)-(1-\beta)f(0)-\beta f(1)}{(1-\beta)f(1)+\beta f(0)-(1-\beta)f(0)-\beta f(1)}}} I(\mathbf{z};\widehat{\mathbf{x}}) \qquad (47) \\
&= \mathcal{R}_{\tilde{d}'}\left(\frac{f(D)-(1-\beta)f(0)-\beta f(1)}{(1-\beta)f(1)+\beta f(0)-(1-\beta)f(0)-\beta f(1)}\right) \qquad (48)
\end{aligned}$$

where in (47), $\tilde{d}'(\mathbf{z},\widehat{\mathbf{x}})$ denotes the expression in the argument of $\mathbf{E}[\cdot]$ and (48) is precisely the classical or "direct" f-separable RDF. The latter is the closed-form solution given in (28).
This completes the proof.

APPENDIX G
PROOF OF EXAMPLE 2

For a given $p(x)$ and $p(z|x)$, we can compute $p(x,z) = p(z|x)p(x)$ and hence $p(z) = \sum_{x\in\{0,1\}} p(x,z)$. Subsequently, we can compute

$$p(x|z) = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}. \tag{49}$$

Moreover, the modified distortion for this example $\tilde{d}(z,\widehat{x}) = \sum_{x\in\{0,1\}} p(x|z)f(d(x,\widehat{x}))$ yields

$$\tilde{d}(z,\widehat{x}) = \begin{bmatrix} f(0) & f(1) \\ \frac{1}{2}(f(0)+f(1)) & \frac{1}{2}(f(0)+f(1)) \\ f(1) & f(0) \end{bmatrix}. \tag{50}$$

To reach our result, we use Lemma 2, (24) for $p^*(\widehat{x}) = \sum_{z\in\{0,e,1\}} p^*(\widehat{x}|z)p(z)$. When applied to our example and after some algebra, (24) yields

$$p^*(\widehat{x}|z) = \begin{bmatrix} \frac{f(1)(1-\frac{\delta}{2})+\frac{\delta}{2}f(0)-f(D)}{(1-\delta)(f(1)-f(0))} & \frac{1}{2} & \frac{f(D)-(1-\frac{\delta}{2})f(0)-\frac{\delta}{2}f(1)}{(1-\delta)(f(1)-f(0))} \\ \frac{f(D)-(1-\frac{\delta}{2})f(0)-\frac{\delta}{2}f(1)}{(1-\delta)(f(1)-f(0))} & \frac{1}{2} & \frac{f(1)(1-\frac{\delta}{2})+\frac{\delta}{2}f(0)-f(D)}{(1-\delta)(f(1)-f(0))} \end{bmatrix},$$

$$p^*(\widehat{x}) = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

Using these closed-form expressions in the definition of mutual information [18], i.e., $I(\mathbf{z};\widehat{\mathbf{x}}) = H(\widehat{\mathbf{x}}) - H(\widehat{\mathbf{x}}|\mathbf{z})$, where

$H(\cdot)$ is the discrete entropy, we obtain (31).
This completes the proof.

Appendix H
Quasi-arithmetic Mean

To get a better understanding of Definition 5, we recall the general definition of the so-called *quasi-arithmetic mean* or *Kolmogorov-Nagumo mean* or *generalized f-mean*.

Definition 6. *(Quasi-arithmetic mean) [26, p. 144] Consider a continuous strictly monotone function f, an interval on the real line $\mathbf{I}$ and the set of real numbers $\mathbb{R}$. If $f : \mathbf{I} \mapsto \mathbb{R}$, then, the quasi-arithmetic mean for $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \in \boldsymbol{I}$ is defined by the function*

$$M_n(\boldsymbol{\xi}) \triangleq f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} f(\xi_i)\right). \tag{51}$$

Note that all known types of mean, e.g., arithmetic, geometric, harmonic and root-mean-square, are of the form (51).

A sequence of functions $\{M_n : \mathbb{R}^n \mapsto \mathbb{R}\}$ defines a *regular type of mean* if the following properties hold:

(i) $M_n(\boldsymbol{\xi})$ is continuous and monotonically increasing in each variable;
(ii) $M_n(\boldsymbol{\xi})$ is a symmetric function;
(ii) The mean of identical numbers is equal to their common value, namely, for $\bar{\boldsymbol{\xi}} = (\underbrace{\xi, \ldots, \xi}_{n\ elements})$, then $M_n(\bar{\boldsymbol{\xi}}) = \xi$ for any n;
(iv) A subset of values can be replaced by their mean without altering the total mean, that is,

$$\begin{aligned} &M_n(\xi_1, \ldots, \xi_n, \hat{\xi}_1, \ldots, \hat{\xi}_m) \\ &= M_{n+m}(M_n(\boldsymbol{\xi}), \ldots, M_n(\boldsymbol{\xi}), \hat{\xi}_1, \ldots, \hat{\xi}_m), \end{aligned} \tag{52}$$

for any m, n.

Next, we state the following theorem.

Theorem 4. *[26, p.144, Theorem] If conditions (i)-(iv) hold, then the mean $M_n(\boldsymbol{\xi})$ has the form of* (51) *with f being a continuous and increasing function and f^{-1} its inverse.*