Successive Coding of Correlated Sources

Jayanth Nayak and Ertem Tuncel, Member, IEEE

Abstract—The rate–distortion (RD) problem for two-layer coding of a pair \((X, Y)\) of correlated sources is considered. The first layer information enables reconstruction of \(X\) within a certain distortion \(D_X\), while reception of both layers additionally enables reconstruction of \(Y\) within distortion \(D_Y\). Although this problem is a special case of the successive refinement problem, the computation of the RD region for this scenario is nontrivial. Using a general class of outer bounds (analogous to Shannon lower bound in the classical RD theory) to the successive refinement rate–distortion region, the successive coding RD region for the case where \((X, Y)\) is a jointly Gaussian pair and the distortion measure is squared-error is explicitly characterized.

Index Terms—Rate–distortion (RD), Shannon lower bound, successive coding, successive refinement.

I. INTRODUCTION

Consider the scenario where a correlated set of sources are available at the same location. Often, one of the sources plays a more important role than the others. For example, if a fire detector simultaneously monitors temperature, humidity, and pressure, temperature is probably the most important parameter to transmit to the base station, and the other two are of secondary nature. Another example might be in video surveillance, where a single camera observes most of the field of interest, and helper cameras handle the few areas that are outside the field of view of the main camera.1 For simplicity, suppose that there are only two sources. Assuming that the secondary source is never accessed without first accessing the primary source, the optimal storage/transmission problem can be cast as a two-layer coding problem. The information in the first layer enables the reconstruction of the main source while accessing both layers additionally allows the reconstruction of the secondary source. In this paper, we analyze this two-layer coding scenario, which we term successive coding.2

Further motivation for considering this scenario comes from transmission of stereophonic audio to disparate receivers. The receivers that are connected by a low-capacity link are content to receive one of the audio channels, while the better receivers require both channels to be reproduced. The natural way to accomplish this is to do successive coding. Successive coding similarly models the transmission of stereo video to disparate receivers when both views are available at a single node.

Successive coding is ultimately related to the famous scenario known in the literature as successive refinement [4], [5], [7], where a source is encoded in two layers while imposing a constraint on the distortion levels resulting from reconstructing the source using either only the first-layer information or information from both layers. In Section II-A, we shall show that the successive coding problem is in fact a special case of the successive refinement problem, and hence its rate–distortion (RD) region is completely characterized. The algorithms in [8] can thus be used to numerically compute the successive coding RD region. However, these numerical methods are computationally intensive, especially when the source alphabet is continuous. On the other hand, analytical evaluation based on Kuhn–Tucker (KT) optimality conditions derived in [8] also seem unlikely due to the complexity of the conditions. In Section III, we shall derive a class of analytically computable outer bounds for the successive refinement RD region using the dual of its associated optimization problem. By translating these results to the successive coding problem, for the squared-error distortion case, we obtain analogues of the Shannon lower bound for single-source coding [1]. The class of outer bounds is guaranteed to contain elements that are tight. In fact, for a pair of jointly Gaussian sources, we shall present an analytical characterization of the successive coding RD region for the squared-error distortion measure in Section IV.

Successive coding is also related to the robust descriptions scenario [3], where a source is encoded by a single encoder and decoded by several receivers. Each receiver imposes a distortion constraint on its reconstruction with respect to an associated distortion measure. The following special case was considered in [10], where the problem was called compression with individual distortion criteria: the source is a pair \((X, Y)\) of jointly Gaussian random variables and of the two decoders, one is interested in reconstructing \(X\), while the other is interested in reconstructing \(Y\). The problem in [10] is also a special case of the successive coding scenario since if we constrain the second-layer rate to be zero in the successive coding problem, it becomes identical to compression with individual distortion criteria.

In this paper, we begin with a formal definition of the successive coding problem in Section II. The relation of this problem to successive refinement will be discussed in Section II-A. Section II-B deals with some definitions and properties of the successive refinement RD function, which will be used in Section III to derive a class of outer bounds for general sources. Our main results are presented, as mentioned earlier, in Sections III

1We are not interested in distributed scenarios where each camera communicates independently with a central receiver. Rather, consider the case where these cameras are connected through a wired network and all views are available at the same location before storage/transmission takes place.

2This successive coding problem is distinct from what was called sequential coding in [9]. In [9], only the first source is available while encoding the first layer, whereas in our scenario, both sources are available.
and IV. We shall conclude in Section V with a summary and discussion of our results.

II. PRELIMINARIES

Let a stationary and memoryless pair of correlated sources produce the sequences \( \{X_t\}_{t=1}^{\infty} \) and \( \{Y_t\}_{t=1}^{\infty} \) according to the joint probability distribution \( p_{XY}(x, y) \), where each \( (X_t, Y_t) \in \mathcal{X} \times \mathcal{Y} \) and for any \( n \geq 1 \),

\[
    p_{XY}(x^n, y^n) = \prod_{t=1}^{n} p_{XY}(x_t, y_t).
\]

In the rest of the paper, we denote the probability distribution of any random variable \( V \) by \( p_V(v) \) and drop the subscript whenever the random variable being referred to is apparent from the argument of the distribution function.

Consider the communication scenario shown in Fig. 1. The sender has access to both sources, and for each source block of length \( n \), produces two layers of encoded bitstreams. Receiver 1 is interested in reconstructing only \( \{X_t\}_{t=1}^{\infty} \) and has at its disposal only the first layer of bits for this purpose. On the other hand, Receiver 2 not only copies this reconstruction of \( \{X_t\}_{t=1}^{\infty} \) using the same set of bits as Receiver 1, but also desires to reconstruct \( \{Y_t\}_{t=1}^{\infty} \) using the entire bit stream. The two reconstructions, denoted \( \{\hat{X}_t\}_{t=1}^{\infty} \) and \( \{\hat{Y}_t\}_{t=1}^{\infty} \), are sequences of symbols in alphabets \( \hat{\mathcal{X}} \) and \( \hat{\mathcal{Y}} \), respectively. Let \( d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to [0, \infty) \) and \( d_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \to [0, \infty) \) be single-letter distortion measures with

\[
    d_{\mathcal{X}}(X^n, \hat{X}^n) \leq \frac{1}{n} \sum_{t=1}^{n} d_{\mathcal{X}}(x_t, \hat{x}_t) \quad \text{and} \quad d_{\mathcal{Y}}(Y^n, \hat{Y}^n) \leq \frac{1}{n} \sum_{t=1}^{n} d_{\mathcal{Y}}(y_t, \hat{y}_t).
\]

We refer to this communication scheme as successive coding of correlated sources.

Definition 1: A quadruple \((R_1, R_2, D_X, D_Y)\) is an achievable RD vector for successive coding of correlated sources if there exists encoders and decoders

\[
    f_1 : \mathcal{X}^n \times \mathcal{Y}^n \to \{1, 2, \ldots, M_1\} \\
    f_2 : \mathcal{X}^n \times \mathcal{Y}^n \to \{1, 2, \ldots, M_2\} \\
    g_1 : \{1, 2, \ldots, M_1\} \to \hat{\mathcal{X}}^n \\
    g_2 : \{1, 2, \ldots, M_1\} \times \{1, 2, \ldots, M_2\} \to \hat{\mathcal{Y}}^n
\]

such that\(^3\)

\[
    \frac{1}{n} \log M_1 \leq R_1 + \epsilon \\
    \frac{1}{n} \log M_2 \leq R_2 + \epsilon \\
    E\{d_{\mathcal{X}}(X^n, g_1(f_1(x^n, y^n)))\} \leq D_X + \epsilon \\
    E\{d_{\mathcal{Y}}(Y^n, g_2(f_1(x^n, y^n), f_2(x^n, y^n)))\} \leq D_Y + \epsilon
\]

for arbitrary \( \epsilon > 0 \) and large enough \( n \).

\(^3\)All logarithms are natural for convenience.

A. SUCCESSIVE CODING AND THE SUCCESSIVE REFINEMENT PROBLEM

We now show that successive coding of correlated sources is a special case of successive refinement of a single source. To see this, consider first the successive refinement scheme shown in Fig. 2. Here, there is a single memoryless source \( \{Z_t\}_{t=1}^{\infty} \) with two reconstructions \( \{\hat{Z}_1\}_{t=1}^{\infty} \) and \( \{\hat{Z}_2\}_{t=1}^{\infty} \), where each \((z_t, \hat{z}_1, \hat{z}_2) \in Z \times \hat{\mathcal{Z}}_1 \times \hat{\mathcal{Z}}_2\). Let

\[
    d_i : Z \times \hat{\mathcal{Z}}_i \to [0, \infty)
\]

and

\[
    d_i(Z^n, \hat{Z}^n_i) \leq \frac{1}{n} \sum_{t=1}^{n} d_i(z_t, \hat{z}_i)
\]

for \( i = 1, 2 \). We can then translate the original problem described above into this setting using

\[
    Z_t = (X_t, Y_t) \\
    \hat{Z}_1 = \hat{X}_t \\
    \hat{Z}_2 = \hat{Y}_t \\
    d_1(z_t, \hat{z}_1) = d_X(x_t, \hat{x}_t) \\
    d_2(z_t, \hat{z}_2) = d_Y(y_t, \hat{y}_t).
\]

Even though \( d_1 \) and \( d_2 \) are possibly different (and incompatible) distortion measures, this scenario can still be considered in the "refinement" framework,\(^4\) and the well-known necessary and sufficient conditions for achievability [5], [7], demanding the existence of \((\hat{Z}_1, \hat{Z}_2)\) jointly distributed with \( Z \) such that

\[
    I(Z; \hat{Z}_1) \leq R_1 \\
    I(Z; \hat{Z}_2) \leq R_1 + R_2 \\
    E\{d_1(Z, \hat{Z}_1)\} \leq D_1 \\
    E\{d_2(Z, \hat{Z}_2)\} \leq D_2
\]

\(^4\)None of the achievability and converse results in [5], [7] dictate \( d_1 = d_2 \).
remain valid. Translating (1)–(4) into our setting, we obtain the following lemma.

**Lemma 1:** A quadruple \((R_1, R_2, D_X, D_Y)\) is an achievable RD vector for successive coding of correlated sources if and only if there exist \(\hat{X}\) and \(\hat{Y}\) jointly distributed with \((X, Y)\) such that

\[
I(X, Y; \hat{X}) \leq R_1 \tag{5}
\]
\[
I(X, Y; \hat{X}, \hat{Y}) \leq R_1 + R_2 \tag{6}
\]
\[
E[d_X(X, \hat{X})] \leq D_X \tag{7}
\]
\[
E[d_Y(Y, \hat{Y})] \leq D_Y. \tag{8}
\]

Even though this translation trivializes the characteristic of the achievable RD region, the computation of the region constitutes a major and non-straightforward problem even for very simple cases such as jointly Gaussian sources with squared-error distortion.

### B. A Lagrangian Formulation of the Successive Refinement RD Problem

We repeat here known results in successive refinement [8]. To compute the rate and distortion levels satisfying (1)–(4) for some \((\hat{Z}_1, \hat{Z}_2)\) jointly distributed with \(Z\), one could use the Lagrangian approach and minimize

\[
L_{\alpha, \beta_1, \beta_2}(p_{\hat{Z}_1, \hat{Z}_2} | \hat{Z}) = \alpha I(Z; \hat{Z}_1) + (1 - \alpha) I(Z; \hat{Z}_1, \hat{Z}_2) + \beta_1 E[d_1(Z; \hat{Z}_1)] + \beta_2 E[d_2(Z; \hat{Z}_2)]. \tag{9}
\]

for all \(\beta_1, \beta_2 \geq 0\) and \(0 \leq \alpha \leq 1\). Denoting by \(L_{\alpha, \beta_1, \beta_2}^*\) the minimum Lagrangian, i.e.,

\[
L_{\alpha, \beta_1, \beta_2}^* = \min_{p_{\hat{Z}_1, \hat{Z}_2} | \hat{Z}} L_{\alpha, \beta_1, \beta_2}(p_{\hat{Z}_1, \hat{Z}_2} | \hat{Z}) \tag{10}
\]

and defining the RD function \(R_{\alpha}(D_1, D_2)\) as the tradeoff between a weighted average of rates and individual distortion levels at each layer, i.e.,

\[
R_{\alpha}(D_1, D_2) \triangleq \min_{p_{\hat{Z}_1, \hat{Z}_2} | \hat{Z}} \{ \alpha I(Z; \hat{Z}_1) + (1 - \alpha) I(Z; \hat{Z}_1, \hat{Z}_2) \}
\]

we have

\[
R_{\alpha}(D_1, D_2) = \max_{\beta_1, \beta_2 \geq 0} \{ L_{\alpha, \beta_1, \beta_2}^* - \beta_1 D_1 - \beta_2 D_2 \}. \tag{11}
\]

Alternatively, if we define

\[
D(\alpha, \beta_1, \beta_2) \triangleq \left\{ \left( E[d_1(Z; \hat{Z}_1)], E[d_2(Z; \hat{Z}_2)^*] \right) : \hat{Z}_1^* \hat{Z}_2^* \right\}
\]

then for any \((D_1, D_2) \in D(\alpha, \beta_1, \beta_2)\)

\[
R_{\alpha}(D_1, D_2) = L_{\alpha, \beta_1, \beta_2}^* - \beta_1 D_1 - \beta_2 D_2. \tag{12}
\]

We shall only concern ourselves with computing or bounding from below \(R_{\alpha}(D_1, D_2)\).

For certain extreme values of the Lagrangian parameters, the successive refinement RD function has a simple form related to \(R_{\alpha}^i(D_i)\), \(i = 1, 2\), where \(R_{\alpha}^i(D_i)\) is the RD function of \(Z\) with respect to the distortion measure \(d_i\) at level \(D_i\). To fully exploit this simplicity, we consider the single-layer Lagrangian minimization

\[
L_{\alpha} = \min_{p_{\hat{Z}_i} | \hat{Z}} \{ I(Z; \hat{Z}_i) + \gamma E[d_i(Z; \hat{Z}_i)] \} \tag{13}
\]

and define

\[
D(\alpha) \triangleq \left\{ E[d_i(Z; \hat{Z}_i)^*] : \hat{Z}_i^* \right\} \text{ achieves } L_{\alpha}^* \}
\]

The following are the extreme cases that simplify the analysis.

1) \(\alpha = 1\): There is no constraint on the second layer rate. Therefore, an optimal strategy would be to encode \(Z\) purely to satisfy the first-layer constraint leading to a first-layer rate of \(R_1^0(D_1)\). Since the second-layer rate can be arbitrarily high without changing the RD cost, we can transmit as many bits as are necessary to satisfy the second-layer constraint. Thus

\[
R_1(D_1, D_2) = R_1^0(D_1) \tag{14}
\]

2) \(\alpha = 0\): \(\beta_1 = 0, \beta_2 < 1\): The constraint on the first layer reconstruction drops out. It can be easily seen that for \(p_{\hat{Z}_1, \hat{Z}_2} | \hat{Z}\) achieving \(L_{\alpha, \beta_1, \beta_2}^*\), \(Z - \hat{Z}_2 \perp \hat{Z}_1\) forms a Markov chain. Additionally, if \(\alpha > 0\), the cost function is minimized only when \(I(Z; \hat{Z}_1) = 0\). Thus, defining \(D_{\alpha, 0, \beta_2}^* \triangleq \min_{p_{\hat{Z}_1, \hat{Z}_2} | \hat{Z}} E[d_1(Z; \hat{Z}_1)]\)

\[
D(\alpha, 0, \beta_2) = \left\{ (D_1, D_2) : D_1 \geq D_{\alpha, 0, \beta_2}^*, D_2 \in D_2(\beta_2) \right\}
\]

where

\[
D_{\alpha, 0, \beta_2}^* = \min_{(\hat{Z}_1, \hat{Z}_2) : I(Z; \hat{Z}_1) = 0} E[d_1(Z; \hat{Z}_1)]. \tag{15}
\]

It is easy to see that \(D_{\alpha}^*(D_1) \leq D_{\alpha, 0}^* \leq D_{\alpha, 0, \beta_2}^*\). Regardless of whether \(\alpha = 0\), for any \((D_1, D_2) \in D(\alpha, 0, \beta_2)\)

\[
R_{\alpha}(D_1, D_2) = (1 - \alpha)R_1^0(D_2). \tag{16}
\]

3) \(\beta_1 = 0, \beta_2 < 1\): The constraint on the second-layer reconstruction drops out. The optimal strategy is to use random variables \((\hat{Z}_1^*, \hat{Z}_2^*)\) satisfying the Markov condition \(Z - \hat{Z}_1 \perp \hat{Z}_2\). Define

\[
D_{\beta_1}^*(D_1) \triangleq \min_{(\hat{Z}_1, \hat{Z}_2) : \beta_1 R_1^0(D_1)} E[d_2(Z; \hat{Z}_2)]. \tag{17}
\]

As in the \(\beta_1 = 0\) case

\[
D_{\beta_1}^*(D_1) \leq D_{\beta_1}^* \triangleq \min_{\hat{Z}_2} E[d_2(Z; \hat{Z}_2)]. \tag{18}
\]

It can easily be seen that

\[
D(\alpha, \beta_1, 0) = \left\{ (D_1, D_2) : D_1 \in D(\beta_1), D_2 \geq D_{\beta_1}^*(D_1) \right\}
\]

and for any \((D_1, D_2) \in D(\alpha, \beta_1, 0)\)

\[
R_{\alpha}(D_1, D_2) = R_{\beta_1}^*(D_1). \tag{19}
\]
Therefore, we will focus on the nontrivial cases, where 0 ≤ α < 1, β1 > 0, and β2 > 0.

There is one more important case where the analysis is straightforward, namely, when the source Z is successively refinable [4]. Using I(Z; Z1, Z2) ≥ I(Z; Z2) with equality if and
only if Z = Z2 > Z1, one can conclude that

\[ L_{α,β_1}^* ≥ αL_{α,β_2}^* \]

with equality if and only if Z2 = Z and achieves L* for all Z2i, Z2\j achieving L* for all Z1\j, respectively, can be combined to form a Markov chain Z = Z2 i = Z1, i.e., the source is successively refinable from D1 ∈ D1(β1/α) to D2 ∈ D2(β2/(1 − α)). In that case

\[ R_0(D1, D2) = αR_2^*(D1) + (1 - α)R_2^*(D2). \]

The following lemma, which is key to analyzing the nontrivial cases was proven in [8].

**Lemma 2 (KT Conditions):** A given p(2|1, Z) achieves L^* if and only if there exist “artificial” conditionals q(2|1, Z) for every Z1 ∈ Z1 with p(2|1) = 0 such that

\[ v_2(\hat{Z}, \hat{Z}) ≤ v_1(\hat{Z}) ≤ 1 \]

for all \( Z1, Z2 ∈ Z2 \), where

\[ v_1(\hat{Z}) = \sum_{z_1} \left[ \frac{p(z)}{f_0(z)} \right]^{\frac{1}{\alpha}} e^{-β_1 d_1(\hat{Z}, z_1)} \]

and

\[ v_2(\hat{Z}, \hat{Z}) = \sum_{z_1} \left[ \frac{p(z)}{f_0(z)} \right]^{\frac{1}{\alpha}} e^{-β_1 d_1(\hat{Z}, z_1) - β_2 d_2(\hat{Z}, z_1)} \]

with \( β_2 = \frac{1}{1−α} \). Here

\[ f_1(\hat{Z}, z_1) = \left\{ \begin{array}{ll}
\sum_{z_2} p(z_2 | z_1) e^{-β_2 d_2(\hat{Z}, z_2)}, & p(\hat{Z}) > 0 \\
\sum_{z_2} q(\hat{Z}, z_1) e^{-β_2 d_2(\hat{Z}, z_2)}, & p(\hat{Z}) = 0
\end{array} \right. \]

and

\[ f_0(\hat{Z}) = \sum_{z_1, z_2} p(z_1, z_2) f_1(\hat{Z}, z_1) e^{-β_2 d_2(\hat{Z}, z_2)}. \]

We proceed towards that end as follows.

\[ L_{α,β_1}^* p(\hat{Z}, \hat{Z}) | Z = -\sum_{z} p(z) \log f_0(z). \]

Observe the difficulty of actually using Lemma 2 for an analytical computation of L^*. There is in fact no known nontrivial case where a solution p(\hat{Z}, \hat{Z}) | Z is "guessed" and tested for optimality using Lemma 2. In [8], it was used to verify (near) optimality of a numerically computed solution.

Also observe that if \( β_2 = 0, v_1(\hat{Z}) = v_2(\hat{Z}, \hat{Z}) \), and \( v_1(\hat{Z}) ≤ 1 \) coincides exactly with the KT condition given in [1]. We refer the reader to [8, Sec. IV] for an interpretation of f0 and f1 as the partition functions in the statistical physics analogy.

**III. A CLASS OF OUTER BOUNDS TO THE SUCCESSIVE REFINEMENT RD REGION**

We prove in the following theorem an alternative characterization for L* which will be useful in deriving lower bounds to it.

**Theorem 1:** Define

\[ Λ = \{ (μ(z), η(z, z_1)) : ∀z_2 ∈ Z, z_1 ∈ Z, z_2 ∈ Z_2, μ(z) > 0, η(z, z_1) > 0, w_2(z_1, z_2) ≤ w_1(z_1) ≤ 1 \} \]

where

\[ w_1(z_1) = \sum_{z} \left[ \frac{p(z)η(z, z_1)}{μ(z)} \right]^{\frac{1}{α}} e^{-β_1 d_1(z, z_1)} \]

\[ w_2(z_1, z_2) = \sum_{z} \left[ \frac{p(z)η(z, z_1)}{μ(z)} \right]^{\frac{1}{α}} e^{-β_1 d_1(z, z_1) - β_2 d_2(z, z_2)}. \]

Then

\[ L_{α,β_1}^* = \max_{(μ(z), η(z, z_1)) ∈ Λ} -\sum_{z} p(z) \log μ(z). \]

**Remark 1:** Although \( μ(z) \) and \( η(z, z_1) \) apparently play the same role as \( f_0(z) \) and \( f_1(z, z_1) \) in Lemma 2, respectively, they do not necessarily correspond to any \( p(z, z_2) \) or \( q(z_2 | z_1) \) as in (19) and (20). Rather, \( μ(z) \) and \( η(z, z_1) \) are considered free functions to optimize over.

**Proof:** It should be clear that \( Λ ≠ ∅ \), because it follows from Lemma 2 that for the optimal \( p(z_1, z_2) \) and accompanying \( q(z_2 | z_1) \), \( f_0(z) \) and \( f_1(z, z_1) \) given by (19) and (20) satisfy (16). Thus, the choice \( μ(z) = f_0(z) \) and \( η(z, z_1) = f_1(z, z_1) \) satisfies \( w_2(z_1, z_2) ≤ w_1(z_1) ≤ 1 \). This fact and (21) imply that it suffices to prove that for any \( p(z_1, z_2) | z \) and \( (μ(z), η(z, z_1)) ∈ Λ \), we have

\[ L_{α,β_1}^* p(z_1, z_2) | Z ≥ -\sum_{z} p(z) \log μ(z). \]

We proceed towards that end as follows.

\[ L_{α,β_1}^* p(z_1, z_2) | Z + \sum_{z} p(z) \log μ(z) \]

\[ = \sum_{z, z_1, z_2} p(z) p(z_1, z_2 | z) \left[ β_1 d_1(z, z_1) + β_2 d_2(z, z_2) + \log μ(z) + α \log \left( \frac{p(z_1 | z)}{μ(z_1)} \right) + (1 - α) \log \left( \frac{p(z_1, z_2 | z)}{μ(z_1, z_2)} \right) \right] \]

\[ = \sum_{z, z_1, z_2} p(z) p(z_1, z_2 | z) \left[ α \log \left( \frac{p(z_1 | z)}{μ(z_1)} e^{-β_1 d_1(z, z_1)} \right) + (1 - α) \log \left( \frac{p(z_1, z_2 | z)}{μ(z_1, z_2)} e^{-β_1 d_1(z, z_1)} \right) \right] \]

\[ = \sum_{z, z_1, z_2} p(z) p(z_1, z_2 | z) \left[ α \log \left( \frac{p(z_1 | z)}{μ(z_1)} e^{-β_1 d_1(z, z_1)} \right) + (1 - α) \log \left( \frac{p(z_1, z_2 | z)}{μ(z_1, z_2)} e^{-β_1 d_1(z, z_1)} \right) \right] \]
≥ \sum_{z, z_1, z_2} p(z)p(\hat{z}_1, \hat{z}_2 | z) \\
\cdot \left[ \alpha \left( 1 - \frac{p(\hat{z}_1 | z)\eta(z, \hat{z}_1)\alpha^{-1}}{p(\hat{z}_2 | z, \hat{z}_1)\mu(z)\eta(z, \hat{z}_1)^\alpha} \right) + (1 - \alpha) \left( 1 - \frac{p(\hat{z}_1, \hat{z}_2 | z)\eta(z, \hat{z}_1, \hat{z}_2)\mu(z)\eta(z, \hat{z}_1)^\alpha}{p(\hat{z}_1, \hat{z}_2 | z)\mu(z)\eta(z, \hat{z}_1)^\alpha} \right) \right] \\
= \alpha \left( 1 - \sum_{\hat{z}_1} p(\hat{z}_1)w_1(\hat{z}_1) \right) \\
+ (1 - \alpha) \left( 1 - \sum_{\hat{z}_1, \hat{z}_2} p(\hat{z}_1, \hat{z}_2)w_2(\hat{z}_1, \hat{z}_2) \right) \\
\geq 0 \quad (26)

where (25) follows from the standard inequality $t - 1 \geq \log t$. The proof is therefore complete. \qed

**Corollary 1:** A pair $(\mu(z), \eta(z, \hat{z}_1)) \in \mathcal{A}$ achieves $I_{\alpha, \beta_1, \beta_2}^*$, i.e.,

$$I_{\alpha, \beta_1, \beta_2}^* = -\sum_x p(x)\log \mu(x) \quad (27)$$

if and only if there exists a marginal $p(\hat{z}_1, \hat{z}_2)$ such that

$$\eta(z, \hat{z}_1) = \sum_{\hat{z}_2} p(\hat{z}_2 | \hat{z}_1) e^{-\beta_2 d_2(z, \hat{z}_2)} \quad (28)$$

for all $\hat{z}_1$ with $p(\hat{z}_1) > 0$, and

$$\mu(z) = \sum_{\hat{z}_1, \hat{z}_2} p(\hat{z}_1, \hat{z}_2)\eta(z, \hat{z}_1) e^{-\beta_1 d_1(z, \hat{z}_1) - \beta_2 d_2(z, \hat{z}_2)} \quad (29)$$

or equivalently

$$\mu(z) = \sum_{\hat{z}_1} p(\hat{z}_1)\eta(z, \hat{z}_1) e^{-\beta_1 d_1(z, \hat{z}_1)} \quad (30)$$

**Proof:** Implicit in the Proof of Theorem 1 are the necessary and sufficient conditions for (27). Specifically, note that since $t - 1 = \log t$ only when $t = 1$, (25) holds if and only if there exists $p(\hat{z}_1, \hat{z}_2 | z)$ satisfying

$$p(\hat{z}_1 | z) = \frac{p(\hat{z}_1)\eta(z, \hat{z}_1) e^{-\beta_1 d_1(z, \hat{z}_1)}}{\mu(z)} \quad (31)$$

and

$$p(\hat{z}_1, \hat{z}_2 | z) = \frac{p(\hat{z}_1, \hat{z}_2)\eta(z, \hat{z}_1) e^{-\beta_1 d_1(z, \hat{z}_1) - \beta_2 d_2(z, \hat{z}_2)}}{\mu(z)} \quad (32)$$

Interestingly, (31) and (32) also imply equality in (26), and are therefore necessary and sufficient for (27). That can be seen by multiplying both sides of (31) and (32) by $p(z)$ and summing over $z, \hat{z}_1, \hat{z}_2$.

Since $p(\hat{z}_1 | z) = \sum_{\hat{z}_2} p(\hat{z}_1, \hat{z}_2 | z)$, (31) and (32) imply (28). Also, summing (32) over both $\hat{z}_1$ and $\hat{z}_2$ yields (29).

Conversely, let $(\mu(z), \eta(z, \hat{z}_1)) \in \mathcal{A}$ satisfy (28) and (29) for some marginal $p(\hat{z}_1, \hat{z}_2)$. It suffices to show that (31) and (32) can be simultaneously satisfied without causing statistical inconsistency. Since (28) and (32) imply (31), this reduces to showing that choosing $p(\hat{z}_1, \hat{z}_2 | z)$ as in (32) is consistent with $p(\hat{z}_1, \hat{z}_2)$. To that end, multiply both sides of (32) with $p(z)$ and sum over $z$ to obtain

$$p(\hat{z}_1, \hat{z}_2) = p(\hat{z}_1, \hat{z}_2) \sum_z p(z) \eta(z, \hat{z}_1) e^{-\beta_1 d_1(z, \hat{z}_1) - \beta_2 d_2(z, \hat{z}_2)} \mu(z)$$

$$= p(\hat{z}_1, \hat{z}_2) \frac{\mu(z)}{w_2(\hat{z}_1, \hat{z}_2)}$$

We must then show $w_2(\hat{z}_1, \hat{z}_2) = 1$ whenever $p(\hat{z}_1, \hat{z}_2) > 0$. Now consider

$$\sum_{\hat{z}_1, \hat{z}_2} p(\hat{z}_1, \hat{z}_2)w_2(\hat{z}_1, \hat{z}_2)$$

$$= \sum_z p(z) \sum_{\hat{z}_1, \hat{z}_2} p(\hat{z}_1, \hat{z}_2)\eta(z, \hat{z}_1) e^{-\beta_1 d_1(z, \hat{z}_1) - \beta_2 d_2(z, \hat{z}_2)} \frac{\mu(z)}{w_2(\hat{z}_1, \hat{z}_2)}$$

$$= \sum_z p(z)$$

$$= 1.$$

Since $w_2(\hat{z}_1, \hat{z}_2) \leq 1$, this implies the desired consistency, and finishes the proof. \qed

Theorem 1 can be generalized to sources with continuous alphabets in a straightforward manner applying techniques in [1] based on variational calculus. This action will be justified thanks to the generalization of the achievability region for successive refinement to continuous alphabets given in [2].

In the next section, we utilize Theorem 1 in the context of successive coding of correlated sources with squared-error distortion measures, and derive outer bounds. We also investigate tightness of these bounds using Corollary 1. To emphasize the successive coding setup, we shall use $\mu(x, y)$ and $\eta(x, y, \hat{x})$ to denote the functions $\mu(z)$ and $\eta(z, \hat{z}_1)$. Translating the conditions for $(\mu, \eta)$ to belong to $\mathcal{A}$ for continuous sources and for the successive coding problem with squared-error distortion, we get

$$1 \geq \int \int \frac{p(x, y)\eta(x, y, \hat{x}) e^{-\beta_1 (x - \hat{x})^2}}{\mu(x, y)} dx dy$$

$$\geq \int \int \frac{p(x, y)\eta(x, y, \hat{x}) e^{-\beta_1 (x - \hat{x})^2 - \beta_2 (y - \hat{y})^2}}{\mu(x, y)} dx dy \quad (33)$$

IV. SUCCESSIVE CODING RD FUNCTION FOR SQUARED-ERROR DISTORTION MEASURES

The squared-error distortion measure is one of the most widely used in practical scenarios. In this section, we shall derive a class of lower bounds on the successive coding RD function for this distortion measure. We shall present an analytical expression for the RD function when the source is a pair of jointly Gaussian random variables. Finally, we derive bounds on the RD function for general sources based on the quadratic Gaussian RD function.

Throughout this section, we shall let $X = \hat{X} = \hat{Y} = (X, Y, \hat{X}) = (x, y, \hat{x})$, $d_X(\hat{x}, x) = (x - \hat{x})^2$, and $d_Y(\hat{y}, y) = (y - \hat{y})^2$. Let $\mathbf{Z} = [X \ Y]^T$ denote a zero-mean random vector. We shall assume that the distribution of $\mathbf{Z}$ is absolutely continuous and that the support of its density $p_\mathbf{Z}(\mathbf{z})$ is the entire real plane. Since scaling one of the sources does not change the RD function if we also scale the distortion level corresponding to that source
in an appropriate manner, we assume for simplicity that $\sigma_X^2 = \sigma_Y^2 = 1$ and $E\{XY\} \geq 0$. Therefore, the covariance matrix of the source can be written as

$$C_{Z} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$ (34)\]

where $0 \leq \rho < 1$ is the correlation coefficient.5.

For the source $Z$ and given $0 \leq \alpha \leq 1$, we are interested in computing $R_\alpha(D_X, D_Y)$. We first specialize the results in Section II-B regarding the extremal values of the Lagrangian parameters to squared-error distortion. Towards that end, we denote the single-layer RD functions of $X$ and $Y$ by $R_X(D_X)$ and $R_Y(D_Y)$, respectively. We shall also denote $D^!(D_2)$ and $D^*(D_2)$ by $D_X^!(D_Y)$ and $D^*_X(D_Y)$, respectively. These functions specialize to

$$D^!(D_Y) = \min_{(X, \hat{X})} E\{(X - \hat{X})^2\}$$

and

$$D^*(D_Y) = \min_{(X, \hat{X})} E\{(Y - \hat{Y})^2\}$$

Finally, we have $D^!_{\alpha max} \triangleq D^!_{\alpha max} = 1$ and $D^*_\alpha max \triangleq D^*_\alpha max = 1$. Thus, computing $R_\alpha(D_X, D_Y)$ for $0 \leq D_X \leq 1, 0 \leq D_Y \leq 1$ suffices.

Using results from Section II-B, we then have

$$R_\alpha(D_X, D_Y) = R_X(D_X)$$

when either $\alpha = 1$ or $D^!(D_X) \leq D_Y \leq 1$, and

$$R_\alpha(D_X, D_Y) = (1 - \alpha)R_Y(D_Y)$$

when either $0 < \alpha < 1$ and $D_X = 1$, or $\alpha = 0$ and $D^!(D_Y) \leq D_X \leq 1$.

We also ignore the cases where either $D_X$ or $D_Y$ are $0$ since such a constraint would result in the RD function being infinite. For $0 \leq \alpha < 1$, we denote by $D^{NZ}_\alpha$ the set of distortion pairs not covered by the earlier cases, that is

$$D^{NZ}_\alpha = \{(D_X, D_Y) : 0 < D_Y \leq D^!(D_X), 0 < D_X < D^*(D_Y) \text{ if } \alpha = 0, \text{ or } 0 < D_X < 1 \text{ if } \alpha > 0\}.$$

We note in passing that the source pair $(X, Y)$ is successively refinable in the classical sense of [4] only in trivial cases. That is because successive refinability requires

$$\hat{X}^* \leftrightarrow X \leftrightarrow Y$$ (37)

$$X \leftrightarrow Y \leftrightarrow \hat{Y}^*$$ (38)

$$(X, Y) \leftrightarrow \hat{Y}^* \leftrightarrow \hat{X}^*$$ (39)

where $\hat{X}^*$ and $\hat{Y}^*$ achieve $R_X(D_X)$ and $R_Y(D_Y)$, respectively. The Markov chain (39) implies for all $(x, y)$ with $p_{XY}(x, y) > 0$ that

$$p_{X \mid 1, XY}(x \mid x, y) \leq p_Y \cdot p_{X \mid Y}(x \mid y) \leq p_{X \mid Y}(x \mid y)$$

Using (37) and (38), (40) can be equivalently stated as

$$p_{X \mid 1, \hat{X}^*}(x \mid x, y) = \int p_Y \cdot p_{X \mid Y}(x \mid y) \, d\hat{y}$$

for all $(x, y)$. Since the right-hand side of (41) is independent of $x$, this implies independence of $X$ and $\hat{X}^*$. That, in turn, is possible if and only if $R_X(D_X) = 0$.

Therefore, the notion of successive refinability does not simplify our analysis, even for jointly Gaussian sources.

A. Lower Bounds on the RD Function for General Sources

In general, for cases other than those discussed earlier, that is when $(D_X, D_Y) \in D^{NZ}_\alpha$, solving the optimization problem for $R_\alpha(D_X, D_Y)$ might be hard. We therefore consider easily computable bounds on this quantity. We use the alternative characterization of the RD function from (11) and Theorem 1 to obtain lower bounds. Given $0 \leq \alpha < 1, \beta_1 > 0$ and $\beta_2 > 0$, let

$$\eta(x, y, \delta) = K_{\beta_2} e^{-\gamma(y - \nu \delta)^2}$$ (43)

for some $0 \leq \gamma \leq \beta_2, K_0 > 0, K_1 > 0$, $\nu$, and

$$B = b \begin{bmatrix} \nu^2 & -\nu \\ -\nu & 1 \end{bmatrix}$$

with $b \geq 0$. Not all such pairs $(\mu, \eta)$ lead to lower bounds on the RD function. More specifically, from (33), we need

$$1 \geq K_{\beta_1} L_\alpha - \int e^{-\frac{1}{2}(x, Bz + 2\xi y)^2 + 2(1-\alpha)\gamma(y - \nu \delta)^2} \, dz$$

$$\geq K_{\beta_1} L_\alpha - \int e^{-\frac{1}{2}(x, Bz + 2\xi y)^2 + 2(1-\alpha)\gamma(y - \nu \delta)^2} \, dz$$

for all $\infty < \eta(x, y) < \infty$. In fact, we shall impose the following stricter conditions on $(\mu, \eta)$:

$$1 = K_{\beta_1} L_\alpha - \int e^{-\frac{1}{2}(x, Bz + 2\xi y)^2 + 2(1-\alpha)\gamma(y - \nu \delta)^2} \, dz$$ (44)

and

$$1 \geq K_{\beta_1} L_\alpha - \int e^{-\frac{1}{2}(x, Bz + 2\xi y)^2 + 2(1-\alpha)\gamma(y - \nu \delta)^2} \, dz$$ (45)

The case $\rho = 1$ is precluded since it would imply that $Y = aX$ for some number $a$ and that the support of the distribution is not the entire real plane. When $\rho = 1$, there is essentially a single source and the usual successive refinement results can be used.
with equality for some pair $-\infty < (\hat{x}, \hat{y}) < \infty$. We shall term the set of $(\mu, \eta)$ that satisfy these conditions as $\hat{\Lambda}$. A set of conditions on $(b, \gamma, \nu)$ for $(\mu, \eta)$ to belong to $\hat{\Lambda}$ are given in the following lemma.

**Lemma 3:** A pair $(\mu, \eta)$ belongs to $\hat{\Lambda}$ if and only if the triple $(b, \gamma, \nu)$ satisfies the following conditions:

\begin{align}
 b &\geq 0 \\
 \gamma &\geq 0 \\
 \gamma &\leq \beta_2' \\
 \phi + \gamma \theta &> 0 \\
 \phi &\geq 0
\end{align}

where

\begin{align}
 \theta &\equiv 2\beta_1 + b\theta^2 \\
 \phi &\equiv b\beta_1 - \alpha\gamma \theta
\end{align}

and

\begin{align}
 K_0 &= \frac{1}{2\pi} \text{det}(E)^{\frac{3}{2}} \text{det}(D)^{\frac{1}{2}} \\
 K_1 &= \sqrt{\frac{\text{det}(E)}{\text{det}(D)}}\end{align}

with

\begin{align}
 E &\equiv B + \begin{bmatrix} 2\beta_1 & 0 \\ 0 & 2(1 - \alpha)\gamma \end{bmatrix} \\
 D &\equiv B + \begin{bmatrix} 2\beta_1 & 0 \\ 0 & 2(\beta_2 - \alpha)\gamma \end{bmatrix}
\end{align}

We shall denote the set of triples $(b, \gamma, \nu)$ satisfying (46a)–(46e) by $\Gamma(\alpha, \beta_1, \beta_2)$.

We provide the proof in Appendix A. The proof employs another matrix $A$, which we repeat here for convenience:

\begin{align}
 A &\equiv D^{-1} \begin{bmatrix} 2\beta_1 & 0 \\ -2\gamma \nu & 2\beta_2 \end{bmatrix}.
\end{align}

It follows from Theorem 1, (42), and Lemma 3 that for any $0 \leq \alpha < 1$, $\beta_1 > 0$, $\beta_2 > 0$, and $(b, \gamma, \nu) \in \Gamma(\alpha, \beta_1, \beta_2)$, we get

\begin{align}
 I_{\alpha, \beta_1, \beta_2}^* &\geq h(X, Y) + \log K_0 - \frac{1}{2} E\{Z^T B Z\} \\
 &= h(X, Y) + \log K_0 - \frac{b}{2} E\{(Y - \nu X)^2\}.
\end{align}

This implies that for any $0 \leq \alpha < 1$

\begin{align}
 R_\alpha(D_X, D_Y) &\geq h(X, Y) + \log K_0 - \frac{b}{2} E\{(Y - \nu X)^2\} \\
 &\quad - \beta_1 D_X - \beta_2 D_Y
\end{align}

for an arbitrary choice of $\beta_1 > 0$, $\beta_2 > 0$, and $(b, \gamma, \nu) \in \Gamma(\alpha, \beta_1, \beta_2)$. The lower bound (52) is tight if and only if (51) is tight and $(D_X, D_Y) \in D(\alpha, \beta_1, \beta_2)$.

We next analyze the conditions for tightness of (51).

**Lemma 4:** Given $0 \leq \alpha < 1$, $\beta_1 > 0$, and $\beta_2 > 0$, a triplet $(b, \gamma, \nu) \in \Gamma(\alpha, \beta_1, \beta_2)$ satisfies (51) with equality if and only if the following conditions are satisfied.

1) There exist random variables $(\hat{X}, \hat{Y})$ such that

\begin{align}
 \hat{Y} &= \nu \hat{X} + W \\
 \begin{bmatrix} X \\ Y \end{bmatrix} &= A \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \tag{54}
\end{align}

where $W$ is independent of $\hat{X}$, and $(N_1, N_2)$ is independent of $(\hat{X}, \hat{Y})$. The random variables $W$ and $(N_1, N_2)$ are zero-mean Gaussian with

\begin{align}
 \sigma_W^2 &= \frac{1}{2\gamma} - \frac{1}{2\beta_2'} \\
 C_N &= D^{-1}.
\end{align}

2) Either $\phi = 0$ or $\beta_2' = \gamma$. In other words

\begin{align}
 \phi(\beta_2' - \gamma) &= 0. \tag{57}
\end{align}

We provide the proof in Appendix B.

**Remark 2:** In fact, Condition 1 in Lemma 4 can be simplified. Using (53) and

\begin{align}
 A \begin{bmatrix} 1 \\ \nu \end{bmatrix} &= \begin{bmatrix} 1 \\ \nu \end{bmatrix}
\end{align}

we can replace (54) with

\begin{align}
 \begin{bmatrix} X \\ Y \end{bmatrix} &= A \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix} = A \begin{bmatrix} 1 \\ \nu \end{bmatrix} \hat{X} + \begin{bmatrix} N_1' \\ N_2' \end{bmatrix}
\end{align}

where $N'$ is independent of $\hat{X}$ and

\begin{align}
 C_{N'} &= D^{-1} + \sigma_W^2 A \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} A^T \\
 &= E^{-1} \tag{58}
\end{align}

where (58) follows from the standard matrix inversion lemma. This relation implies that

\begin{align}
 C_{Z} &= \delta \begin{bmatrix} 1 & \nu \\ \nu & \nu^2 \end{bmatrix} + E^{-1} \tag{59}
\end{align}

where $\delta \equiv E\{\hat{X}^2\} \geq 0$.

**Remark 3:** If (53) and (54) are satisfied, then

\begin{align}
 Y - \nu X &= \begin{bmatrix} -\nu & 1 \end{bmatrix} \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} + \begin{bmatrix} -\nu & 1 \end{bmatrix} \begin{bmatrix} N_1 \\ N_2 \end{bmatrix} \\
 &= \frac{2\beta_1}{\phi + \beta_2'\theta} \begin{bmatrix} -\nu & 1 \end{bmatrix} \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} + N_2 - \nu N_1 \\
 &= \frac{2\beta_1 \beta_2'}{\phi + \beta_2'\theta} W + N_2 - \nu N_1. \tag{60}
\end{align}

This simplifies the evaluation of $E\{(Y - \nu X)^2\}$ in (51) and (52).

*Consequently, $W$ is also independent of $(N_1, N_2)$.\*
B. RD Function for the Quadratic Gaussian Problem

In this subsection, we shall assume that \( Z \) is a pair of jointly Gaussian sources. We begin this subsection by handling the trivial cases. When \( Z \) is Gaussian, we have

\[
D^A_X(D_Y) = 1 - \rho^2(1 - D_Y) \quad (61)
\]

\[
D^B_X(D_Y) = 1 - \rho^2(1 - D_X) \quad (62)
\]

and the RD function for Gaussians becomes

\[
R^A_{\alpha}(D_X, D_Y) = \frac{1}{2} \log \frac{1}{D_X} \quad \text{if either } \alpha = 1 \text{ or } D^B_X(D_Y) \leq D_Y \leq 1,
\]

\[
R^B_{\alpha}(D_X, D_Y) = \frac{1 - \alpha}{2} \log \frac{1}{D_Y} \quad \text{if either } 0 < \alpha < 1, D_X = 1 \text{ or } \alpha = 0, D^A_X(D_Y) \leq D_X \leq 1.
\]

We now characterize \( R^A_{\alpha}(D_X, D_Y) \) for \( (D_X, D_Y) \in \mathcal{D}^N_\alpha \). When \( Z \) is Gaussian, \( \hat{X} \) is necessarily a Gaussian for (51) to be tight, and the first condition in Lemma 4 reduces to (59). The distortion levels induced by \( p_{\hat{X}|X,Y} \) satisfying the conditions of Lemma 4 are

\[
D_X = \sigma^2_{X|Y} = \frac{b + 2\gamma(1 - \alpha)}{2(\phi + \gamma \theta)} \quad (63)
\]

\[
D_Y = \sigma^2_{Y|X} = \frac{2(1 - \alpha)}{2(\phi + \beta \theta)} \quad (64)
\]

where to obtain (64), we use the fact that either \( \phi = 0 \) or \( \phi = \nu \hat{X} \). Since \( (D_X, D_Y) \in \mathcal{D}(\alpha, \beta_1, \beta_2) \), we can substitute (60), (63), and (64) into (52) to obtain after some further algebra that

\[
R^A_{\alpha}(D_X, D_Y) = \frac{1}{2} \log(1 - \rho^2) + \frac{\alpha}{2} \log \det(\mathbf{E}) + \frac{1 - \alpha}{2} \log \det(\mathbf{D}). \quad (65)
\]

Since it is the distortion levels \( (D_X, D_Y) \) that are usually fixed, we solve for the other parameters in terms of the distortions. If \( \rho = 0 \), the two sources are independent and coding solely the first source in the first layer and the second source in the second layer is optimal. Therefore

\[
R^A_{\alpha}(D_X, D_Y) = R_X(D_X) + (1 - \alpha)R_Y(D_Y). \quad (66)
\]

We obtain the same result from our approach by assigning \( \beta_1 = \frac{1}{D_X}, \beta_2 = \frac{1}{D_Y}, b = \alpha, \gamma = \frac{1}{2}, \) and \( \nu = 0 \).

To handle the \( 0 < \rho < 1 \) case, we prove the following lemma.

Lemma 5: Given \( 0 \leq \alpha < 1, 0 < \rho < 1, \) and \( (D_X, D_Y) \in \mathcal{D}^N_\alpha \), a set of parameters \((\beta_1, \beta_2)\) and \((b, \gamma, \nu)\) in \( \Gamma(\alpha, \beta_1, \beta_2) \) leads to (65) if and only if

\[
\xi \triangleq (1 - \delta)(1 - \nu^2 - \delta - (\rho - \nu \delta))^2 > 0 \quad (67a)
\]

\[
\phi = \frac{f_\alpha(\nu)}{2 \gamma^2 \nu (1 - \alpha)} \geq 0 \quad (67b)
\]

\[
\rho \geq \nu \delta \quad (67c)
\]

\[
\rho > \frac{\nu}{\gamma} \quad (67d)
\]

\[
\nu < \frac{1}{\rho} \quad (67e)
\]

where

\[
\gamma \leq \beta_2 \quad (67f)
\]

\[
\phi(\beta_2 - \gamma) = 0 \quad (67g)
\]

\[
\delta = 1 - D_X \quad (67h)
\]

\[
b = \frac{\nu \xi - \nu}{\nu \xi - \nu} \quad (67i)
\]

\[
\gamma = \frac{\nu - \rho}{2\nu(1 - \alpha \xi)} \quad (67j)
\]

\[
\beta_1 = 1 - \nu \rho \frac{2\nu}{2\nu(1 - \alpha \xi)} \quad (67k)
\]

\[
\beta_2 = \frac{1 - \nu^2 \delta(\nu - \rho)(1 - \nu^2 \delta)(\nu - \rho)}{2\nu D_Y \xi} \quad (67l)
\]

Proof: We first show that the conditions for equality imply (67), (67f) and (67g) are restatements of (66c) and (57), respectively. Similarly, observing from (59) that \( \xi = \det(\mathbf{E}^{-1}) \), (67a) is only a restatement of \( \det(\mathbf{E}^{-1}) > 0 \). Now, the matrix (59) expands to

\[
1 = \delta + \frac{b + 2\gamma(1 - \alpha)}{2(\phi + \gamma \theta)} \quad (68)
\]

\[
1 = \nu^2 \delta + \frac{\theta - \nu^2 b}{2(\phi + \gamma \theta)} \quad (69)
\]

\[
\rho = \nu \delta + \frac{\nu \theta}{2(\phi + \gamma \theta)} \quad (70)
\]

Comparing (68) and (63), we get (67h). Observe that (70) implies \( \nu > 0 \). Rearranging (70) and using the fact that \( \det(\mathbf{E}^{-1}) = \frac{1}{\xi^2 \nu^2 \delta^2} \), we get (67i). Then, by the nonnegativity of \( b \), we get (67c). From (69) and (70), we have

\[
1 - \nu \rho = \frac{\theta - \nu^2 b}{2(\phi + \gamma \theta)} = 2\beta_2 \xi
\]

which gives (67k) and (67e). Similarly, from (68) and (70), we get (67j). Comparing (64) and (69), we can write

\[
D_Y = (1 - \nu^2 \delta) \frac{\phi + \gamma \theta}{\phi + \beta_2 \theta}. \quad (67d)
\]

This gives (67l) and (67d), where to obtain the former we use (67) together with (88). Finally, (67b) follows by substituting for \( b, \beta_1, \gamma \) in the expression for \( \phi \).

By reversing these steps, we can also prove that (67) implies the conditions for equality. \( \square \)

We therefore see that the RD problem can be completely solved if we can choose a \( \nu \) satisfying (67) for a given \( (D_X, D_Y) \). For every \( \alpha \), we can divide the region \( \mathcal{D}^N_\alpha \) into various regions depending on the behavior of \( \nu \). Specifically, there are two regions when \( \alpha > 0 \):

1) \( \phi = 0, \beta_2 > \gamma > 0 \)

2) \( \beta_2 = \gamma > 0 \),

When \( \alpha = 0 \), the first region splits into two subregions, one of which corresponds to \( \beta_1 = 0 \), which lies outside \( \mathcal{D}^N_\alpha \) and has hence already been taken care of. The other subregion corresponds to \( b = 0 \).
The following lemma is crucial in analyzing the first case, which requires us to solve \( f_\alpha(\nu) = 0 \) due to (67b).

**Lemma 6:** For every \( 0 < \alpha < 1 \) and \( 0 < D_X < 1 \), \( f_\alpha(\nu) \) has a unique root \( \nu_0 \) satisfying (67c)-(67e) and (67a), except when \( \nu = 0 \) and \( D_X \geq 1 - \rho^2 \), in which case \( f_\alpha(\nu) \) has no such root.

The proof is given in Appendix C.

Now, define
\[
g_\alpha(D_X) \triangleq 1 - \nu_0^2(1 - D_X)
\]
for \( 0 < \alpha < 1 \) and \( \alpha = 0, D_X < 1 \),
\[
g_\alpha(D_X) \triangleq \lim_{\alpha \to 0} g_\alpha(D_X) = 1 - \frac{1}{\rho^2}(1 - D_X)
\]
for \( D_X \geq 1 - \rho^2 \), and finally
\[
g_1(D_X) \triangleq 1 - \rho^2(1 - D_X).
\]
We are now ready to state the main theorem of this section.

**Theorem 2:** The successive coding RD function for Gaussian sources with squared-error distortion with \( 0 < D_X < 1 \) and \( 0 < D_Y < 1 \) is
\[
R_\alpha(D_X, D_Y) = \frac{1}{2} \log \frac{1 - \rho^2}{D_X(1 - \rho^2(1 - D_X)) - (\rho - \nu(1 - D_X))^2} + \frac{1 - \alpha}{2} \left[ \frac{1}{\log \frac{1 - \nu^2(1 - D_X)}{D_Y}} \right]^{-}
\]
(71)
where
\[
\nu = \begin{cases} 
\nu_0, & \text{if } (D_X, D_Y) \in \mathcal{D}_\alpha^1 \\
\nu^*, & \text{if } (D_X, D_Y) \in \mathcal{D}_\alpha^2 \\
\frac{1}{\rho}, & \text{if } \alpha = 0, D_Y(1 - D_X) \leq D_Y < 1 \\
\sqrt{D_Y^2 + 1}, & \text{if } D_Y(1 - D_X) \leq D_Y < 1
\end{cases}
\]
with
\[
\nu^* \triangleq \sqrt{1 - \frac{D_Y}{1 - D_X}}.
\]
(72)
Also,
\[
\mathcal{D}_\alpha^1 \triangleq \{(D_X, D_Y) \in \mathcal{D}_\alpha^{N \alpha} : 0 < D_Y < g_\alpha(D_X)\}
\]
and
\[
\mathcal{D}_\alpha^2 \triangleq \{(D_X, D_Y) \in \mathcal{D}_\alpha^{N \alpha} : g_\alpha(D_X) \leq D_Y < g_1(D_X)\}.
\]

**Proof:** As mentioned in the Proof of Lemma 5, we have from (59) that \( \det(E) = \frac{1}{\xi} \). Similarly
\[
\det(D) = 2(\phi + \beta_2 \theta)
\]
\[
= 2(\phi + \gamma \theta) \frac{\beta_2}{\gamma}
\]
\[
= 1 - \nu^2(1 - D_X)
\]
(74)
(75)
where (74) follows from (88), and (75) from (67j) and (67l). Thus, using (67a), (65) can be rewritten as in (71).

It then remains to show (72). Comparing (67j) and (67l), it can be seen that the case \( \phi = 0, \beta_2 > \gamma > 0 \) corresponds to \( \nu = \nu_0 \) and \( (D_X, D_Y) \in \mathcal{D}_\alpha^1 \) for any \( 0 < \alpha < 1 \) and for \( \alpha = 0 \) with \( D_X < 1 - \rho^2 \).

For the case \( \beta_2 = \gamma > 0 \) and (67j) and (67l) yield \( \nu = \nu^* \). This condition occurs when \( (D_X, D_Y) \in \mathcal{D}_\alpha^2 \), because as discussed in the Proof of Lemma 6, \( f_\alpha(\nu^*) \geq 0 \) and \( \xi > 0 \) in the interval \( \rho < \nu^* < \nu_0 \) when \( 0 < \alpha < 1 \) and \( \alpha = 0, D_X < 1 - \rho^2 \), and in the interval \( \rho < \nu^* < \frac{1}{\rho} \) when \( \alpha = 0 \) and \( D_X \geq 1 - \rho^2 \).

The other two values of \( \nu \) in (72) handle the cases where \( (D_X, D_Y) \notin \mathcal{D}_\alpha^{N \alpha} \). Note that \( g_1(D_X) = D_Y(D_X) \).

We observe that \( g_\alpha(D_X) \) is increasing in \( \nu \) for all fixed \( 0 < D_X < 1 \). To see that, rewrite \( f_\alpha(\nu) \) as
\[
f_\alpha(\nu) = (\rho - \nu)(1 - \nu(1 - D_X)) - \nu(1 - \nu)(1 - \nu^2(1 - D_X))
\]
Hence, for every \( D_X \) and \( \nu \) satisfying (67c)-(67e), \( f_\alpha(\nu) \) decreases as \( \nu \) increases. In other words, the interval of \( \nu \) satisfying (67c)-(67e) and \( f_\alpha(\nu) \geq 0 \) shrinks with increasing \( \alpha \).

Since \( \nu \) is fixed as in (73), this implies that the region \( g_\alpha(D_X) \leq D_Y < g_1(D_X) \) shrinks with increasing \( \alpha \). Since \( g_1(D_X) \) stays constant, \( g_\alpha(D_X) \) must increase. Note also that when \( \alpha = 1 \), the unique root of \( f_\alpha(\nu) \) satisfying (67c)-(67e) and (67a) is \( \nu = \rho \) and therefore \( \lim_{\alpha \to 1} g_\alpha(D_X) = g_1(D_X) \).

Fig. 3 shows \( g_\alpha(D_X) \) for various values of \( \alpha \) and \( \rho \). Given \( \rho \) and \( D_Y \), \( g_\alpha(D_X) \) increases as \( \alpha \) increases as mentioned earlier. Also, the region corresponding to \( \alpha = \phi = 0 \) and \( D_X < 1 - \rho^2 \) shrinks towards the point \( D_X = D_Y = 0 \) from \( 0 < D_X < 1, 0 < D_Y < 1 \) to an empty region as \( \rho \) increases from 0 to 1.

One can substitute the parameters satisfying (67) into (53) and (54) to obtain
\[
I(X; Y; \hat{X}) \leq \frac{1}{2} \log \frac{1 - \rho^2}{D_X(1 - \rho^2(1 - D_X)) - (\rho - \nu(1 - D_X))^2}
\]
(76)
\[
I(Y; X; \hat{Y}) \leq \frac{1}{2} \log \frac{1 - \rho^2}{D_Y(1 - \rho^2(1 - D_Y)) - (\rho - \nu(1 - D_Y))^2}
\]
(77)
for the optimal \( p_{X|Y}^*, p_{Y|X}^* \), where \( \nu \) is as in (72).

Fig. 4 plots the \( (R_1, R_1 + R_2) \) tradeoff using (76) and (77) for \( 0 \leq \alpha \leq 1 \) and for various values of \( \rho \). Irrespective of the relative values of \( D_X \) and \( D_Y \), for all values of \( \rho \), the curves meet at the same minimum \( R_1 = R_X(D_X) \). However, the required second-layer rate decreases with increasing \( \rho \) as expected. It is interesting to study the behavior of the points corresponding to \( \alpha = 0 \) for various pairs \( (D_X, D_Y) \). Of particular interest is the “corner” point \( (R_\alpha^1, R_\alpha^1 + R_\beta^2) \), the point with the lowest \( R_1 \) such that \( \alpha = 0 \). Observe from (76) and (77) that whether or not \( R_\alpha^2 = 0 \) depends solely on the value of \( \frac{1 - \nu^2(1 - D_X)}{D_Y} \).

Now, \( (67f), (67j), \) and (67l) collectively imply that
\[
\frac{1 - \nu^2(1 - D_X)}{D_Y} \leq 1, \quad \text{if } \beta_2 = \gamma > 0
\]
(78)
Using this, we analyze the three cases \( D_X < D_Y, D_X = D_Y, \) and \( D_X > D_Y \) separately.
Fig. 3. $g_\alpha(D_X)$ for $\alpha = 0, 0.1, \ldots, 1.0$. Observe that $g_\alpha(D_X)$ is increasing in $\alpha$.

Fig. 4. The tradeoff of achievable $R_1$ and $R_1 + R_2$ for $\rho = 0, 0.1, \ldots, 1.0$. The $\times$ mark the points with smallest $R_1$ for which $\alpha = 0$.

- $D_X < D_Y$: When $\rho < \nu^\alpha(1 - D_X)$, the point $(D_X, D_Y)$ is in the region with $\phi = 0$, $\beta_2 > \gamma$, and $\beta_2 = \gamma$. Consequently, $R_2^* > 0$, or equivalently, $(R_1^*, R_1^* + R_2^*)$, lies above the $R_1 = R_1 + R_2$ line. For all $\rho \geq \nu^\alpha(1 - D_X)$,
\( R_0^\theta = 0 \), since \( \beta_2 = \gamma \). However, if \( \rho \) is further increased to \( \rho \geq \sqrt[\theta]{\beta_1, \beta_2} = 0 \), and the curve corresponding to the various \( \alpha \) is a line parallel to the ordinate axis passing through \((R_X(D_X), R_Y(D_X))\).

- \( D_X = D_Y \): The behavior is pretty similar to that in the \( D_X < D_Y \) case except for the fact that since \( \beta_1 = 1 \), the \( \rho = 1 \) curve is the only one leading to a line parallel to the ordinate axis passing through \((R_X(D_X), R_Y(D_X))\).

- \( D_X > D_Y \): \( R_0^\theta > 0 \) whenever \( \rho < \sqrt[\theta]{1/D_X} \) as before. However, the \( R_0^\theta = 0 \) interval for \( \rho \) is limited to \( \sqrt[\theta]{(1/D_X)} \leq \rho \leq 1/\sqrt[\theta]{\beta_1} \). When \( \rho > 1/\sqrt[\theta]{\beta_1} \), the point \((D_X, D_Y)\) goes into the \( \beta_1 = 0 \) region. Consequently, \( R_0^\theta + R_2^\theta \) remains fixed at \( R_Y(D_Y) \), and \( R_0^\theta \) starts decreasing after \( \rho \geq 1/\sqrt[\theta]{\beta_1} \). Finally, when \( \rho \to 1 \), the curve becomes the union of two half-lines meeting at right angles at the point \((R_X(D_X), R_Y(D_Y))\).

Critical values of \( \rho \) where the behavior changes in each case are also shown in Fig. 4.

We close this section by making an observation about the optimal \((\hat{X}, \hat{Y})\). The fact that \((\mu(x,y), \eta(x,y,\hat{x})\) of the form (42)–(43) achieves the RD function implies that \( \hat{X}, \hat{Y} \) satisfy the Markov condition \( X - (\hat{X}, \hat{Y}) - Y \). This can be shown using the expression for \( p(\hat{X}, \hat{Y}) \) from (32):

\[
p(\hat{Y}|x, \hat{x}) = \frac{p(\hat{X}, \hat{Y}) \cdot x, \hat{x}}{\int p(\hat{X}, \hat{Y}) \cdot x, \hat{x} d\hat{Y}} = \frac{p(\hat{X}) \cdot x, \hat{x}}{\int p(\hat{X}) \cdot x, \hat{x} d\hat{Y}} = \frac{p(\hat{X}) \cdot x, \hat{x}}{\int p(\hat{X}) \cdot x, \hat{x} d\hat{Y}}.
\]

The Markov condition follows since the right-hand side (RHS) is independent of \( x \).

C. The Gaussian RD Function and General Sources

The successive coding RD function for jointly Gaussian sources exhibits some extremal properties as given in the following theorem.

**Theorem 3:** Let \( Z \) be an general source pair as defined at the beginning of this section. Let \( Z^* \) be a Gaussian source pair with the same covariance matrix as \( Z \) and let \( R_0^\theta(D_X, D_Y) \) be its RD function. If \( D_0^\theta \) denotes the region corresponding to nontrivial Lagrangian parameters for the Gaussian, for \( 0 \leq \alpha < 1 \), \( (D_X, D_Y) \in D_0^\theta \), the RD function for \( Z \) satisfies

\[
R_0^\theta(D_X, D_Y) - D(Z||Z^*) \leq R_0^\theta(D_X, D_Y) \leq R_0^\theta(D_X, D_Y).
\]

(78)

**Remark 4:** The theorem could be extended by using appropriate \((\mu, \eta)\) pairs that are optimal for Gaussian sources for the points lying outside \( D_0^\theta \) and we can obtain the same upper and lower bounds. However, the lower bound so obtained is worse than the trivial lower bound in terms of the single-layer RD functions \( R_X(D_X) \) and \( R_Y(D_Y) \), that is

\[
R_0(D_X, D_Y) \geq \max[R_X(D_X), (1-\alpha)R_Y(D_Y)].
\]

(79)

If \( R_X(D_X) \) and \( R_Y(D_Y) \) are not easily computable and the RD functions in (79) are replaced by their respective Shannon lower bounds, we would still obtain better lower bounds than those in (78) when \((D_X, D_Y) \not\in D_0^\theta \).

**Proof:** We start with the lower bound. Since the membership of a triple \((h, \gamma, \nu)\) in \( \Gamma(\beta_1, \beta_2) \) does not depend on the source distribution, we can use the parameters that achieve \( R_0^\theta(D_X, D_Y) \) in (52) for \( Z \), giving

\[
R_0(D_X, D_Y) \geq h(Z^*) + R_0^\theta(D_X, D_Y) - h(Z^*).
\]

But \( h(Z^*) - h(Z) = D(Z||Z^*) \) and the lower bound follows.

From (54), we have that for Gaussian sources, the source random variable \( Z^* \), and the reproduction random variable \( Z^* \) are jointly Gaussian. Therefore, the optimal forward channel for Gaussian source is of the form

\[
\tilde{Z} = GZ^* + U
\]

where \( U \) is Gaussian and independent of \( Z^* \). Let \( \tilde{Z} \) be the output of the same forward channel when \( Z \) is the input. We see that the distortion constraints are satisfied by \( Z^* \). Comparing the rates, we have

\[
R_0^\theta(D_X, D_Y) - R_0(D_X, D_Y) \geq \alpha I(Z^*; X^*) + (1-\alpha)I(Z^*; Z^*) - \alpha I(Z; \hat{X}) - (1-\alpha)I(Z; \hat{Z})
\]

\[
= \alpha(h(X^*) - h(\hat{X})) + (1-\alpha)(h(\hat{Z}^*) - h(\hat{Z})) \geq 0
\]

since Gaussian random variables have the largest differential entropy among all random variables with a given covariance matrix and since the sum of two random vectors is Gaussian if and only if they are both Gaussian.

This theorem implies that the Gaussian is the source that is hardest to code among all sources with a given covariance matrix. Further, this theorem implies that for some choice of parameters \((\beta_1, \beta_2)\), the lower bound in (52) differs from the RD function of \( Z \) by at most \( D(Z||Z^*) \).

V. Conclusion

In this paper, we analyzed a special case of the successive refinement problem, where the first- and the second-layer descriptions are, respectively, for the primary and secondary source, both of which are available at the encoder simultaneously. While the characterization of the RD function follows from results on successive refinement, computation of this function is difficult. Therefore, using the dual formulation of the successive refinement RD optimization problem, we derived a class of lower bounds for the successive refinement RD function. By specializing these bounds to the successive coding RD problem and analyzing the conditions for tightness of this bound, we also obtained an explicit characterization of the RD function for Gaussian sources with squared error as the distortion measure. The analysis of the Gaussian RD function also led to simple upper and lower bounds on the RD function for general sources under the same distortion measure. This is similar to what happens in single-source coding, where the RD function for a general source is bounded from below.
by the Shannon lower bound and bounded from above by the RD function for a Gaussian source with the same covariance matrix as the original source. In both single-source coding and successive coding, the difference between the upper bound and the lower bound is the relative entropy between the original source and its corresponding Gaussian source.

APPENDIX A
PROOF OF LEMMA 3

Defining

\[ s = \begin{bmatrix} 1 \\ \nu \end{bmatrix} \]

the exponent in (44) can be rewritten as

\[ z^T B z + 2\beta_1 (x - \hat{x})^2 + 2(1 - \alpha)\gamma(y - \nu \hat{x})^2 \]
\[ = z^T E z - 2\hat{x} z^T \begin{bmatrix} 2\beta_1 \\ 2(1 - \alpha)\gamma \nu \end{bmatrix} \]
\[ + \hat{x}^2 \left( \frac{2\beta_1}{2(1 - \alpha)\gamma \nu} \right) \]
\[ = (z - s \hat{x})^T E (z - s \hat{x}) \]
\[ + 2\hat{x} E \left( \frac{2\beta_1}{2(1 - \alpha)\gamma \nu} - E s \right) \]
\[ + \hat{x}^2 \left( \frac{2\beta_1}{2(1 - \alpha)\gamma \nu} - s^T E s \right) \]
\[ = (z - s \hat{x})^T E (z - s \hat{x}). \]

Thus, for (44) to be satisfied, we need to choose the parameters \((b, \gamma, \nu)\) such that \(E\) is positive definite, in which case

\[ K_0 K_1^{1-\alpha} = \frac{\sqrt{\det(E)}}{2\pi}. \] (80)

Since

\[ D = E + \begin{bmatrix} 0 & 0 \\ 0 & 2(\beta_2 - \gamma) \end{bmatrix} \]

\(D\) is positive definite if \(E\) is, and we can define

\[ A \triangleq D^{-1} \begin{bmatrix} 2\beta_1 \\ -2\alpha \gamma \nu \\ 2\beta_2 \end{bmatrix} \]
\[ F \triangleq 2\beta_1 (\beta_2 - \gamma) \begin{bmatrix} 2(1 - \alpha)\gamma \nu \\ 0 \\ 2\beta_2 \end{bmatrix} - A^T DA. \]

Manipulating the exponent in (45), we obtain

\[ z^T B z + 2\beta_1 (x - \hat{x})^2 + 2\beta_2 (y - \hat{y})^2 - 2\gamma \nu (y - \nu \hat{x})^2 \]
\[ = (z - A \hat{x})^T D (z - A \hat{x}) + \hat{x}^T F \hat{x}. \] (81)

Then, (45) becomes

\[ \frac{\sqrt{\det(E)}}{K_1 \sqrt{\det(D)}} e^{-\frac{1}{2} z^T F z} \leq 1 \]

which is satisfied, together with (80), if \(F\) is positive semi-definite and \(K_0\) and \(K_1\) are as in (47) and (48). Since (45) has to be satisfied with equality for some pair \((\hat{x}, \hat{y})\), (47) and (48) are also necessary.

The matrices \(A\) and \(F\) can be expressed explicitly as

\[ A = \frac{1}{\phi + \beta_2 \gamma \nu} \begin{bmatrix} \phi + 2\beta_1 \beta_2 \gamma \nu \\ \beta_1 \beta_2 \gamma \nu \nu \end{bmatrix} \]
\[ F = \frac{2\beta_1 \phi}{\phi + \beta_2 \gamma \nu} \begin{bmatrix} \nu^2 & -\nu \\ -\nu & 1 \end{bmatrix}. \]

Also \(\det(D) = 2(\phi + \beta_2 \gamma \nu)\) and \(\det(E) = 2(\phi + \gamma \theta)\). The conditions \(E > 0\), and \(F \geq 0\) then translate to

\[ \phi > 0 \]
\[ \phi + \gamma \theta > 0 \]
\[ \theta > 0 \]
\[ b + 2(1 - \alpha) \gamma > 0. \]

The first two conditions are the same as (46d) and (46e), while the third is a consequence of \(\beta_1 > 0\). The final condition is implied by (46a)–(46e) since if \(b + 2(1 - \alpha) \gamma = 0\) (which is possible only if \(b\) and \(\gamma\) are zero), so is \(\phi + \gamma \theta\). Therefore, (46a)–(46e) form a necessary and sufficient set of conditions for \((\mu, \eta) \in \bar{L}\).

APPENDIX B
PROOF OF LEMMA 4

We use the conditions for tightness of the lower bound from Corollary 1.

Note first that \(\beta_2 > 0\) implies that \(\gamma\) is also nonzero, because otherwise (43) would translate to \(\eta(x, y, \hat{x}) = K_1\), and that, in turn, would prohibit (28).

From (28), we obtain

\[ \eta(x, y, \hat{x}) = K_1 e^{-(y - \hat{y})^2} = \int p(\hat{y} | \hat{x}) e^{-(y - \hat{y})^2} d\hat{y} \]

or

\[ K_1 \sqrt{\frac{\beta_2}{\gamma}} \frac{2(1 - \alpha) \gamma \nu}{\sqrt{\pi}} e^{-(y - \hat{y})^2} = \int p(\hat{y} | \hat{x}) \sqrt{\frac{\beta_2}{\gamma}} e^{-(y - \hat{y})^2} d\hat{y}. \] (87)

Since the RHS of (87) is a convolution of two distributions, the left-hand side (LHS) must also be a valid distribution, which implies

\[ K_1 = \sqrt{\frac{\gamma}{\beta_2}}. \]

Comparing this with (48), we obtain

\[ \frac{\phi + \gamma \theta}{\phi + \beta_2 \gamma \nu} = \frac{\gamma}{\beta_2} \]

and therefore (57). Further, (87) also implies that \(p(\hat{y} | \hat{x})\) is a Gaussian distribution. Since convolution of distributions signifies addition of random variables, by comparing the mean and variance of the random variables corresponding to the two sides of (87), we get (53) and that \(\gamma_{11} = \frac{1}{\beta_2} = \frac{2\gamma}{\beta_2} \nu^2\).

From (29), the second condition for tightness, we have that

\[ p(x, y) \]
\[ = K_0 K_1^{1-\alpha} \int \int p(\hat{x}, \hat{y}) \cdot e^{-\frac{1}{2} z^T B z + \frac{1}{2} \beta_2 (y - \hat{y})^2 - \gamma (y - \nu \hat{x})} d\hat{x} d\hat{y} \]
\[ = K_0 K_1^{1-\alpha} \int \int p(\hat{x}, \hat{y}) e^{-\frac{1}{2} (z - A \hat{x})^T D (z - A \hat{x})} d\hat{x} d\hat{y}. \] (89)
\[ \int \int p(\hat{x}, \hat{y}) \frac{e^{-\frac{1}{2}((x-A\hat{x})^TDX-(x-A\hat{x}))}}{2\pi \sqrt{\det(D^{-1})}} \, d\hat{x} d\hat{y} \]  

(90)

where (89) follows from (81) and that

\[ z^T Fz = 0 \]  

(91)

for all \( z = (\hat{x}, \hat{y}) \) with \( p(\hat{x}, \hat{y}) > 0 \). The proof of (91) is trivial if \( \phi = 0 \). If, on the other hand, \( \phi > 0 \), then \( \gamma = \beta \hat{x} \) and, hence, \( \sigma_W^2 = 0 \), implying \( Y = \nu \hat{X} \). Identity (91) then follows from

\[ [1 \nu] F [1 \nu] = 0. \]

The proof is finished by observing that (90) is possible if and only if (54) holds with \( C_N = D^{-1} \).

**APPENDIX C**

**PROOF OF LEMA 6**

Since \( f_\alpha(\nu) \) is a cubic equation, it has at most three real solutions. We first compute

\[
\begin{align*}
    f_\alpha(\frac{\mu}{\rho}) &= -\alpha \frac{\mu}{\rho^2} (1-\delta)(\delta - \rho^2) \\
    f_\alpha(\rho) &= \rho(1-\alpha)(1-\delta)(1-\rho^2) \\
    f_\alpha(\frac{1}{\rho}) &= -\alpha \frac{1-\rho^2}{\rho^3} (\rho^2 - \delta).
\end{align*}
\]

Now, observe that \( f_\alpha(\rho) > 0 \) always, \( f_\alpha(\frac{\mu}{\rho}) \) has the same sign as \( \alpha(\rho^2 - \delta) \), and \( f_\alpha(\frac{1}{\rho}) \) has the same sign as \( \alpha(\delta - \rho^2) \).

Let us first tackle \( 0 < \alpha < 1 \). When \( DX > 1 - \rho^2 \), or equivalently, \( \rho^2 > \delta \), we have

\[ \rho < \frac{1}{\rho} < \frac{1}{\mu}. \]

Thus, \( f_\alpha(\rho) > 0, f_\alpha(\frac{\mu}{\rho}) < 0 \), and \( f_\alpha(\frac{1}{\rho}) > 0 \). Therefore, there is a unique root \( \rho < \nu_0 < \frac{1}{\rho} \) satisfying (67c)–(67e). Further, since \( \xi \) is quadratic in \( \nu \) with a negative coefficient for \( \nu^2 \), it must be positive at \( \nu_0 \) because it is positive at both \( \nu = \rho \) and \( \nu = \frac{1}{\rho} \).

Similarly, when \( DX < 1 - \rho^2 \), since \( \rho < \frac{\mu}{\rho} < \frac{1}{\rho} \) and \( f_\alpha(\rho) > 0, f_\alpha(\frac{\mu}{\rho}) < 0 \), and \( f_\alpha(\frac{1}{\rho}) > 0 \) there is a unique root \( \rho < \nu_0 < \frac{\mu}{\rho} \) satisfying (67c)–(67e). We can also conclude \( \xi > 0 \) as above since \( \xi > 0 \) at \( \nu = \frac{\mu}{\rho} \).

When \( DX = 1 - \rho^2 \), we have \( \rho < \frac{1}{\rho} = \frac{\mu}{\rho} \) implying \( f_\alpha(\frac{1}{\rho}) = 0 \). However, for \( \nu = \frac{1}{\rho}, \xi = 0 \). To show the existence (and uniqueness) of a root \( \rho < \nu_0 < \frac{1}{\rho} \), we consider \( h_\alpha(\nu) = \frac{f_\alpha(\nu)}{1-\rho^2} \), and observe that \( h_\alpha(\rho) > 0 \) and \( h_\alpha(\frac{1}{\rho}) < 0 \).

Now for \( \alpha = 0, f_\alpha(\nu) \) becomes a quadratic equation whose roots are \( \frac{1}{\rho} \) and \( \frac{\mu}{\rho} \). If \( DX < 1 - \rho^2, \nu_0 = \xi \) since \( \xi < \frac{1}{\rho} \). However, when \( DX \geq 1 - \rho^2, \) since \( \frac{1}{\rho} < \xi \), one cannot satisfy (67c) and (67e) simultaneously with either \( \nu = \frac{1}{\rho} \) or \( \nu = \frac{\mu}{\rho} \). In both cases, \( \xi > 0 \).

**REFERENCES**


**Jayanth Nayak** received the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 2005. He is currently a Research Staff Member at Mayachitra, Inc., in Santa Barbara. His research interests include information theory, pattern recognition, and computer vision.

**Ertem Tuncel** (S’99–M’04) received the Ph.D. degree in electrical and computer engineering from University of California, Santa Barbara, in 2002. In 2003, he joined the Department of Electrical Engineering, University of California, Riverside, where he is currently an Associate Professor. His research interests include rate–distortion theory, multiterminal source coding, joint source–channel coding, zero-error information theory, and content-based retrieval in high-dimensional databases.

Prof. Tuncel received the 2007 National Science Foundation CAREER Award.