Computation and Analysis of the N-Layer Scalable Rate-Distortion Function

Ertem Tuncel, *Student Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*—Methods for determining and computing the rate-distortion (RD) bound for N-layer scalable source coding of a finite memoryless source are considered. Optimality conditions were previously derived for two layers in terms of the reproduction distributions q_{y_1} and $q_{y_2|y_1}$. However, the ignored and seemingly insignificant boundary cases, where $q_{y_1} = 0$ and $q_{y_2|y_1}$ is undefined, have major implications on the solution and its practical application. We demonstrate that, once the gap is filled and the result is extended to N-layers, it is, in general, impractical to validate a tentative solution, as one has to verify the conditions for *all conceivable* $q_{y_{i+1},\ldots,y_N|y_1,\ldots,y_i}$ at each $(y_1, \ldots, y_i)$ such that $q_{y_1,\ldots,y_i} = 0$.**

As an alternative computational approach, we propose an iterative algorithm that converges to the optimal joint reproduction distribution $q_{y_1,\ldots,y_N}$, if initialized with $q_{y_1,\ldots,y_N} > 0$ everywhere. For nonscalable coding $(N = 1)$, the algorithm specializes to the Blahut–Arimoto algorithm. The algorithm may be used to directly compute the RD bound, or as an optimality testing procedure by applying it to a perturbed tentative solution q. We address two additional difficulties due to the higher dimensionality of the RD surface in the scalable $(N > 1)$ case, namely, identifying the sufficient set of Lagrangian parameters to span the entire RD bound; and the problem of efficient navigation on the RD surface to compute a particular RD point.

***Index Terms*—Alternating minimization, Kuhn–Tucker optimality conditions, rate distortion (RD), scalable source coding, successive refinement.**

I. Introduction

Scalable source coding has received much attention in the last decade, especially after the advances in heterogeneous networks such as the Internet, because it enables serving a diverse set of users with differing bandwidth constraints. In scalable source coding, N descriptions, ranging from coarse to fine, are embedded into a single bit stream. Hence, users with a low-bandwidth connection can reproduce the signal at reasonable quality, although they only access a subset of the bit stream, while high bandwidth users can achieve high-quality reproduction of the source.

The early treatment of the problem of scalable coding within rate-distortion (RD) theory is due to Koshelev [10], [11], and Equitz and Cover [8]. These papers were concerned with the conditions under which scalable coding is possible without compromising the RD performance. Koshelev used the term *divisibility*, and Equitz and Cover coined the term *successive refinability*. Here, we follow [12] and employ the term "successive refinement without rate loss" to distinguish from "plain" successive refinement. Rimoldi [14] addressed the more general question and discovered necessary and sufficient conditions for the achievability of any sequence of rates and distortions. Later, Effros [7] extended these results for stationary ergodic and nonergodic sources. In an interesting recent work, Lastras and Berger [12] proved that for continuous reproduction alphabets and difference distortion measures, it is possible to *universally* bound the rate loss (the extra rate penalty paid for using a scalable coding scheme). Specifically, they showed that for the squared error distortion measure, the rate loss is bounded by half a bit at each layer, i.e., for an arbitrary source, there exists a scalable source coder achieving distortions $\{D_i\}_{i=1}^N$ and rates $\{R(D_i) + 1/2\}_{i=1}^N$, where $R(D)$ denotes the nonscalable RD function. This important result leaves open a few questions. It is unknown whether similar bounds exist for other cases (e.g., finite-alphabet sources with finite reproduction alphabets). Another concern is that the rate loss may become significant at low-resolution applications, i.e., where the rate is comparable to, or is lower than, 1/2 bits.

In this paper, we consider exact computation of the N-layer scalable RD surface for finite-alphabet sources. This problem is first analyzed in [7, Sec. V] for the case of $N = 2$, where, a nonlinear system of equations and inequalities in terms of the optimal reproduction distribution $q^*_{y_1, y_2}$ is formed. This system parallels the nonscalable RD optimality conditions, and is typically employed to find the optimum by the trial of "tentative solutions" that satisfy a subset of the inequalities with equality, until the one also satisfying the remaining inequalities is found. (See [2, Sec. 2.6] for a detailed description of such an approach for the nonscalable RD analysis.) However, unless symmetry or other properties of the problem help in reducing the space of possible tentative solutions, this approach becomes impractically complex as the size of the reproduction alphabet grows. For an extreme example, if the source and the reproduction alphabets are continuous, one has to finely discretize the reproduction space, and the number of tentative solutions to test grows beyond reasonable computational means. Moreover, the optimality conditions of [7] contain a small but crucial gap. In fact, they are correct only if one assumes that $q_{y_1} > 0$ for all y_1. These optimality conditions are ambiguous for a test q_{y_1, y_2} such that $q_{y_1} = 0$ for some y_1, and do not specify whether such values of y_1 can be omitted, or whether it suffices to satisfy the

Manuscript received November 7, 2000; revised December 3, 2002. This work was supported in part by the National Science Foundation under Grants EIA-9986057 and EIA-0080134, the University of California MICRO Program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, Inc., and Qualcomm, Inc. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000.

The authors are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, CA 93106-9560 USA (e-mail: ertem@ece.ucsb.edu; rose@ece.ucsb.edu).

Communicated by P. Narayan, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2003.810627

conditions for an arbitrary $q_{y_2|y_1}$. While in most problems involving joint distributions this would merely be a formal objection, we shall see that in this case it has not only theoretical but also major practical consequences. These two observations motivate our work.

In Sections III and IV, we temporarily fix $N = 2$, and derive the main results for two-layer scalable source coding. This choice is made in order to minimize the notational burden and simplify the presentation. It is easy to see that the main tools used in the proofs of theorems and lemmas in those sections are by no means restrictive to $N = 2$, i.e., they are easily generalizable to $N > 2$. In Section V, we present extensions to N-layer scalable source coding, accompanied by brief proof sketches where appropriate.

In Section III, we present and prove convergence and optimality of an approach based on an iterative algorithm for the computation of the RD bound. Not surprisingly, the algorithm is a generalization of the well-known Blahut–Arimoto algorithm [3], [1], which was introduced for the nonscalable ($N = 1$) RD computation problem. The proposed algorithm is initialized with an arbitrary reproduction distribution with $q_{y_1, \ldots, y_N} > 0$ for all $(y_1, \ldots, y_N)$, and monotonically approaches the optimal reproduction distribution $q^*_{y_1, \ldots, y_N}$. We also discuss two relevant problems in Section III, namely, the sufficient set of Lagrangian parameters to compute the entire RD bound, and the problem of efficient navigation on the RD surface to compute a particular point.

In Section IV, we fill the above mentioned gap in the optimality conditions by carefully handling the cases where $q^*_{y_1, \ldots, y_i} = 0$. The revised optimality conditions surprisingly require us to try *all conceivable* $q_{y_{i+1}, \ldots, y_N|y_1, \ldots, y_i}$ for each $q^*_{y_1, \ldots, y_i} = 0$, in order to ensure optimality of a tentative solution. In most cases, this requirement represents an impractical computational burden. Alternatively, beside its obvious use to directly compute the RD bound, the proposed iterative algorithm may be used to test tentative solutions while circumventing this problem. To test whether a given $q_{y_1, \ldots, y_N}$ is nearly optimal, one can simply perturb $q_{y_1, \ldots, y_N}$ (to ensure positivity everywhere) and run the iterative algorithm. This fact suggests that for $N > 1$, the proposed algorithm is more useful than the optimality conditions themselves, as even checking the optimality of a *guessed* solution would normally require the utilization of an iterative algorithm. More importantly, the algorithm is, to the best of our knowledge, the only existing tool to find the global optimum $q^*_{y_1, \ldots, y_N}$, in general.

II. Preliminaries and Notation

Let $\{X_t\}_{t=1}^{\infty}$ be a sequence of independent and identically distributed (i.i.d.) random variables with marginal distribution p_x. Throughout this work, we assume the source alphabet $\mathcal{X}$, and the reproduction alphabets $\mathcal{Y}_i$, $i = 1, \ldots, N$, are finite. Let $d_i(x, y_i)$: $\mathcal{X} \times \mathcal{Y}_i \longrightarrow [0, \infty)$ denote the ith-layer distortion measure, which extends to blocks of length n as

$$d_i(x^n, y_i^n) = \frac{1}{n}\sum_{t=1}^{n} d_i(x_t, y_{it}).$$

Typically, the same reproduction alphabets and distortion measures are used throughout the layers, i.e., $\mathcal{Y}_i = \mathcal{Y}$, and $d_i(x, y_i) = d(x, y_i)$ for all $i = 1, \ldots, N$. However, such restrictions are not necessary and are not assumed so as not to obscure the fact that all the results in this work (with the exception of the result in Section III-C) are valid for the general case of possibly different reproduction alphabets and distortion measures.

An N-layer scalable block code $(\phi_1, \ldots, \phi_N, \psi_1, \ldots, \psi_N)$ consists of encoding functions

$$\phi_i: \mathcal{X}^n \longrightarrow \mathcal{M}_i$$

which maps the source to index set $\mathcal{M}_i$, and decoding functions

$$\psi_i: \mathcal{M}_1 \times \cdots \times \mathcal{M}_i \longrightarrow \mathcal{Y}_i^n.$$

A $2N$-tuple of rates and distortions, $\boldsymbol{R} = \{R_i\}_{i=1}^N$ and $\boldsymbol{D} = \{D_i\}_{i=1}^N$, is called *scalably achievable* if for every $\delta > 0$ and sufficiently large n, there exists a block code $(\phi_1, \ldots, \phi_N, \psi_1, \ldots, \psi_N)$ such that

$$\frac{1}{n}\log|\mathcal{M}_1|\cdots|\mathcal{M}_i| \le R_i + \delta$$

and

$$E\, d_i(X^n, \psi_i(\phi_1(X^n), \ldots, \phi_i(X^n)))\} \le D_i + \delta.$$

The region of scalably achievable rates and distortions, as characterized by Rimoldi [14], consists of all $(\boldsymbol{R}, \boldsymbol{D})$ such that there exists a conditional distribution $Q_{y_1, \ldots, y_N|x}$ satisfying

$$\begin{array}{rcl} E\, d_i(X, Y_i) & \le & D_i \\ I(X; Y_1, \ldots, Y_i) & \le & R_i \end{array} \qquad i = 1, \ldots, N. \tag{1}$$

We are interested in computing the boundary of this region, which is easily shown to be convex (see, e.g., [7]). Therefore, by performing the Lagrangian minimization

$$L^*_{\boldsymbol{\alpha}, \boldsymbol{\beta}} = \inf_{Q_{y_1, \ldots, y_N|x}} L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q) \tag{2}$$

where

$$L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q) = \sum_{i=1} \alpha_i I(X; Y_1, \ldots, Y_i) + \beta_i E d_i(X, Y_i)$$

for all positive $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^N$, and $\boldsymbol{\beta} = \{\beta_i\}_{i=1}^N$, we completely traverse the points on the boundary of the N-layer scalable RD surface. Since the minimization above is over a compact set and $L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q)$ is a continuous function of $Q_{y_1, \ldots, y_N|x}$, the minimum is achieved by a distribution $Q^*_{y_1, \ldots, y_N|x}$, and we may formally replace the infimum by a minimum. For a given $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, let $(\boldsymbol{R}, \boldsymbol{D})_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$ denote the point corresponding to $Q^*_{y_1, \ldots, y_N|x}$. It then follows that the vector $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ may be interpreted [7] as the normal of the hyperplane supporting the achievability region at $(\boldsymbol{R}, \boldsymbol{D})_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$.

As an aside, we note that in many cases of interest, we do not have the freedom to choose both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The following examples of standard practical considerations illustrate how $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ may in fact be fixed by the scenario.

- There are N channels operating at a fixed rate vector $\boldsymbol{r}$, where the ith channel carries the incremental description of the ith layer. We define $R_i = \sum_{j=1}^{i} r_i$, and denote by p_i the probability that the user accesses only the first i layers (due to limitations of the connection). Here, it is reasonable to minimize the expected distortion observed by the user, i.e.,

$$D_{\rm avg}(\boldsymbol{R}) \triangleq \min_{Q_{y_1,\ldots,y_N|x}} D_{\rm avg}(Q) = \min_{Q_{y_1,\ldots,y_N|x}} \sum_{i=1}^{N} p_i E\, d_i(X, Y_i) \tag{3}$$

where the minimization is over $Q_{y_1,\ldots,y_N|x}$ such that

$$I(X; Y_1, \ldots, Y_i) \le R_i, \qquad \text{for } i = 1, \ldots, N.$$

We can compute $D_{\rm avg}(\boldsymbol{R})$ by solving (2) at $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ such that $\beta_i = p_i$. For a given $\boldsymbol{\alpha}$, the resulting optimal point $(\boldsymbol{R}_{\boldsymbol{\alpha}}, D_{\rm avg}(\boldsymbol{R}_{\boldsymbol{\alpha}}))$ is a point whose normal to the region of all achievable $(\boldsymbol{R}, D_{\rm avg})$ is in the direction of $(\alpha_1, \ldots, \alpha_N, 1)$.

- N successive descriptions of the source with a prespecified distortion vector $\boldsymbol{D}$ are needed. We denote by p_i the probability that by transmitting only the first i layers, the user is satisfied. In this case, to minimize the expected load of the channel, one must solve

$$R_{\rm avg}(\boldsymbol{D}) \triangleq \min_{Q_{y_1,\ldots,y_N|x}} R_{\rm avg}(Q) = \min_{Q_{y_1,\ldots,y_N|x}} \sum_{i=1}^{N} p_i I(X; Y_1, \ldots, Y_i) \tag{4}$$

where the minimization is over $Q_{y_1,\ldots,y_N|x}$ such that $E\, d_i(X, Y_i) \le D_i$ for $i = 1, \ldots, N$. The solution is obtained by solving (2) at $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ such that $\alpha_i = p_i$. For a given $\boldsymbol{\beta}$, the resulting optimal point $(R_{\rm avg}(\boldsymbol{D}_{\boldsymbol{\beta}}), \boldsymbol{D}_{\boldsymbol{\beta}})$ is a point whose normal to the region of all achievable $(R_{\rm avg}, \boldsymbol{D})$ is in the direction of $(1, \beta_1, \ldots, \beta_N)$.

We return to the Lagrangian of (2) and expand the expression for $L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q)$ to obtain

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q) = \sum_{x} \sum_{y_1,\ldots,y_N} p_x Q_{y_1,\ldots,y_N|x} \sum_{i=1}^{N} \cdot \left\{ \alpha_i \log \frac{Q_{y_1,\ldots,y_i|x}}{Q_{y_1,\ldots,y_i}} + \beta_i\, d_i(x, y_i) \right\} \tag{5}$$

where $Q_{y_1,\ldots,y_i}$ is the marginal distribution corresponding to $Q_{y_1,\ldots,y_i|x}$, i.e.,

$$Q_{y_1,\ldots,y_i} = \sum_{x} p_x Q_{y_1,\ldots,y_i|x}.$$

We will also find useful the functional $L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q)$ defined as

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q) \triangleq \sum_{x} \sum_{y_1,\ldots,y_N} p_x Q_{y_1,\ldots,y_N|x} \sum_{i=1}^{N} \cdot \left\{ \alpha_i \log \frac{Q_{y_1,\ldots,y_i|x}}{q_{y_1,\ldots,y_i}} + \beta_i\, d_i(x, y_i) \right\} \tag{6}$$

where $q_{y_1,\ldots,y_N}$ is a *free* distribution, i.e., not necessarily equal to the true marginal $Q_{y_1,\ldots,y_N}$. However, the equality $q_{y_1,\ldots,y_N} = Q_{y_1,\ldots,y_N}$ obviously yields

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q) = L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q).$$

Let $\mathcal{D}(p_z \| q_z)$ denote the standard divergence (or the Kullback–Leibler distance) between distributions, i.e.,

$$\mathcal{D}(p_z \| q_z) = \sum_z p_z \log \frac{p_z}{q_z}. \tag{7}$$

Motivated by the form of (6), we also define the "weighted scalable" divergence between distributions $p_{y_1,\ldots,y_N}$ and $q_{y_1,\ldots,y_N}$ as

$$\mathcal{D}_{\boldsymbol{\alpha}}(p_{y_1,\ldots,y_N} \| q_{y_1,\ldots,y_N}) \triangleq \sum_{i=1}^{N} \alpha_i \mathcal{D}(p_{y_1,\ldots,y_i} \| q_{y_1,\ldots,y_i}) \tag{8}$$

and between distributions $p_{x,y_1,\ldots,y_N}$ and $q_{x,y_1,\ldots,y_N}$ as

$$\overline{\mathcal{D}}_{\boldsymbol{\alpha}}(p_{x,y_1,\ldots,y_N} \| q_{x,y_1,\ldots,y_N}) \triangleq \mathcal{D}_{[0\,\boldsymbol{\alpha}]}(p_{x,y_1,\ldots,y_N} \| q_{x,y_1,\ldots,y_N}) = \sum_{i=1}^{N} \alpha_i \mathcal{D}(p_{x,y_1,\ldots,y_i} \| q_{x,y_1,\ldots,y_i}). \tag{9}$$

Note from the foregoing that for a general distribution $p_{y_1,\ldots,y_N}$ (or $p_{y_1,\ldots,y_N|x}$), we use $p_{y_1,\ldots,y_i}$ (or $p_{y_1,\ldots,y_i|x}$) to denote the corresponding distribution obtained by summing over $y_{i+1}, \ldots, y_N$.

III. Minimization of the RD Lagrangian

In this section and in Section IV, we temporarily fix $N = 2$, and derive the main results for two-layer scalable source coding. In Section V, we present extensions to N-layer scalable source coding.

A. Alternating Minimization Lemmas

We begin by applying a technique introduced by Blahut [3], to recast the problem as a double minimization.

Lemma 1:

$$L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \min_{Q_{y_1,y_2|x}} \min_{q_{y_1,y_2}} L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q). \tag{10}$$

Proof: It is straightforward to show from (6) that

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q) = L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q) + \mathcal{D}_{\boldsymbol{\alpha}}(Q_{y_1,y_2} \| q_{y_1,y_2}).$$

Hence, $L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q) \ge L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q)$ with equality if and only if $q_{y_1,y_2} = Q_{y_1,y_2}$. Therefore,

$$\min_{q_{y_1,y_2}} L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q) = L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q). \qquad \square$$

Since the inner minimization in (10) is always finite, we may reverse the order of minimization to obtain the following corollary.

Corollary 1:

$$L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \min_{q_{y_1,y_2}} \min_{Q_{y_1,y_2|x}} L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q). \tag{11}$$

In the following lemma, we determine the optimal conditional distribution $Q^*_{y_1,y_2|x}(q)$ for the inner minimization in (11), i.e.,

$$Q^*(q) \triangleq \arg \min_{Q_{y_1,y_2|x}} L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q).$$

Lemma 2: $Q^*_{y_1, y_2|x}(q)$ is given by

$$Q^*_{y_1, y_2|x}(q) = \frac{q_{y_1, y_2} e^{-\beta_1' d_1(x,y_1) - \beta_2' d_2(x,y_2) - \alpha_1' \log f_1(x,y_1)}}{f_0(x)} \tag{12}$$

where

$$f_0(x) = \sum_{y_1, y_2} q_{y_1, y_2} e^{-\beta_1' d_1(x, y_1) - \beta_2' d_2(x, y_2) - \alpha_1' \log f_1(x, y_1)} \tag{13}$$

$$f_1(x, y_1) = \sum_{y_2} q_{y_2|y_1} e^{-\beta_2' d_2(x, y_2)} \tag{14}$$

and

$$\beta_1' = \frac{\beta_1}{\alpha_1 + \alpha_2}, \quad \beta_2' = \frac{\beta_2}{\alpha_2}, \quad \alpha_1' = \frac{\alpha_1}{\alpha_1 + \alpha_2}. \tag{15}$$

Note that $f_0(x)$ can be simplified further

$$\begin{aligned} f_0(x) &= \sum_{y_1, y_2} q_{y_1, y_2} e^{-\beta_1' d_1(x, y_1) - \beta_2' d_2(x, y_2) - \alpha_1' \log f_1(x, y_1)} \\ &= \sum_{y_1} q_{y_1} e^{-\beta_1' d_1(x, y_1) - \alpha_1' \log f_1(x, y_1)} \\ &\quad \cdot \sum_{y_2} q_{y_2|y_1} e^{-\beta_2' d_2(x, y_2)} \\ &= \sum_{y_1} q_{y_1} e^{-\beta_1' d_1(x, y_1) - \alpha_1' \log f_1(x, y_1)} f_1(x, y_1) \\ &= \sum_{y_1} q_{y_1} e^{-\beta_1' d_1(x, y_1) + (1-\alpha_1') \log f_1(x, y_1)}. \end{aligned} \tag{16}$$

Also, observe that the corresponding distribution $Q^*_{y_1|x}(q)$ is given by the formula

$$Q^*_{y_1|x}(q) = \frac{q_{y_1} e^{-\beta_1' d_1(x,y_1) + (1-\alpha_1') \log f_1(x,y_1)}}{f_0(x)}. \tag{17}$$

Now we proceed with the proof of Lemma 2.

Proof: The claim will follow immediately once we show

$$L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q) = L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*(q), q) + \overline{\mathcal{D}}_{\boldsymbol{\alpha}}(p_x Q_{y_1, y_2|x} \| p_x Q^*_{y_1, y_2|x}(q))$$

because then $L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q) \geq L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*(q), q)$, with equality if and only if $Q = Q^*(q)$. Toward this end, we first expand $L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*(q), q)$

$$\begin{aligned} L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*(q), q) &= \sum_{x, y_1, y_2} p_x Q^*_{y_1, y_2|x}(q) \\ &\quad \cdot \left\{ \alpha_1 \log \frac{Q^*_{y_1|x}(q)}{q_{y_1}} + \alpha_2 \log \frac{Q^*_{y_1,y_2|x}(q)}{q_{y_1,y_2}} \right. \\ &\quad \left. + \beta_1 d_1(x, y_1) + \beta_2 d_2(x, y_2) \right\} \\ &= -(\alpha_1 + \alpha_2) \sum_x p_x \log f_0(x) \\ &\quad + (\alpha_1 - \alpha_1 \alpha_1' - \alpha_2 \alpha_1') \\ &\quad \cdot \sum_{x, y_1} p_x Q^*_{y_1|x}(q) \log f_1(x, y_1) \\ &\quad + (\beta_1 - \alpha_1 \beta_1' - \alpha_2 \beta_1') \\ &\quad \cdot \sum_{x, y_1} p_x Q^*_{y_1|x}(q) d_1(x, y_1) \\ &\quad + (\beta_2 - \alpha_2 \beta_2') \\ &\quad \cdot \sum_{x, y_2} p_x Q^*_{y_2|x}(q) d_2(x, y_2) \\ &= -(\alpha_1 + \alpha_2) \sum_x p_x \log f_0(x) \end{aligned}$$

where the last equality is justified by observing the identities

$$\beta_1'(\alpha_1 + \alpha_2) = \beta_1 \tag{18}$$

$$\beta_2' \alpha_2 = \beta_2 \tag{19}$$

$$\alpha_1(1 - \alpha_1') = \alpha_2 \alpha_1'. \tag{20}$$

We also use the identities (18) –(20) to simplify the following expansion:

$$\begin{aligned} &\overline{\mathcal{D}}_{\boldsymbol{\alpha}}(p_x Q_{y_1, y_2|x} \| p_x Q^*_{y_1, y_2|x}(q)) \\ &\quad = (\alpha_1 + \alpha_2) \sum_x p_x \log f_0(x) + \overline{\mathcal{D}}_{\boldsymbol{\alpha}}(p_x Q_{y_1, y_2|x} \| p_x q_{y_1, y_2}) \\ &\qquad + \alpha_1 \sum_{x, y_1} p_x Q_{y_1|x} \\ &\qquad \cdot \{\beta_1' d_1(x, y_1) - (1 - \alpha_1') \log f_1(x, y_1)\} \\ &\qquad + \alpha_2 \sum_{x, y_1, y_2} p_x Q_{y_1, y_2|x} \\ &\qquad \cdot \{\beta_1' d_1(x, y_1) + \beta_2' d_2(x, y_2) + \alpha_1' \log f_1(x, y_1)\} \\ &\quad = (\alpha_1 + \alpha_2) \sum_x p_x \log f_0(x) + L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q) \end{aligned}$$

which yields the desired result. □

Corollary 2:

$$L^*_{\boldsymbol{\alpha}, \boldsymbol{\beta}} = \min_{q_{y_1, y_2}} -(\alpha_1 + \alpha_2) \sum_x p_x \log f_0(x). \tag{21}$$

B. An Iterative Algorithm

An immediate algorithm motivated by the alternating minimization lemmas above is the following.

1) Initialize with $q^{(0)}_{y_1, y_2} > 0$ for all $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$. Set $n = 1$.
2) Compute
$$Q^{(n)}_{y_1, y_2|x} = \frac{q^{(n-1)}_{y_1,y_2} e^{-\beta_1' d_1(x,y_1) - \beta_2' d_2(x,y_2) - \alpha_1' \log f_1^{(n-1)}(x,y_1)}}{f_0^{(n-1)}(x)} \tag{22}$$
where $f_0^{(n-1)}(x)$ and $f_1^{(n-1)}(x, y_1)$ are computed by substituting $q^{(n-1)}_{y_1, y_2}$ into (13) and (14), respectively.
3) Compute
$$q^{(n)}_{y_1, y_2} = \sum_x p_x Q^{(n)}_{y_1, y_2|x}. \tag{23}$$
4) $n \leftarrow n + 1$. Go to 2) if not converged.

The construction by alternating minimization ensures that

$$\begin{aligned} L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(1)}, q^{(0)}\right) &\geq L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(1)}, q^{(1)}\right) \\ &\geq L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(2)}, q^{(1)}\right) \\ &\geq \cdots \end{aligned}$$

and given the bound $L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q) \geq L^*_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$, the above sequence must converge. We next show that it converges to $L^*_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$. In fact, this algorithm is a generalization of the well-known Blahut–Arimoto algorithm [3], [1] which computes the RD curve for the case of nonscalable coding. Convergence of the Blahut–Arimoto algorithm to the minimum Lagrangian was proven by Csiszár [5], and his proof is extended here to scalable RD.

Theorem 1: The sequence $q^{(0)}, Q^{(1)}, q^{(1)}, Q^{(2)}, \ldots,$ as generated by the iterative algorithm, converges to

$$(Q^*, q^*) = \arg \min_{Q_{y_1, y_2|x}, q_{y_1, y_2}} L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q).$$

Proof: Let us fix

$$q_{y_1, y_2} = Q_{y_1, y_2} = \sum_x p_x Q_{y_1, y_2|x}$$

throughout the proof. (Note that normally, q is treated as a free distribution anywhere else in the paper.) It is convenient to consider the posterior (backward) probabilities

$$r_{x|y_1, y_2} = \frac{p_x Q_{y_1,y_2|x}}{q_{y_1,y_2}}. \tag{24}$$

Similarly, define $r^{(n)}_{x|y_1, y_2}$ as

$$r^{(n)}_{x|y_1, y_2} = \frac{p_x Q^{(n)}_{y_1,y_2|x}}{q^{(n)}_{y_1,y_2}}. \tag{25}$$

We will now show that

$$\begin{aligned}\mathcal{D}_{\boldsymbol{\alpha}}&\left(q_{y_1, y_2} \middle\| q^{(n-1)}_{y_1, y_2}\right) - \mathcal{D}_{\boldsymbol{\alpha}}\left(q_{y_1, y_2} \middle\| q^{(n)}_{y_1, y_2}\right)\\ &= L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q)\\ &\quad + \overline{\mathcal{D}}_{\boldsymbol{\alpha}}\left(r_{x|y_1, y_2} q_{y_1, y_2} \middle\| r^{(n)}_{x|y_1, y_2} q_{y_1, y_2}\right).\end{aligned} \tag{26}$$

Toward that end, let us expand $\mathcal{D}(r_{x|y_1} q_{y_1} \| r^{(n)}_{x|y_1} q_{y_1})$ first

$$\begin{aligned}\mathcal{D}&\left(r_{x|y_1} q_{y_1} \middle\| r^{(n)}_{x|y_1} q_{y_1}\right)\\ &= -\mathcal{D}\left(q_{y_1} \middle\| q^{(n)}_{y_1}\right) + \sum_{x, y_1} p_x Q_{y_1|x} \log \frac{Q_{y_1|x}}{Q^{(n)}_{y_1|x}}\\ &= -\mathcal{D}\left(q_{y_1} \middle\| q^{(n)}_{y_1}\right) + \sum_{x, y_1} p_x Q_{y_1|x}\\ &\qquad \cdot \log \frac{Q_{y_1|x} f_0^{(n-1)}(x)}{q^{(n-1)}_{y_1} e^{-\beta_1' d_1(x, y_1) + (1-\alpha_1') \log f_1^{(n-1)}(x, y_1)}}\\ &= \mathcal{D}\left(q_{y_1} \middle\| q^{(n-1)}_{y_1}\right) - \mathcal{D}\left(q_{y_1} \middle\| q^{(n)}_{y_1}\right) + \sum_x p_x \log f_0^{(n-1)}(x)\\ &\quad + \sum_{x, y_1} p_x Q_{y_1|x} \left\{\log \frac{Q_{y_1|x}}{q_{y_1}} + \beta_1' d_1(x, y_1) - (1-\alpha_1')\right.\\ &\qquad \left. \cdot \log f_1^{(n-1)}(x, y_1)\right\}.\end{aligned}$$

We similarly expand $\mathcal{D}(r_{x|y_1, y_2} q_{y_1, y_2} \| r^{(n)}_{x|y_1, y_2} q_{y_1, y_2})$ as

$$\begin{aligned}\mathcal{D}&\left(r_{x|y_1, y_2} q_{y_1, y_2} \middle\| r^{(n)}_{x|y_1, y_2} q_{y_1, y_2}\right)\\ &= \mathcal{D}\left(q_{y_1, y_2} \middle\| q^{(n-1)}_{y_1, y_2}\right) - \mathcal{D}\left(q_{y_1, y_2} \middle\| q^{(n)}_{y_1, y_2}\right)\\ &\quad + \sum_x p_x \log f_0^{(n-1)}(x) + \sum_{x, y_1, y_2} p_x Q_{y_1, y_2|x}\\ &\quad \cdot \left\{\log \frac{Q_{y_1,y_2|x}}{q_{y_1,y_2}} + \beta_1' d_1(x, y_1) + \beta_2' d_2(x, y_2)\right.\\ &\qquad \left. + \alpha_1' \log f_1^{(n-1)}(x, y_1)\right\}.\end{aligned}$$

After noting from (21) that

$$L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(n)}, q^{(n-1)}\right) = -(\alpha_1 + \alpha_2) \sum_x p_x \log f_0^{(n-1)}(x)$$

the claim (26) follows. This implies, in particular, that

$$\begin{aligned}L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}&\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*, q^*)\\ &\le \mathcal{D}_{\boldsymbol{\alpha}}\left(q^*_{y_1, y_2} \middle\| q^{(n-1)}_{y_1, y_2}\right) - \mathcal{D}_{\boldsymbol{\alpha}}\left(q^*_{y_1, y_2} \middle\| q^{(n)}_{y_1, y_2}\right)\end{aligned} \tag{27}$$

and, hence,

$$\begin{aligned}\sum_{n=1}^{K}&\left[L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*, q^*)\right]\\ &\le \mathcal{D}_{\boldsymbol{\alpha}}\left(q^*_{y_1, y_2} \middle\| q^{(0)}_{y_1, y_2}\right) - \mathcal{D}_{\boldsymbol{\alpha}}\left(q^*_{y_1, y_2} \middle\| q^{(K)}_{y_1, y_2}\right)\\ &\le \mathcal{D}_{\boldsymbol{\alpha}}\left(q^*_{y_1, y_2} \middle\| q^{(0)}_{y_1, y_2}\right).\end{aligned} \tag{28}$$

Since we select $q^{(0)}_{y_1, y_2} > 0$ everywhere, and the reproduction space $\mathcal{Y}_1 \times \mathcal{Y}_2$ is finite, the right-hand side is bounded from above, i.e., $\mathcal{D}_{\boldsymbol{\alpha}}(q^*_{y_1, y_2} \| q^{(0)}_{y_1, y_2}) < \infty$. Therefore, as $K \to \infty$, we have an infinite series of positive terms on the left-hand side, which is guaranteed to be convergent. This, in turn, implies that the argument of the series converges to 0, i.e.,

$$\lim_{n\to\infty} \left[L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*, q^*)\right] = 0. \tag{29}$$

□

Remark: In fact, the inequality

$$\begin{aligned}L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}&\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q)\\ &\le \mathcal{D}_{\boldsymbol{\alpha}}\left(q_{y_1, y_2} \middle\| q^{(n-1)}_{y_1, y_2}\right) - \mathcal{D}_{\boldsymbol{\alpha}}\left(q_{y_1, y_2} \middle\| q^{(n)}_{y_1, y_2}\right)\end{aligned}$$

can be recast as

$$\begin{aligned}L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}&\left(Q^{(n)}, q^{(n-1)}\right) + L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q, q^{(n)}\right)\\ &\le L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q, q^{(n-1)}\right) + L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q)\end{aligned}$$

where $q_{y_1, y_2} = \sum_x p_x Q_{y_1, y_2|x}$. Since

$$L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(n)}, q^{(n)}\right) \le L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q^{(n)}, q^{(n-1)}\right)$$

and also $L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q)$ is minimized by $q_{y_1, y_2} = \sum_x p_x Q_{y_1, y_2|x}$, this in turn implies

$$\begin{aligned}L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}&\left(Q^{(n)}, q^{(n)}\right) + L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q, q^{(n)}\right)\\ &\le L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}\left(Q, q^{(n-1)}\right) + L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q)\end{aligned} \tag{30}$$

for arbitrary $Q_{y_1, y_2|x}$ and q_{y_1, y_2}. The inequality (30) shows that, for all Q, the distance function $L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q, q)$ satisfies the so-called "five points property," which was introduced in the classical paper of Csiszár and Tusnády [6]. Therefore (cf. [6, Theorem 2 and subsequent remarks]), convergence to $L^*_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$ is guaranteed by (30). Further, convergence of the pair $(Q^{(n)}, q^{(n)})$ to *some* (Q^*, q^*) for which

$$L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*, q^*) = L^*_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$$

follows from [6, Theorem 3].

Corollary 3: From (27), it follows that

$$\begin{aligned}L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}&\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(Q^*, q^*)\\ &\le (\alpha_1 + \alpha_2) \log \left[\max_{y_1, y_2} \frac{q^{(n)}_{y_1,y_2}}{q^{(n-1)}_{y_1,y_2}}\right].\end{aligned} \tag{31}$$

To see that (31) indeed holds, observe that

$$\frac{q_{y_1}^{(n)}}{q_{y_1}^{(n-1)}} = \frac{\sum_{y_2} q_{y_1,y_2}^{(n)}}{\sum_{y_2} q_{y_1,y_2}^{(n-1)}} \leq \max_{y_2} \frac{q_{y_1,y_2}^{(n)}}{q_{y_1,y_2}^{(n-1)}}$$

where the last is a well-known inequality for strictly positive numbers.

The significance of this corollary is that we can use (31) as a stopping criterion for the algorithm. Even though, in general, we do not know the optimal distributions Q^* and q^*, we can always calculate the right-hand side in (31), and use it as an upper bound on the distance of the achieved Lagrangian from its optimal value $L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}}$.

C. Sufficient Set of Lagrangian Parameters

It is clear from construction that the degree of freedom for the general scenario of (2) is 3, instead of 4 (or $2N-1$, instead of $2N$, in general). That is because if we multiply $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by a constant c, that would not change the direction of the normal vector to the supporting hyperplane, and hence, $(\boldsymbol{R}, \boldsymbol{D})_{c\boldsymbol{\alpha}, c\boldsymbol{\beta}} = (\boldsymbol{R}, \boldsymbol{D})_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$. So, as discussed in [7], it is possible to constrain $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by $\alpha_1 + \alpha_2 + \beta_1 + \beta_2 = 1$ or by $\alpha_1 = 1$, etc. In this subsection, we show that in most interesting cases, we can further reduce the set of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ which suffices to compute the entire RD surface.

Let $\boldsymbol{\mathcal{L}}$ be defined as

$$\boldsymbol{\mathcal{L}} = \left\{ (\boldsymbol{\alpha}, \boldsymbol{\beta}) : \frac{\beta_2}{\alpha_2} \geq \frac{\beta_1}{\alpha_1} \geq \gamma_0 \right\} \quad (32)$$

where γ_0 is the first "critical slope" for the computation of the nonscalable RD function $R(D)$ (see [2] and [15]), i.e., the most negative slope γ where

$$D_\gamma = D_{\max} \triangleq \min_{y \in \mathcal{Y}} \sum_x p_x\, d(x, y).$$

Here, $D_{\max}$ denotes the supremum of all distortion values satisfying $R(D) > 0$, or in other words, the minimum of all values satisfying $R(D) = 0$.

Theorem 2: Let $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Y}$ and

$$d_1(x, y) = d_2(x, y) = d(x, y).$$

Then, for each quadruple (R_1, R_2, D_1, D_2) on the RD surface with $R_2 > R_1 > 0$ and $D_{\max} > D_1 > D_2$, there exists a normal vector $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\mathcal{L}}$ for which $(\boldsymbol{R}, \boldsymbol{D}) = (\boldsymbol{R}, \boldsymbol{D})_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$.

Proof: If (R_1, R_2, D_1, D_2) is achievable by a two-layered scalable coder, then so are the quadruples

$$(R_1, R_1, D_1, D_1)$$
$$(R_2, R_2, D_2, D_2)$$
$$(0, R_2, D_{\max}, D_2).$$

Let the minimization (2) with Lagrangian parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ yield (R_1, R_2, D_1, D_2). Hence,

$$L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \alpha_1 R_1 + \alpha_2 R_2 + \beta_1 D_1 + \beta_2 D_2.$$

Let us define $L'_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and $L''_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ as the values of the Lagrangians corresponding to parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and quadruples

$$(R_1, R_1, D_1, D_1) \quad \text{and} \quad (R_2, R_2, D_2, D_2)$$

respectively. Then, by definition

$$0 \leq L'_{\boldsymbol{\alpha},\boldsymbol{\beta}} - L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \alpha_2(R_1 - R_2) + \beta_2(D_1 - D_2) \quad (33)$$

$$0 \leq L''_{\boldsymbol{\alpha},\boldsymbol{\beta}} - L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \alpha_1(R_2 - R_1) + \beta_1(D_2 - D_1) \quad (34)$$

and, therefore,

$$\frac{\beta_2}{\alpha_2}(D_1 - D_2) \geq R_2 - R_1 \geq \frac{\beta_1}{\alpha_1}(D_1 - D_2)$$

which implies $\frac{\beta_2}{\alpha_2} \geq \frac{\beta_1}{\alpha_1}$.

Similarly, let $L'''_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ correspond to the Lagrangian for the point $(0, R_2, D_{\max}, D_2)$. Then

$$L'''_{\boldsymbol{\alpha},\boldsymbol{\beta}} - L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \beta_1(D_{\max} - D_1) - \alpha_1 R_1 \geq 0$$

which implies

$$\begin{aligned}\frac{\beta_1}{\alpha_1}(D_{\max} - D_1) &\geq R_1 \\ &\geq R(D_1) \\ &= R(D_1) - R(D_{\max}) \\ &\geq \gamma_0(D_{\max} - D_1)\end{aligned}$$

where the last inequality is from the definition of γ_0 and from convexity of $R(D)$. The result $\frac{\beta_1}{\alpha_1} \geq \gamma_0$ follows. □

We conclude that by solving (2) for $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\mathcal{L}}$, we traverse every nontrivial point (i.e., points such that $R_2 > R_1 > 0$ and $D_{\max} > D_1 > D_2$) on the RD surface.

D. An Example Scenario

We consider a scalable coder that consists of only two layers: a base layer and an enhancement layer. The base layer is to operate at rate R_1 and the enhancement layer at rate R_2, where $R_2 > R_1$. We assume that, with probability p, the receiver only receives the base-layer information. The objective is to determine the minimum achievable average distortion $D_{\text{avg}}(\boldsymbol{R})$ for all values $\boldsymbol{R} = (R_1, R_2)$. To attack this problem, we consider (2), $\beta_1 = p$, $\beta_2 = 1 - p$, and run the proposed iterative algorithm for all α_1 and α_2 such that $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\mathcal{L}}$. For each choice of α_1 and α_2, we get the point on the surface of $D_{\text{avg}}(\boldsymbol{R})$ whose normal is parallel to $[1, \alpha_1, \alpha_2]$.

In Fig. 1, we present the $D_{\text{avg}}(\boldsymbol{R})$ function for a discrete memoryless source (DMS) with source and reproduction alphabet $\{0, 1, 2\}$ and $p_x = \{s, 1-2s, s\}$. The distortion measure is given by $d(x, y) = |x - y|$. This configuration is known as the Gerrish example [9]. (It appeared in previous discussions of successive refinement, e.g., [7], [8].) We have chosen $s = 0.45$, in which case there exists a nonempty subset of distortion values where it is not possible to achieve successive refinement without rate loss (see [8]). We also have chosen $p = 0.3$. Note that, for $R_1 = R_2$, the function specializes to the standard nonscalable distortion-rate curve.

E. Navigation Over the RD Surface

In the two-layer example given earlier, we have demonstrated how to determine $D_{\text{avg}}(R_1, R_2)$ for all R_1, R_2. However, if the objective is to determine $D_{\text{avg}}(R_1, R_2)$ for a *specific* pair (R_1, R_2), it would be clearly inefficient to simply run the algorithm for all $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\mathcal{L}}$. In the nonscalable case, this task is relatively easy: run the Blahut–Arimoto algorithm for some initial slope parameter for $R(D)$, and if the resulting rate is greater or less than the desired rate, decrease or increase the slope parameter, respectively, until the attained rate converges to the desired rate. Typically, such practical search should involve an interval partitioning technique. However, in the scalable case, we have to control a "normal vector" instead of a slope parameter, and

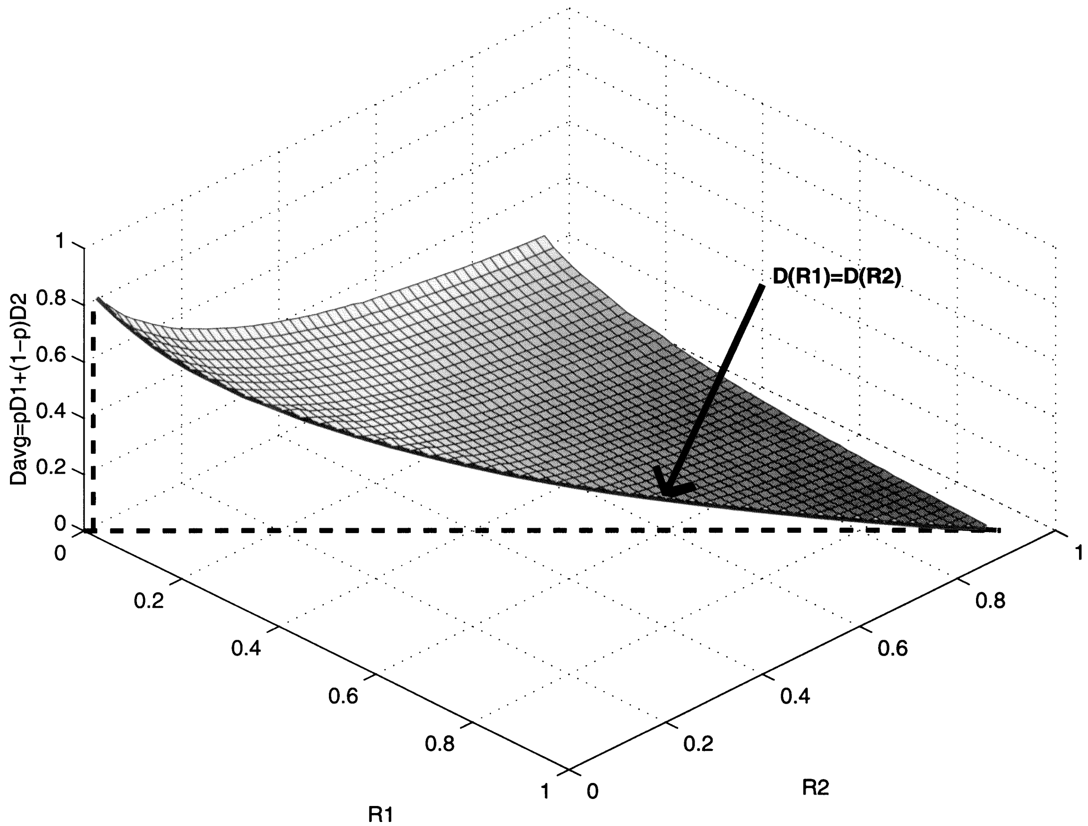

Fig. 1. Visualization of $D_{\text{avg}}(\boldsymbol{R})$.

hence we have a continuum of possible "directions" to update $\boldsymbol{\alpha}$. We propose the following method.

- Start with an arbitrary $\boldsymbol{\alpha}^{(0)} > 0$ and $k = 0$. Let $\boldsymbol{R} = (R_1, R_2)$ be the target rate vector.
- Repeat until convergence:
  — run the iterative algorithm to determine $\boldsymbol{R}_{\boldsymbol{\alpha}^{(k)}}$;
  — update $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + \epsilon_k(\boldsymbol{R}_{\boldsymbol{\alpha}^{(k)}} - \boldsymbol{R})$, where $\epsilon_k > 0$;
  — $k \leftarrow k + 1$.

The following theorem is concerned with the convergence of the method.

Theorem 3: If $D_{\text{avg}}(\boldsymbol{R})$ is a strictly convex function in the set $\{\boldsymbol{R}\colon D_{\text{avg}}(\boldsymbol{R}) > 0\}$, then in the preceding method, ϵ_k can be chosen properly so as to satisfy

$$\lim_{k\to\infty} \boldsymbol{R}_{\boldsymbol{\alpha}^{(k)}} = \boldsymbol{R}.$$

Before we present the proof, we give the following lemma, which is a simple application of the duality principle in optimization theory [13].

Lemma 3:

$$D_{\text{avg}}(\boldsymbol{R}) = \max_{\alpha_1, \alpha_2} \left\{ L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} - \alpha_1 R_1 - \alpha_2 R_2 \right\} \tag{35}$$

for all $\boldsymbol{R}$. The maximum is achieved by (α_1^*, α_2^*), where $(1, \alpha_1^*, \alpha_2^*)$ is the normal to $D_{\text{avg}}(\boldsymbol{R})$. Moreover, $L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is a concave function of $\boldsymbol{\alpha}$, and hence (35) is a convex optimization problem.

Proof: Follows from the convexity of $D_{\text{avg}}(\boldsymbol{R})$. For elaboration, see [4]. □

Proof of Theorem 3: Since $D_{\text{avg}}(\boldsymbol{R})$ is a strictly convex function in the set $\{\boldsymbol{R}\colon D_{\text{avg}}(\boldsymbol{R}) > 0\}$, it follows that for each $\boldsymbol{\alpha}$, there is a unique $\boldsymbol{R}_{\boldsymbol{\alpha}}$, and that $L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is differentiable with respect to (w.r.t.) $\boldsymbol{\alpha}$ everywhere. It further can be shown by algebraic manipulation that

$$\nabla_{\boldsymbol{\alpha}} L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \boldsymbol{R}_{\boldsymbol{\alpha}}.$$

Hence, the update $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + \epsilon_k(\boldsymbol{R}_{\boldsymbol{\alpha}^{(k)}} - \boldsymbol{R})$ is the update for a gradient-ascent algorithm to maximize (35). ϵ_k can be chosen as in any line-search method (e.g., see [13]). □

Remark: If $D_{\text{avg}}(\boldsymbol{R})$ is not strictly convex at the target rate vector $\boldsymbol{R}$, then it must be planar, and there are infinitely many solutions for the corresponding $\boldsymbol{\alpha}$. Under that circumstance, the above algorithm breaks down. However, in nonscalable RD analysis, this phenomenon is seen only when the size of the reproduction alphabet is larger than that of the source (see [2, Example 2.7.3]), and this fact is most likely to be generalizable to scalable RD. On the other hand, the most interesting cases of RD analysis typically involve identical source and reproduction alphabets.

As a final note, it should be emphasized that by replacing $\boldsymbol{R}$ with $\boldsymbol{D}$, D_{avg} with R_{avg}, and $\boldsymbol{\alpha}$ with $\boldsymbol{\beta}$ in the above discussion we obtain the equivalent results for "target $\boldsymbol{D}$ navigation."

IV. Necessary and Sufficient Conditions for Optimality

In this section, we investigate analytical evaluation of the scalable RD function. The first attempt toward this direction was made in [7, Sec. V]. There, the solution to minimization of (2) is presented as the solution for a system of equations and inequalities in $q^*_{y_1, y_2}$. The conditions for optimality are embedded in the system. (See [7, inequalities (25) and (26)]) However, those conditions are ambiguous for a test q_{y_1, y_2} such that $q_{y_1} = 0$ at some y_1. It is not clear if one should ignore those values of y_1, or whether satisfaction of the conditions for an arbitrary $q_{y_2|y_1}$ indicates optimality of q_{y_1, y_2}.

This motivates us to derive the optimality conditions by carefully handling the cases where $q_{y_1} = 0$. The optimality conditions presented in this section are indeed identical to the those derived in [7] for the case where the distribution under test satisfies $q_{y_1} > 0$ for all y_1. However, the revised optimality conditions surprisingly require us to try *all* conceivable $q_{y_2|y_1}$ for each $q_{y_1} = 0$. As we show by an example later in this section, if we ignore those y_1 such that $q_{y_1} = 0$, or rely on the conditions of [7] after trial of an arbitrary $q_{y_2|y_1}$, then a suboptimal tentative solution will be declared optimal.

Recall that (21) provides us with an equivalent minimization problem for $L^*_{\boldsymbol{\alpha}, \boldsymbol{\beta}}$, which is more compact and has fewer parameters to optimize. We refer to the expression to be minimized in (21)

$$L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(q) \triangleq -(\alpha_1 + \alpha_2) \sum_x p_x \log f_0(x) \tag{36}$$

as the *Helmholtz free energy* of the system, in order to emphasize relations with statistical physics. The analogy is direct and simple in the nonscalable case (see [15] and [16]) where the RD Lagrangian is, in fact, the Helmholtz free energy of a physical system whose energy is the distortion and whose inverse temperature is the slope parameter. Thus, finding a point on the RD curve is equivalent to reaching isothermal equilibrium in the physical analogy. In the scalable case the description is similar, albeit complicated by the existence of multiple temperatures and the interaction between layers.

In this section, we apply calculus of variations on (21) to derive the necessary and sufficient conditions for a given q_{y_1, y_2} to minimize the free energy.

For a better intuition about the interpretation of the functions $f_0(x)$ and $f_1(x, y_1)$, we observe that the optimal random encoding function in the first and second layers are given by

$$Q^*_{y_1|x}(q) = \frac{q_{y_1} e^{-\beta_1' d_1(x,y_1) + (1-\alpha_1') \log f_1(x,y_1)}}{f_0(x)} \tag{37}$$

and

$$Q^*_{y_2|x, y_1}(q) = \frac{Q^*_{y_1, y_2|x}(q)}{Q^*_{y_1|x}(q)} = \frac{q_{y_2|y_1} e^{-\beta_2' d_2(x,y_2)}}{f_1(x, y_1)} \tag{38}$$

respectively. These forms, together with (16) and (14), give $f_0(x)$ and $f_1(x, y_1)$ the interpretation of partition functions, in the statistical physics analogy.

A necessary condition for the optimality of a given q_{y_1, y_2} is that all perturbations increase the free energy. We formalize an ϵ-perturbation of q_{y_1, y_2} as

$$q^\epsilon_{y_1, y_2} \triangleq (1-\epsilon) q_{y_1, y_2} + \epsilon r_{y_1, y_2}$$

where $\epsilon > 0$, and require

$$\left. \frac{\partial L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(q^\epsilon)}{\partial \epsilon} \right|_{\epsilon = 0} \geq 0 \tag{39}$$

for all admissible reproduction distributions r_{y_1, y_2}. Further, if the free energy is a convex function of q, then the "first-order" condition (39) is sufficient as well. Therefore, we start by proving the convexity of the free energy in Section IV-A; we continue by deriving the optimality conditions in Section IV-B; we then show in Section IV-C that a suboptimal tentative solution could be declared optimal by the conditions of [7]; and end this section with a demonstration of the difficulty in evaluation of the (Kuhn–Tucker) optimality conditions.

A. Convexity of the Free Energy

To prove the convexity of the free energy (36), it suffices to show that $f_0(x)$ is a concave function of q_{y_1, y_2}. Toward this end, we prove the following lemma.

Lemma 4: Let $p_{y, z}$ be a joint distribution, and let $0 \leq \gamma \leq 1$. If $g(\cdot)$ is a strictly positive concave function, then the function

$$h(p_{y, z}) = p_y g(p_{z|y})^\gamma$$

is concave in $p_{y, z}$.

Proof: Let $r_{y, z} = \lambda p_{y, z} + (1-\lambda) q_{y, z}$. Then

$$r_y = \lambda p_y + (1-\lambda) q_y$$

and

$$\begin{aligned} r_{z|y} &= \frac{r_{y,z}}{r_y} \\ &= \frac{\lambda p_{y,z} + (1-\lambda) q_{y,z}}{r_y} \\ &= \lambda \frac{p_y}{r_y} p_{z|y} + (1-\lambda) \frac{q_y}{r_y} q_{z|y}. \end{aligned}$$

Now, if $g(\cdot)$ is concave, obviously so is $g(\cdot)^\gamma$ with $0 \leq \gamma \leq 1$, and

$$\begin{aligned} h(r_{y, z}) &= r_y g(r_{z|y})^\gamma \\ &= r_y g\left(\lambda \frac{p_y}{r_y} p_{z|y} + (1-\lambda) \frac{q_y}{r_y} q_{z|y}\right)^\gamma \\ &\geq r_y \left(\lambda \frac{p_y}{r_y} g(p_{z|y})^\gamma + (1-\lambda) \frac{q_y}{r_y} g(q_{z|y})^\gamma\right) \\ &= \lambda p_y g(p_{z|y})^\gamma + (1-\lambda) q_y g(q_{z|y})^\gamma \\ &= \lambda h(p_{y, z}) + (1-\lambda) h(q_{y, z}). \qquad \square \end{aligned}$$

We use this lemma to prove that $f_0(x)$ is concave in q_{y_1, y_2} as follows. From the definition of $f_0(x)$ in (16), we observe that in order to apply the lemma it suffices to show the concavity of $f_1(x, y_1)$ in $q_{y_2|y_1}$. Indeed, by (14), $f_1(x, y_1)$ is a *linear* function of, and consequently concave in, the conditional distribution $q_{y_2|y_1}$.

B. Derivation of the Optimality Conditions

We proceed by expanding (39). First, observe that, we have not yet clarified the definition of $f_1(x, y_1)$ for values of y_1 such that $q_{y_1} = 0$. In the iterative algorithm of Section III-B, this ambiguity does not cause a problem, because we start with a distribution $q^{(0)}_{y_1, y_2} > 0$ everywhere, and the iterations never reach a case where $q^{(n)}_{y_1} = 0$. Moreover, in the free energy formula (36), it does not matter how, or even whether or not, $f_1(x, y_1)$ is defined for y_1 such that $q_{y_1} = 0$. We choose to leave it undefined for those cases, and recast the formula of $f_0(x)$ as

$$f_0(x) = \sum_{y_1: q_{y_1} > 0} q_{y_1} e^{-\beta'_1 d_1(x, y_1) + (1-\alpha'_1) \log f_1(x, y_1)}. \quad (40)$$

Now, denote by $f^\epsilon_0(x)$ and $f^\epsilon_1(x, y_1)$ the perturbed versions of $f_0(x)$ and $f_1(x, y_1)$, respectively. The domain of $f^\epsilon_1(x, y_1)$ is potentially larger than that of $f_1(x, y_1)$, due to the inclusion of the possibly nonempty set

$$\mathcal{Y}_0(q, r) \triangleq \{y_1: q_{y_1} = 0, r_{y_1} > 0\}.$$

For all $y_1 \in \mathcal{Y}_0(q, r)$, note that $q^\epsilon_{y_1, y_2} = \epsilon r_{y_1, y_2}$, and, therefore, $q^\epsilon_{y_2|y_1} = r_{y_2|y_1}$. Hence, for $y_1 \in \mathcal{Y}_0(q, r)$, we have

$$f^\epsilon_1(x, y_1) = g_1(x, y_1) \triangleq \sum_{y_2} r_{y_2|y_1} e^{-\beta'_2 d_2(x, y_2)}. \quad (41)$$

Remark the independence of $g_1(x, y_1)$ from ϵ.

We will also need the following limits:

$$\lim_{\epsilon \to 0} f^\epsilon_1(x, y_1) = \begin{cases} f_1(x, y_1), & q_{y_1} > 0 \\ g_1(x, y_1), & y_1 \in \mathcal{Y}_0(q, r) \end{cases} \quad (42)$$

and

$$\lim_{\epsilon \to 0} f^\epsilon_0(x) = f_0(x). \quad (43)$$

We are finally ready to present the explicit conditions resulting from (39) as a theorem.

Theorem 4—Kuhn–Tucker Optimality Conditions: A given q_{y_1, y_2} is optimal if and only if

$$v(y_1, y_2) \le 1, \qquad \forall\, y_1, y_2 \text{ with } q_{y_1} > 0 \quad (44)$$

$$w(y_1) \le 1, \qquad \forall\, y_1 \text{ with } q_{y_1} = 0, \text{ and } \forall\, r_{y_2|y_1} \quad (45)$$

where

$$v(y_1, y_2) \triangleq \sum_x \frac{p_x e^{-\beta'_1 d_1(x,y_1) - \beta'_2 d_2(x,y_2) - \alpha'_1 \log f_1(x,y_1)}}{f_0(x)} \quad (46)$$

and

$$w(y_1) \triangleq \sum_x \frac{p_x e^{-\beta'_1 d_1(x,y_1) + (1-\alpha'_1) \log g_1(x,y_1)}}{f_0(x)}. \quad (47)$$

Proof: With the help of (41)–(43), we evaluate the derivatives related to (39) as

$$\left.\frac{\partial L_{\boldsymbol{\alpha}, \boldsymbol{\beta}}(q^\epsilon)}{\partial \epsilon}\right|_{\epsilon=0} = -(\alpha_1 + \alpha_2) \sum_x \frac{p_x}{f_0(x)} \left.\frac{\partial f^\epsilon_0(x)}{\partial \epsilon}\right|_{\epsilon=0}$$

$$\begin{aligned}\left.\frac{\partial f^\epsilon_0(x)}{\partial \epsilon}\right|_{\epsilon=0} &= -f_0(x) + \sum_{y_1: q_{y_1} > 0} r_{y_1} e^{-\beta'_1 d_1(x, y_1) + (1-\alpha'_1) \log f_1(x, y_1)} \\ &\quad + \sum_{y_1 \in \mathcal{Y}_0(q, r)} r_{y_1} e^{-\beta'_1 d_1(x, y_1) + (1-\alpha'_1) \log g_1(x, y_1)} \\ &\quad + (1-\alpha'_1) \sum_{y_1: q_{y_1} > 0} q_{y_1} e^{-\beta'_1 d_1(x, y_1) - \alpha'_1 \log f_1(x, y_1)} \\ &\quad \cdot \left.\frac{\partial f^\epsilon_1(x, y_1)}{\partial \epsilon}\right|_{\epsilon=0}\end{aligned}$$

and

$$\left.\frac{\partial f^\epsilon_1(x, y_1)}{\partial \epsilon}\right|_{\epsilon=0} = \frac{1}{q_{y_1}} \sum_{y_2} (r_{y_1, y_2} - r_{y_1} q_{y_2|y_1}) e^{-\beta'_2 d_2(x, y_2)}$$

where the last equality is for y_1 such that $q_{y_1} > 0$ only. Combining all these derivative evaluations, we expand (39) as

$$\begin{aligned} 1 \ge &\sum_{y_1 \in \mathcal{Y}_0(q, r)} r_{y_1} w(y_1) \\ &+ (1-\alpha'_1) \sum_{y_1: q_{y_1} > 0} \sum_{y_2} r_{y_1, y_2} v(y_1, y_2) \\ &+ \alpha'_1 \sum_{y_1: q_{y_1} > 0} \sum_{y_2} r_{y_1} q_{y_2|y_1} v(y_1, y_2). \end{aligned} \quad (48)$$

Therefore, q_{y_1, y_2} is optimal if and only if (48) is satisfied for all r_{y_1, y_2}.

Now, if (44) and (45) are both satisfied, then so is (48) for any r_{y_1, y_2}. Conversely, if (48) is true for all r_{y_1, y_2}, then by choosing deterministic distributions for r_{y_1, y_2}, i.e.,

$$r_{y_1, y_2} = \begin{cases} 1, & y_1, y_2 = z_1, z_2 \\ 0, & \text{otherwise} \end{cases}$$

we can show that

$$(\alpha_1 + \alpha_2) \ge \alpha_2 v(y_1, y_2) + \alpha_1 \sum_{y'_2} q_{y'_2|y_1} v(y_1, y'_2) \quad (49)$$

for all y_1, y_2 with $q_{y_1} > 0$. Multiplying both sides by $q_{y_2|y_1}$ and summing over y_2, we get

$$1 \ge \sum_{y_2} q_{y_2|y_1} v(y_1, y_2).$$

On the other hand, since from (46), $\sum_{y_1, y_2} q_{y_1, y_2} v(y_1, y_2) = 1$, we have equality above for all y_1 such that $q_{y_1} > 0$. Therefore, from (49), (44) follows. Similarly, if we substitute in (48) $r_{y_1} = 1$ at some y_1 such that $q_{y_1} = 0$, we precisely obtain (45). □

Remarks:

1) Condition (44) is equivalent to the optimality conditions claimed in [7]. However, as we show in the next subsection, satisfaction of (44) is not sufficient for optimality of q_{y_1, y_2}.

2) Checking condition (44) involves only substitution of the candidate q_{y_1, y_2} in (46), and hence, is straightforward. However, checking (45) involves substitution of all conceivable $r_{y_2|y_1}$ in (47) for *every* y_1 such that $q_{y_1} = 0$, which is computationally catastrophic. Alternatively, one can think of maximizing $w(y_1)$ over all $r_{y_2|y_1}$ and checking whether the attained maximum is less than 1. Using (41) and (47), together with Lemma 4, this is easily shown to be a convex optimization problem, i.e., $w(y_1)$ is a concave function of $r_{y_2|y_1}$. Solving this new problem, we get the optimality conditions

$$\sum_x \frac{p_x e^{-\beta_1' d_1(x,y_1)-\beta_2' d_2(x,y_2)-\alpha_1' \log g_1(x,y_1)}}{f_0(x)} \le w(y_1)$$

with equality whenever $r_{y_2|y_1} > 0$. Note that this condition is not easy to check either. However, it leads us to an alternative set of Kuhn–Tucker conditions, which we present below.

Lemma 5—Alternative Optimality Conditions: A given q_{y_1,y_2} is optimal if and only if there exists a $q_{y_2|y_1}$ for all $q_{y_1} = 0$ such that when the domain of $f_1(x, y_1)$ is extended to all x, y_1, we obtain

$$v(y_1, y_2) \le v(y_1) \le 1 \tag{50}$$

for all y_1, y_2, where $v(y_1, y_2)$ is defined as in (46), and

$$v(y_1) \triangleq \sum_x \frac{p_x e^{-\beta_1' d_1(x,y_1)+(1-\alpha_1')\log f_1(x,y_1)}}{f_0(x)}. \tag{51}$$

Checking this version of the optimality conditions is not easier than the original form in Theorem 4, as in order to verify the optimality of q_{y_1,y_2}, one has to *find* some artificial $q_{y_2|y_1}$ for all $q_{y_1} = 0$ such that (50) is satisfied. Note that once the domain of $f_1(x, y_1)$ is assumed to be $\mathcal{X} \times \mathcal{Y}_1$, definitions of $f_1(x, y_1)$ and $g_1(x, y_1)$ become identical if $r_{y_2|y_1}$ in (41) is replaced by $q_{y_2|y_1}$.

Another alternative for checking optimality is to run an iterative algorithm that is guaranteed to converge to $r^*_{y_2|y_1}$, i.e., the maximizer of $w(y_1)$. In Section III-B, we already provided an iterative algorithm which can be utilized for this purpose: perturb the given q_{y_1,y_2} to ensure that $q_{y_1,y_2} > 0$, run the proposed iterative algorithm until convergence, and check if the free energy of the solution is arbitrarily close to that of the original q_{y_1,y_2}.

Although the most prominent application of the Kuhn–Tucker conditions is to verify optimality of a given q_{y_1,y_2}, the alternative version given by (50) is also useful in dealing with solutions for product sources and sum distortion measures [2, Sec. 2.8]. Specifically, using Lemma 5, it is easy to prove that the boundary point $(\boldsymbol{R}, \boldsymbol{D})_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ can be computed by summing up the RD vectors $(\boldsymbol{R}_k, \boldsymbol{D}_k)_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ which are computed for the component problems for $k = 1, \ldots, K$ independently.

C. Distinction With Previous Results

In this subsection, we will demonstrate that the simpler conditions of [7] are not sufficient for optimality. Let us reconsider the example of Section III-D. Now let $p = 0.1803$, $\alpha_1 = 0.8197$, $\alpha_2 = 1.9126$. Running the algorithm proposed in Section III-B, we can obtain the correct solution for $q^*_{y_1,y_2}$: see the matrix at the bottom of the page, where columns and rows represent the first- and the second-layer symbols, respectively. Note that the values of $q^*_{y_1,y_2}$ are numerically computed in finite precision, and are hence approximate. However, the error is negligibly small since $v(y_1, y_2) \le 1 + 1.3323 \times 10^{-15}$ for all y_1, y_2. Also, using the stopping criterion (31), we ensure that the Lagrangian cost $L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(q) = 0.8285$ is at most 3.6402×10^{-15} away from the true optimal $L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}}$.

Now consider a different tentative solution q_{y_1,y_2}

	0	1	2
0	1/2	0	0
1	0	0	0
2	1/2	0	0

This simple solution yields $v(0, y_2) \le 1$ for all y_2. However, testing (45) for the value $w(1)$ with the choice of conditional distribution $r_{y_2|1} = \{1/4, 0, 3/4\}$, we obtain $w(1) = 1.01330374$. Hence, the Kuhn–Tucker conditions are violated and this tentative solution is *suboptimal*. Also, the Lagrangian cost $L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(q) = 0.8465$ is considerably larger than the optimal cost $L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. However, the conditional distribution choice $r_{y_2|1} = \{1, 0, 0\}$ and $r_{y_2|2} = \{1, 0, 0\}$ yields $w(1) = 0.97562679$ and $w(2) = 0.9575114$, respectively. Therefore, had we either ignored $w(1)$ and $w(2)$ or evaluated them only for the latter choice of conditional distribution, we would have considered this suboptimal solution to be optimal.

D. The Difficulty in Evaluating the Kuhn–Tucker Conditions

Let us reconsider the example of Section IV-C with the tentative solution q_{y_1,y_2}

	0	1	2
0	0.00674733662552	0.49323604588916	0
1	0	0	0
2	0.00206622293909	0.49795039454623	0

When we test $v(y_1, y_2)$ for $y_1 \in \{0, 1\}$ and $y_2 \in \{0, 1, 2\}$, we see that $v(y_1, y_2) \le 1$, and, hence, the first part of the Kuhn–Tucker conditions (44) is not violated.[1] However, to test the second part (45), we have to evaluate $w(2)$ for all $r_{y_2|2}$. In Fig. 2, we present $w(2)$ as a function of $r_{0|2}$ and $r_{1|2}$. (Note that we made use of the fact that $r_{0|2} + r_{1|2} + r_{2|2} = 1$ to draw the figure in three dimensions.) We see that $w(2)$ exceeds 1 for a region of conditional distributions $r_{y_2|2}$, and, therefore, q_{y_1,y_2} is not optimal. Furthermore, since this region is very small, the time complexity of checking (45) is in the same order as that of an exhaustive search over the entire simplex of $r_{y_2|2}$. One could

[1]Similarly, due to finite precision, we have $v(y_1, y_2) \le 1 + 2.2204 \times 10^{-15}$ for $y_1 \in \{0, 1\}$ and for all y_2.

	0	1	2
0	0.09749464111237	0.37302587944721	0.02947947944042
1	0	0	0
2	0.02947947944042	0.37302587944721	0.09749464111237

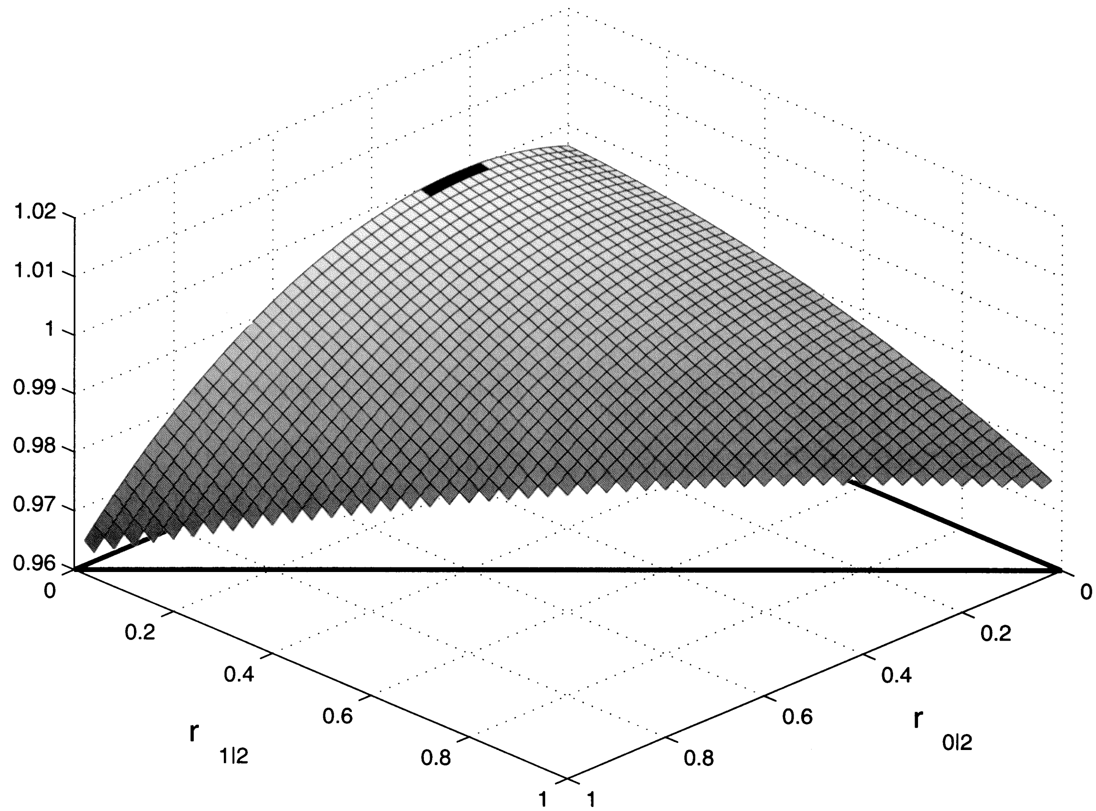

Fig. 2. Kuhn–Tucker condition $w(2)$ as a function of $r_{0|2}$ and $r_{1|2}$. The small region where $w(2) > 1$ is filled in.

argue that we can use some iterative maximization algorithm to find the region where $w(2) > 1$. However, rather than employ an iterative algorithm for the evaluation of $w(2)$ at a *single* tentative solution, we might as well use the iterative algorithm proposed in Section III-B, which *directly* produces the optimal solution.

V. Generalization to N-Layers

In this section, we demonstrate that the results generalize to N layers, where $N > 2$. Most proofs are fundamentally similar to the $N = 2$ case, except that they require cumbersome notation. We provide proof sketches when needed.

Lemma 1—N-Layer:

$$L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \min_{Q_{y_1,\dots,y_N|x}} \min_{q_{y_1,\dots,y_N}} L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q). \tag{52}$$

Proof: Follows after observing

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q) = L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q) + \mathcal{D}_{\boldsymbol{\alpha}}(Q_{y_1,\dots,y_N} \| q_{y_1,\dots,y_N}). \quad \square$$

In Lemma 2, we determine the optimal conditional distribution $Q^*_{y_1,\dots,y_N|x}(q)$ for the outer minimization of Lemma 1.

Lemma 2—N-Layer:

$$Q^*_{y_1,\dots,y_N|x}(q) = \frac{q_{y_1,\dots,y_N} e^{\left\{-\sum_{i=1}^N \beta'_i d_i(x,y_i) - \sum_{i=1}^N \alpha'_i \log f_i(x,y_1,\dots,y_i)\right\}}}{f_0(x)} \tag{53}$$

where $f_N(x, y_1, \dots, y_N) = 1$, and

$$f_i(x, y_1, \dots, y_i) = \sum_{y_{i+1},\dots,y_N} q_{y_{i+1},\dots,y_N|y_1,\dots,y_i} \cdot e^{\left\{-\sum_{j=i+1}^N \beta'_j d_j(x,y_j) - \sum_{j=i+1}^N \alpha'_j \log f_j(x,y_1,\dots,y_j)\right\}} \tag{54}$$

for $0 < i < N$. Also,

$$f_0(x) = \sum_{y_1,\dots,y_N} q_{y_1,\dots,y_N} \cdot e^{\left\{-\sum_{i=1}^N \beta'_i d_i(x,y_i) - \sum_{i=1}^N \alpha'_i \log f_i(x,y_1,\dots,y_i)\right\}} \tag{55}$$

and

$$\beta'_i = \frac{\beta_i}{\alpha_i + \cdots + \alpha_N}, \qquad \alpha'_i = \frac{\alpha_i}{\alpha_i + \cdots + \alpha_N}. \tag{56}$$

Proof: It suffices to show by substitution that

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q) = L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q^*(q), q) + \overline{\mathcal{D}}_{\boldsymbol{\alpha}}\left(p_x Q_{y_1,\dots,y_N|x} \middle\| p_x Q^*_{y_1,\dots,y_N|x}(q)\right). \quad \square$$

Corollary 2—N-Layer:

$$L^*_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \min_{q_{y_1,\dots,y_N}} -(\alpha_1 + \cdots + \alpha_N) \sum_x p_x \log f_0(x). \tag{57}$$

Before proceeding further, let us rearrange (54) and (55) into simpler recursive forms

$$f_i(x, y_1, \dots, y_i) = \sum_{y_{i+1}} q_{y_{i+1}|y_1,\dots,y_i} \cdot e^{\left\{-\beta'_{i+1} d_{i+1}(x,y_{i+1}) + (1-\alpha'_{i+1}) \log f_{i+1}(x,y_1,\dots,y_{i+1})\right\}} \tag{58}$$

$$f_0(x) = \sum_{y_1} q_{y_1} e^{\left\{-\beta_1' d_1(x, y_1) + (1-\alpha_1')\log f_1(x, y_1)\right\}}. \quad (59)$$

The above identities may be verified by substituting

$$q_{y_{i+1},\dots,y_N|y_1,\dots,y_i} = q_{y_{i+1}|y_1,\dots,y_i} q_{y_{i+2},\dots,y_N|y_1,\dots,y_{i+1}}$$

$$q_{y_1,\dots,y_N} = q_{y_1} q_{y_2,\dots,y_N|y_1}$$

in (54) and (55), respectively, and taking the summations over $y_{i+2}, \dots, y_N$ and $y_2, \dots, y_N$ first. As in the case of $N = 2$, these alternative forms shed more light on the statistical physics analogy by allowing the observation

$$Q^*_{y_{i+1}|x,y_1,\dots,y_i}(q) = \frac{Q^*_{y_1,\dots,y_{i+1}|x}(q)}{Q^*_{y_1,\dots,y_i|x}(q)} = \frac{q_{y_{i+1}|y_1,\dots,y_i}}{f_i(x, y_1, \dots, y_i)} \cdot e^{\left\{-\beta_{i+1}' d_{i+1}(x, y_{i+1}) + (1-\alpha_{i+1}')\log f_{i+1}(x, y_1,\dots,y_{i+1})\right\}} \quad (60)$$

for all $0 < i < N$.

Theorem 1—N-Layer: The sequence

$$q^{(0)}, Q^{(1)}, q^{(1)}, Q^{(2)}, \dots$$

as generated by the iterative algorithm

- initialize with $q^{(0)}_{y_1,\dots,y_N} > 0$ for all $y_1, \dots, y_N$;
- iterate until convergence:
  — compute $Q^{(n)}_{y_1,\dots,y_N|x} = Q^*_{y_1,\dots,y_N|x}(q)$;
  — compute $q^{(n)}_{y_1,\dots,y_N} = \sum_x p_x Q^{(n)}_{y_1,\dots,y_N|x}$;

converges to

$$(Q^*, q^*) = \arg\min_{Q_{y_1,\dots,y_N|x},\, q_{y_1,\dots,y_N}} L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q, q).$$

The proof of this theorem is identical to that of Theorem 1, and therefore is omitted. It suffices to show

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q^*, q^*) \le \mathcal{D}_{\boldsymbol{\alpha}}\left(q^*_{y_1,\dots,y_N} \,\middle\|\, q^{(n-1)}_{y_1,\dots,y_N}\right) - \mathcal{D}_{\boldsymbol{\alpha}}\left(q^*_{y_1,\dots,y_N} \,\middle\|\, q^{(n)}_{y_1,\dots,y_N}\right)$$

which also yields the following stopping criterion as a corollary.

Corollary 3—N-Layer:

$$L_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left(Q^{(n)}, q^{(n-1)}\right) - L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(Q^*, q^*) \le (\alpha_1 + \cdots + \alpha_N)\log\left[\max_{y_1,\dots,y_N} \frac{q^{(n)}_{y_1,\dots,y_N}}{q^{(n-1)}_{y_1,\dots,y_N}}\right].$$

To generalize the results of Section III-C, we redefine $\boldsymbol{\mathcal{L}}$ as

$$\boldsymbol{\mathcal{L}} = \left\{(\boldsymbol{\alpha}, \boldsymbol{\beta}) \colon \frac{\beta_N}{\alpha_N} \ge \frac{\beta_{N-1}}{\alpha_{N-1}} \ge \cdots \ge \frac{\beta_1}{\alpha_1} \ge \gamma_0\right\}. \quad (61)$$

Theorem 2—N-Layer: Let $\mathcal{Y}_i = \mathcal{Y}$, and $d_i(x, y_i) = d(x, y_i)$, for $i = 1, \dots, N$. Then, for each RD vector pair $(\boldsymbol{R}, \boldsymbol{D})$ on the RD surface with

$$R_N > \cdots > R_1 > 0 \quad \text{and} \quad D_{\max} > D_1 > \cdots > D_N$$

there exists a normal vector $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\mathcal{L}}$ for which $(\boldsymbol{R}, \boldsymbol{D}) = (\boldsymbol{R}, \boldsymbol{D})_{\boldsymbol{\alpha},\boldsymbol{\beta}}$.

Proof: If $(\boldsymbol{R}, \boldsymbol{D})$ is achievable by an N-layered scalable coder, then so are

$$(R_1, \dots, R_{i-1}, R_i, R_i, R_{i+2}, \dots, R_N, D_1, \dots, D_{i-1}, D_i, D_i, D_{i+2}, \dots, D_N)$$

for all $0 < i \le N$, and

$$(0, R_2, \dots, R_N, D_{\max}, D_2, \dots, D_N).$$

The rest of the proof is identical to that of Theorem 2. □

The discussion in Section III-E on the navigation on the RD surface is, in fact, valid for general N, and, therefore, we will not repeat the arguments, theorems, and lemmas for $N > 2$.

Convexity of the free energy (57) is provable by arguments similar to those of (21). From (58) and (59), it suffices to show that $f_1(x, y_1)$ is concave in $q_{y_2,\dots,y_N|y_1}$, for which it suffices to show that $f_2(x, y_1, y_2)$ is concave in $q_{y_3,\dots,y_N|y_1,y_2}$, and so on. This sequence of reasoning leads to the concavity of $f_0(x)$ in $q_{y_1,\dots,y_N}$, after observing that $f_{N-1}(x, y_1, \dots, y_{N-1})$ is a linear (and hence a concave) function of $q_{y_N|y_1,\dots,y_{N-1}}$.

Lemma 5–N-Layer: A given $q_{y_1,\dots,y_N}$ is optimal if and only if there exists a $q_{y_{i+1}|y_1,\dots,y_i}$ for all $q_{y_1,\dots,y_i} = 0$, such that when the domain of $f_i(x, y_1, \dots, y_i)$ is extended to all $x, y_1, \dots, y_i$, we obtain

$$v(y_1, \dots, y_N) \le v(y_1, \dots, y_{N-1}) \le \cdots \le v(y_1) \le 1 \quad (62)$$

for all $y_1, \dots, y_N$, where $v(y_1, \dots, y_i)$ is defined as

$$v(y_1, \dots, y_i) = \sum_x \frac{p_x}{f_0(x)} e^{-\sum_{j=1}^{i} \beta_j' d_j(x, y_j)} \cdot e^{\left\{-\sum_{j=1}^{i-1} \alpha_j' \log f_j(x, y_1,\dots,y_j) + (1-\alpha_i')\log f_i(x, y_1,\dots,y_i)\right\}}. \quad (63)$$

Proof: As in the proof of Theorem 4, we expand

$$\left.\frac{\partial L_{\boldsymbol{\alpha},\boldsymbol{\beta}}(q^\epsilon)}{\partial \epsilon}\right|_{\epsilon=0} \ge 0$$

by substituting the derivatives

$$\left.\frac{\partial f_0^\epsilon(x)}{\partial \epsilon}\right|_{\epsilon=0} \quad \text{and} \quad \left.\frac{\partial f_i^\epsilon(x, y_1, \dots, y_i)}{\partial \epsilon}\right|_{\epsilon=0}$$

where $q^\epsilon = (1-\epsilon)q + \epsilon r$, $f_0^\epsilon(x)$, and $f_i^\epsilon(x, y_1, \dots, y_i)$ are the perturbed versions of q, $f_0(x)$, and $f_i(x, y_1, \dots, y_i)$, respectively. To proceed further, we need to introduce the functional $g_i(x, y_1, \dots, y_i)$ for $(y_1, \dots, y_i)$ such that $q_{y_1,\dots,y_i} = 0$, but $r_{y_1,\dots,y_i} > 0$

$$g_i(x, y_1, \dots, y_i) = \sum_{y_{i+1}} r_{y_{i+1}|y_1,\dots,y_i} \cdot e^{\left\{-\beta_{i+1}' d_{i+1}(x, y_{i+1}) + (1-\alpha_{i+1}')\log g_{i+1}(x, y_1,\dots,y_{i+1})\right\}}$$

with $g_N(x, y_1, \dots, y_N) = 1$ as the recursion termination rule. Also, let

$$w(y_1, \dots, y_i) = \sum_x \frac{p_x}{f_0(x)} e^{\left\{-\sum_{j=1}^{i} \beta_j' d_j(x, y_j)\right\}} \cdot e^{\left\{-\sum_{j=1}^{i-1} \alpha_j' \log f_j(x, y_1,\dots,y_j) + (1-\alpha_i')\log g_i(x, y_1,\dots,y_i)\right\}}.$$

After some manipulation, the expansion becomes

$$(\alpha_1 + \cdots + \alpha_N) \geq \sum_{i=1}^{N} \Bigg\{ \alpha_i \sum_{\substack{y_1, \ldots, y_i: \\ q_{y_1, \ldots, y_i} > 0}} r_{y_1, \ldots, y_i} v(y_1, \ldots, y_i) + (\alpha_i + \cdots + \alpha_N) \sum_{\substack{y_1, \ldots, y_i: \\ q_{y_1, \ldots, y_i} = 0 \\ q_{y_1, \ldots, y_{i-1}} > 0}} r_{y_1, \ldots, y_i} w(y_1, \ldots, y_i) \Bigg\}$$

from which, with similar arguments as in the proof of Theorem 4, we conclude

$$v(y_1, \ldots, y_N) = 1, \ \forall (y_1, \ldots, y_N) \text{ with } q_{y_1, \ldots, y_N} > 0 \tag{64}$$

$$w(y_1, \ldots, y_i) \leq 1, \ \forall (y_1, \ldots, y_i) \text{ with } q_{y_1, \ldots, y_i} = 0, \ q_{y_1, \ldots, y_{i-1}} > 0, \ \text{and } \forall r_{y_{i+1}, \ldots, y_N | y_1, \ldots, y_i}. \tag{65}$$

Paralleling the remark after Theorem 4, the lemma follows after maximizing $w(y_1, \ldots, y_i)$ in terms of $r_{y_{i+1}, \ldots, y_N | y_1, \ldots, y_i}$. □

VI. Conclusion

We proposed an iterative algorithm for the computation of N-layer scalable RD bound. The algorithm is guaranteed to converge to a solution point on the RD surface, provided that it is initialized with a reproduction distribution that is positive everywhere. We rigorously derived the optimality (Kuhn–Tucker) conditions for the reproduction distribution. To our surprise, the resultant conditions are, in general, computationally impractical to check, in contrast to the case $(N = 1)$ of nonscalable coding. Alternatively, the proposed iterative algorithm may be utilized as an optimality testing procedure by applying it to the perturbed tentative reproduction distribution (perturbation is necessary to ensure that the reproduction is positive everywhere). Hence, the proposed algorithm is more useful than the optimality conditions, in the sense that checking the optimality conditions normally requires an iterative algorithm to be used.

We also derived the sufficient set of Lagrangian parameters to visit all the points on the RD surface, and devised an efficient algorithm for navigating over the RD surface so as to reach a target point. These two problems are relatively easy for nonscalable coding, but complications occur due to the increased dimensionality of the RD surface in the case of scalable coding.

References

[1] S. Arimoto, "An algorithm for computing capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT–18, pp. 14–20, Jan. 1972.

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

[3] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT–18, pp. 460–473, July 1972.

[4] ——, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[5] I. Csiszár, "On the computation of rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT–20, pp. 122–124, Jan. 1974.

[6] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statist. Decisions*, pp. 205–237, 1984. Supplement Issue 1.

[7] M. Effros, "Distortion-rate bounds for fixed- and variable-rate multiresolution source coding," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1887–1910, Sept. 1999.

[8] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[9] A. M. Gerrish, "Estimation of information rates," Ph.D. dissertation, Yale Univ., New Haven, CT, 1963.

[10] V. Koshelev, "Hierarchical coding of discrete sources," *Probl. Pered. Inform.*, vol. 16, no. 3, pp. 31–49, 1980.

[11] ——, "An evaluation of the average distortion for discrete schemes of sequential approximation," *Probl. Pered. Inform.*, vol. 17, no. 3, pp. 20–33, 1981.

[12] L. Lastras and T. Berger, "All sources are nearly successively refinable," *IEEE Trans. Inform. Theory*, vol. 47, pp. 918–926, Mar. 2001.

[13] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984.

[14] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.

[15] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1939–1952, Nov. 1994.

[16] ——, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proc. IEEE*, vol. 86, pp. 2210–2239, Nov. 1998.