On Approximating the Rate Regions for Lossy Source Coding with Coded and Uncoded Side Information

WeiHsin Gu†, Soumya Jana‡, and Michelle Effros†

†Department of Electrical Engineering
California Institute of Technology
Pasadena, California, CA 91125, USA
Email: {wgu, effros}@caltech.edu

‡Department of Electrical Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
Email: jana@uiuc.edu

***Abstract*— We derive new algorithms for approximating the rate regions for a family of source coding problems that includes lossy source coding, lossy source coding with uncoded side information at the receiver (the Wyner-Ziv problem), and an achievability bound for lossy source coding with coded side information at the receiver. The new algorithms generalize a recent approximation algorithm by Gu and Effros from lossless to lossy coding. In each case, prior information theoretic descriptions of the desired regions are available but difficult to evaluate for example sources due to their dependence on auxiliary random variables. Our algorithm builds a linear program whose solution is no less than the desired lower bound and no greater than $(1+\epsilon)$ times that optimal value. These guarantees are met even when the optimal value is unknown. Here $\epsilon > 0$ is a parameter chosen by the user; the algorithmic complexity grows as $O(\epsilon^{-M})$ as ϵ approaches 0, where $M > 4$ is a constant that depends on the source coding problem and the alphabet sizes of the sources.**

I. Introduction

The derivation of rate regions for lossy and lossless source coding problems is a central goal of network source coding theory research. While a network source coding problem is often considered to be solved once an achievable rate region and matching converse are demonstrated, these results become useful in practice only when we can evaluate them for example sources. For some problems, like Slepian and Wolf's lossless multiple access source coding problem [1], evaluating the optimal rate region for example sources is trivial since the information theoretic bound gives an explicit rate region characterization. For other problems, including lossy source coding, lossless source coding with coded side information at the receiver [2],[1] and the family of lossy source coding problems described by Jana and Blahut in [3], the information theoretic characterization describes an optimization problem whose solution is the desired bound. These optimization problems are often difficult to solve for example sources.

[0]This material is based upon work partially supported by NSF Grant No. CCF-0325324 and Caltech's Lee Center for Advanced Networking.

[1]Source coding with coded side information at the receiver may be viewed as a type of lossy coding problem since perfect reconstruction of the side information is not required.

While single-letter characterizations and alphabet-size bounds for auxiliary random variables are often motivated by concerns about rate region evaluation, the evaluation problem itself has received surprisingly little attention in the literature. Most existing algorithms follow the strategy proposed by Blahut [4] and Arimoto [5]. When applied to rate-distortion bound evaluation, this iterative descent approach progressively updates solutions for the marginal $p(z)$ on the reproduction alphabet and the conditional $p(z|x)$ on the reproduction given the source. The convexity of the objective function results in the algorithm's guaranteed convergence to the optimal solution [6]. Calculating the algorithmic complexity of this approach would require a bound on the number of iterations required to achieve convergence to the optimal solution (or a sufficiently accurate approximation).

In [7], Gu and Effros offer an alternative approach for rate region calculation. The proposed algorithm involves building a linear program whose solution approximates the optimal rate region to within a guaranteed factor of $(1+\epsilon)$ times the optimal solution. The goal of achieving $(1+\epsilon)$-accuracy using a polynomial-time algorithm is related to Csiszár and Körner's definition of computability, which they propose as a critical component of any future definition for a single-letter characterization [8, p.259–260].

The algorithm in [7] gives a $(1+\epsilon)$-approximation of the rate region for lossless source coding with coded side information at the decoder [2]; we here generalize that approach to a lossy incast source coding problem described by Jana and Blahut in [3] and the achievable rate region for the lossy coded side information problem described by Berger et al. in [9]. Incast problems are multiple access source coding problems with one or more transmitters and a single receiver that wishes to reconstruct all sources in the network. (Reconstruction of possible receiver side information is trivial.) The lossy incast problem from [3], differs from traditional incast problems in that the sources may be statistically dependent and exactly one source is reconstructed with loss (subject to a distortion constraint) while the others source reconstructions are lossless. The lossy source coding and Wyner-Ziv problems meet this model of lossy incast problems. The rate region for this lossy incast problem relies on a single auxiliary random variable [3]. The achievable rate region for the lossy coded side information problem relies on a pair of auxiliary random variables [9].

Section II describes the algorithmic strategy. Section III describes the approximation algorithm for our lossy incast problem. Since describing the problem in its most general form increases notational complexity without adding much insight, we give details only for the Wyner-Ziv problem. Section IV tackles the lossy coded side information achievability bound using tools developed for the incast problem.

II. Outline of Strategy

In all of the problems studied here, we begin with known information theoretic descriptions that rely on one or more auxiliary random variables. Optimization of each auxiliary random variable requires optimization of that variable's conditional distribution given one or more source random variables. Direct optimization is difficult since the desired rates are not convex or concave in the conditional distributions.

The central observation for our algorithm is that for any fixed conditional distribution on the source given a single auxiliary random variable, all rates and distortions are linear in the auxiliary random variable's marginal distribution. As a result, for any given conditional distribution, we can efficiently optimize the marginal on the auxiliary random variable using a linear program. Since the true conditional distribution of the source given the auxiliary random variable is unknown, we quantize the space of conditional distributions and find the best marginal with respect to a conditional distribution that exhibits each of these quantized distributions as the conditional given some value $z \in \mathcal{Z}$. The solution is at least as good as the solution that would be obtained if we were to first quantize the optimal conditional distribution and then run the linear program for that quantized conditional. As a result, to prove that the algorithm yields a $(1+\epsilon)$ approximation, we need only show that quantizing the optimal conditional distribution on the source given the auxiliary random variable would yield performance within a factor $(1+\epsilon)$ of the optimum.

For any finite alphabet $\mathcal{A}$, we quantize distribution $\{q(a)\}_{a\in\mathcal{A}}$ to distribution $\{\widehat{q}(a)\}_{a\in\mathcal{A}}$ as follows. First, fix parameters $\delta, \eta > 0$ and $c := 1+\eta/|\mathcal{A}|$. These parameters are related to the approximation constant ϵ in a manner described in later sections. Fix $a_0 \in \arg\max_{a\in\mathcal{A}} q(a)$. Then

$$\widehat{q}(a) := \begin{cases} 0 & \text{if } a \neq a_0 \text{ and } q(a) < \delta \\ c^{-n} & \text{if } a \neq a_0,\ q(a) \geq \delta, \text{ and} \\ & \quad c^{-n} \leq q(a) < c^{-n+1} \\ 1 - \sum_{a\neq a_0} \hat{q}(a) & \text{if } a = a_0, \end{cases} \tag{1}$$

which describes one of

$$N(\delta, \eta, |\mathcal{A}|) \leq |\mathcal{A}| \left(\frac{-\log\delta}{\log(1+\frac{\eta}{|\mathcal{A}|})} + 1 \right)^{|\mathcal{A}|-1}$$

distributions. This approach quantizes smaller probability values more finely than larger probability values but maps the smallest probability values to zero. The impact of quantizing the smallest values of $q(a)$ to zero is limited since $q\log(1/q)$ approaches 0 as q approaches 0. The variation in the quantization cell size for $q(a)$ is motivated by Lemma 1.

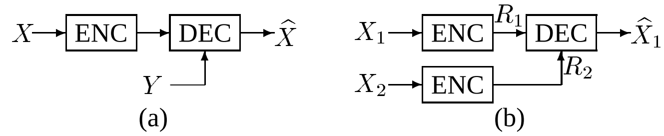

Fig. 1. The (a) Wyner-Ziv and (b) lossy coded side information problems.

Lemma 1: [7, Lemma 1] Given distributions $\{q(a)\}_{a\in\mathcal{A}}$ and $\{\widehat{q}(a)\}_{a\in\mathcal{A}}$ on finite alphabet $\mathcal{A}$. If $|q(a)-\widehat{q}(a)| \leq \epsilon q(a)$ for all $a \in \mathcal{A}$, then

$$|H(q) - H(\hat{q})| \leq \epsilon H(q) + \epsilon \log \frac{e}{1-\epsilon}.$$

III. The Wyner-Ziv Rate Region

Let $\mathcal{X}$ and $\mathcal{Y}$ denote the finite alphabets for sources X and Y. The Wyner-Ziv rate-distortion bound

$$\begin{aligned} R_{X|\{Y\}}(D) &= \min_{Z\in\Psi(X,Y)} I(X;Z|Y) \\ \Psi(X,Y) &:= \Big\{ Z \Big| Z \to X \to Y, \\ &\qquad \exists\phi \text{ s.t. } Ed(X,\psi(Y,Z)) \leq D \Big\}. \end{aligned}$$

specifies the minimal rate for describing source X to a receiver that knows side information Y and reconstructs X with expected distortion no greater than D [10]. (See Fig. 1(a).) The Lagrangian

$$J(\lambda) := I(X;Z|Y) + \lambda \min_{\phi} Ed(X,\psi(Y,Z))) \tag{2}$$

captures the desired constrained optimization.

Let $\mathcal{Z} = \{1,\ldots,N(\delta,\eta,|\mathcal{X}|)\}$ be the alphabet for auxiliary random variable Z, and for each $z \in \mathcal{Z}$ let $\{Q_{X|Z}(x|z)\}_{x\in\mathcal{X}}$ be a distinct distribution from our quantized collection (1). We wish to find the marginal $\{P_Z(z)\}$ that minimizes $J(\lambda)$ for any $\lambda > 0$. Since $I(X;Z|Y) = H(X|Y) - H(X|Y,Z)$ and

$$\begin{aligned} &\min_{\phi} Ed(X,\psi(Y,Z)) \\ &\quad = \min_{\psi} \sum_{z\in\mathcal{Z}} P_Z(z) E[d(X,\psi(Y,Z))|Z=z] \\ &\quad = \sum_{z\in\mathcal{Z}} P_Z(z) \min_{\psi} E[d(X,\psi(Y,z))|Z=z], \end{aligned}$$

and the constraints $\sum_{z\in\mathcal{Z}} P_Z(z) = 1$, $P_Z(z) \geq 0$ for all $z \in \mathcal{Z}$, and $\sum_{z\in\mathcal{Z}} P_Z(z) Q_{X|Z}(x|z) = p(x)$ for all but one $x \in \mathcal{X}$ are all linear functions of $P_Z(z)$, we optimize $\{P_Z(z)\}_{z\in\mathcal{Z}}$ for $\{Q_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ using linear programming. The proof of Theorem 1 appears in the Appendix.

Theorem 1: The proposed algorithm yields a $(1+\epsilon)$-approximation algorithm for the Wyner-Ziv rate region in time $O(\epsilon^{-4(|\mathcal{X}|+1)})$ as ϵ approaches 0.

No matter what the initial size of $\mathcal{Z}$, the solution to the linear program satisfies $P_Z(z) = 0$ for all but $|\mathcal{X}|$ values of

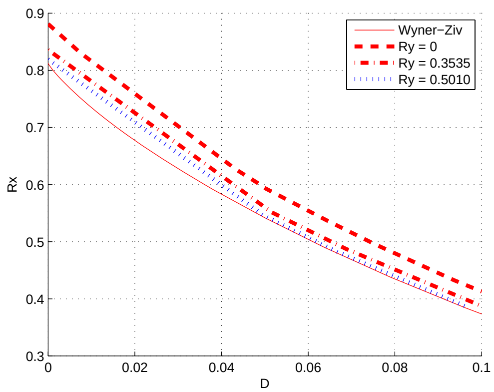

Fig. 2. Example $(1+\epsilon)$ approximations for the Wyner-Ziv and Berger et al. achievable rate regions.

$z \in \mathcal{Z}$ by the following argument. The linear program has $|\mathcal{Z}|$ variables and $|\mathcal{X}| + |\mathcal{Z}|$ constraints. Since there exists a solution for any linear program at a boundary point, there exists an optimal marginal $\{P_Z^*(z)\}$ for which $|\mathcal{Z}|$ constraints are satisfied with equality. At most $|\mathcal{X}| - 1$ are constraints of the form $\sum_{z\in\mathcal{Z}} P_Z^*(z) Q_{X|Z}(x|z) = p(x)$ for some $x \in \mathcal{X}$, and one constraint ensures that $\sum_{x\in\mathcal{X}} P_Z^*(z) = 1$. The remaining $|\mathcal{Z}| - |\mathcal{X}|$ constraints take the form $P_Z^*(z) = 0$.

Figure 2 shows the Wyner-Ziv rate region (solid line) and our algorithm's $(1+\epsilon)$-approximation (circles) when $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ with joint distribution $\{0.06, 0.24, 0.42, 0.28\}$, $d(x, \hat{x})$ is the Hamming distortion measure, and $\epsilon = 0.1$. The example demonstrates that the approximation is often tighter than the $(1+\epsilon)$ worst-case guarantee. While the Wyner-Ziv region is not difficult to calculate for these simple binary sources, the difficulty of the calculation increases with the alphabet size.

IV. The Lossy Coded Side Information Region

In [9], Berger et al. derive an achievability result for the lossy coded side-information problem illustrated in Fig. 1(b). Let $\mathcal{X}_1$ and $\mathcal{X}_2$ denote the finite alphabets for sources X_1 and X_2, respectively. The region proposed by Berger et al. is the convex hull of the rates

$$R_1 \geq I(X_1; Z_1|Z_2), \qquad R_2 \geq I(X_2; Z_2),$$

for $(Z_1, Z_2) \in \Psi(X_1, X_2)$, where

$$\Psi(X_1, X_2) \quad := \quad \Big\{(Z_1, Z_2)\Big|\, Z_1 \to X_1 \to X_2 \to Z_2, \\ \exists\psi \text{ s.t. } Ed(X, \psi(Z_1, Z_2)) \leq D\Big\}.$$

We find the desired lower convex hull using Lagrangian

$$J(\lambda_1, \lambda_2, \lambda_3) \quad := \quad \lambda_1 I(X_1; Z_1|Z_2) + \lambda_2 I(X_2; Z_2) \\ +\lambda_3 \min_{\psi} Ed(X_1, \psi(Z_1, Z_2)).$$

Calculating the optimal rate region for a given pair of sources (X_1, X_2) requires joint optimization of conditionals

$$\{Q_{Z_i|X_i}(z_i|x_i)\}_{(x_i,z_i)\in\mathcal{X}_i\times\mathcal{Z}_i}, \quad i \in \{1, 2\}$$

where $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are the alphabets for auxiliary random variables Z_1 and Z_2. Since joint optimization of these conditional distributions is tricky, we define a sequence of conditional distributions $\{Q_{Z_2|X_2}(z_2|x_2)\}_{(x_2,z_2)\in\mathcal{X}_2\times\mathcal{Z}_2}$ from the quantized class defined in (1) and then optimize $\{Q_{Z_1|X_1}(z_1|x_1)\}_{(x_1,z_1)\in\mathcal{X}_1\times\mathcal{Z}_1}$ for each. Comparing these optimal solutions yields the best pair of conditionals among all possible solutions in the class considered.

The number of possible conditionals on Z_2 given X_2 in the quantized class is $N(\delta, \eta, |\mathcal{X}_2|)^{|\mathcal{Z}_1|}$. To make this value as small as possible, we begin by bounding the alphabet size $|\mathcal{Z}_2|$. For any fixed conditional distributions $\{Q_{Z_1|X_1}(z_1|x_1)\}_{(x_1,z_1)\in\mathcal{X}_1\times\mathcal{Z}_1}$ and $\{Q_{X_2|Z_2}(x_2|z_2)\}_{(x_2,z_2)\in\mathcal{X}_2\times\mathcal{Z}_2}$, both $J(\lambda_1, \lambda_2, \lambda_3)$ and the distribution constraints are linear in $\{P_{Z_2}(z_2)\}_{z_2\in\mathcal{Z}_2}$. An argument analogous to the one in Section III then demonstrates that there exists an optimal solution to this linear equation in which $P_{Z_2}(z_2) = 0$ for all but at most $|\mathcal{X}_2|$ values of z_2 – giving $|\mathcal{Z}_2| \leq |\mathcal{X}_2|$. [2]

Let $\mathcal{Z}_1 = \{1, \ldots, N(\delta, \eta, |\mathcal{X}_1|)\}$, and for each $z_1 \in \mathcal{Z}_1$ let $\{Q_{X_1|Z_1}(x_1|z_1)\}_{x_1\in\mathcal{X}_1}$ be a distinct distribution from our quantized collection (1). Let $\mathcal{Z}_2 = \{1, \ldots, |\mathcal{X}_2|\}$. For each of the $N(\delta, \eta, |\mathcal{X}_2|)^{|\mathcal{X}_2|}$ conditionals $\{Q_{X_2|Z_2}(x_2|z_2)\}_{(x_2,z_2)\in\mathcal{X}_2\times\mathcal{Z}_2}$ in the class defined by (1), we run a linear program to optimize $J(\lambda_1, \lambda_2, \lambda_3)$ subject to the distribution constraints. The algorithm outputs the best of these solutions. The proof of Theorem 2 appears in the Appendix.

Theorem 2: The proposed algorithm runs in time $O(\epsilon^{-4(|\mathcal{X}_1|+|\mathcal{X}_2|^2+1)})$ and guarantees a $(1+\epsilon)$-approximation.

The broken lines in Figure 2 show our algorithm's $(1+\epsilon)$-approximation for the achievable rate region from [9] for the example considered in the previous section. Each curve plots rate R_1 against distortion for a fixed value of R_2. In this case, the optimal region is not easily available, but the given solution is guaranteed to meet our $(1+\epsilon)$-approximation bound.

V. Conclusions

The proposed family of algorithms enables systematic calculation of the rate regions for a large class of source coding problems. The ability to calculate these regions is useful because it allows us to determine the limits of what is possible in a variety of applications – thereby enabling an objective assessment of the performance of source coding algorithms.

The given approach may also be useful for resolving theoretical questions. The coded side-information problem provides a potential example. In [11], Berger and Tung derive an inner bound (here called the Berger-Tung bound) for the lossy multiple access source coding problem. While the formulations are quite different, in [12] Jana and Blahut prove the equivalence of the inner bounds from [9] and [11] A long-standing open question is whether the bound is tight. One possible means of proving the looseness of the bound would be to calculate it for random variables (X, Y) and

[2] We can similarly show that $|\mathcal{Z}_1| \leq |\mathcal{X}_1|$, though that result is not applied in the argument that follows.

compare the resulting region to the normalized region for the random variables (X^n, Y^n) where (X_i, Y_i) are drawn i.i.d. according to the same distribution as (X, Y). If these values differ for any n, then the region is not tight. (The experiment would be inconclusive if the values are the same.) Since direct calculation of these values is difficult even for $n = 2$, the proposed algorithm may enable a solution to this problem.

APPENDIX

A. Proof of Theorem 1

Let $J^*(\lambda) = \min_{\{P_Z(z)\}_{z\in\mathcal{Z}}} J(\lambda)$ be the optimal value of $J(\lambda)$ for the Wyner-Ziv rate region, and let $\widehat{J}(\lambda)$ be the value computed by the algorithm proposed in Section III. Then $\widehat{J}(\lambda) \geq J^*(\lambda)$ since the algorithm finds an auxiliary random variable Z achieving the given Lagrangian. We next find (η, δ) to guarantee that $\widehat{J}(\lambda) \leq (1+\epsilon)J^*(\lambda)$.

Recall that $J(\lambda) := I(X;Z|Y) + \lambda \min_\psi Ed(X, \psi(Y, Z))$. Before bounding $\widehat{J}(\lambda) - J^*(\lambda)$, rewrite $J(\lambda)$ as

$$\begin{aligned} J(\lambda) &= H(X|Y) - H(Y|X) + H(Y|Z) - H(X|Z) \\ &\quad + \lambda \min_\psi Ed(X, \psi(Y,Z)) \\ &= H(X|Y) - H(Y|X) + \sum_{z\in\mathcal{Z}} P_Z(z) H(Y|Z=z) \\ &\quad - \sum_{z\in\mathcal{Z}} P_Z(z) \Bigg[H(X|Z=z) \\ &\quad + \sum_x p(y|x) Q_{X|Z}(x|z) d(x, \psi^*(y,z)) \Bigg], \end{aligned}$$

where $\psi^*(y, z)$ is the optimizing reproduction of X given (y, z) and conditional $\{Q_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$. Fix $\{Q_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ and let $\{\widehat{Q}_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ be the quantized conditional. Let

$$\begin{aligned} Q_{Y|Z}(y|z) &= \sum_x p(y|x) Q_{X|Z}(x|z) \\ \widehat{Q}_{Y|Z}(y|z) &= \sum_x p(y|x) \widehat{Q}_{X|Z}(x|z) \end{aligned}$$

be the corresponding conditionals on Y given Z. Then

$$\begin{aligned} (1-\eta)Q_{X|Z}(x|z) - \delta \leq \quad \widehat{Q}_{X|Z}(x|z) &\leq (1+\eta)Q_{X|Z}(x|z) \\ (1-\eta)Q_{Y|Z}(y|z) - \delta \leq \quad \widehat{Q}_{Y|Z}(y|z) &\leq (1+\eta)Q_{Y|Z}(y|z) \end{aligned}$$

for all x, y, z. Finally, let $\{P_Z^*(z)\}_{z\in\mathcal{Z}}$ be the marginal on auxiliary random variable Z that achieves $J^*(\lambda)$ and define $\tau := \eta \log \frac{e}{1-\eta}$. By Lemma 1, when $(\max\{|\mathcal{X}|, |\mathcal{Y}|\})\,\delta \log \frac{1}{\delta} < \tau$

$$\begin{aligned} |H(\widehat{Q}_{X|Z=z}) - H(Q_{X|Z=z})| &\leq \eta H(Q_{X|Z=z}) + 2\tau \\ |H(\widehat{Q}_{Y|Z=z}) - H(Q_{Y|Z=z})| &\leq \eta H(Q_{Y|Z=z}) \end{aligned}$$

for every $z \in \mathcal{Z}$.

Let $\mathcal{Z}' := \mathcal{Z} \cup \{z_x\}_{x\in\mathcal{X}}$ and set

$$\widehat{Q}_{X|Z}(t|z_x) = \begin{cases} 1, & \text{if } t = x \\ 0, & \text{otherwise} \end{cases}$$

$$\widehat{P}_Z(z) = \begin{cases} (1-\eta)P_Z^*(z), & \text{if } z \in \mathcal{Z} \\ P_X(x) - \sum_{z\in\mathcal{Z}} \widehat{Q}_{X|Z}(x|z)\widehat{P}_Z(z), & \text{if } z = z_x. \end{cases}$$

$$\widehat{\psi}(y,z) = \begin{cases} \psi^*(y,z) & \text{if } z \in \mathcal{Z} \\ x & \text{if } z = z_x \end{cases}$$

Since $\widehat{J}(\lambda)$ is optimized over all quantized distributions, $\widehat{J}(\lambda) \leq J(\lambda)|_{\widehat{Q}_{X|Z}, \widehat{P}_Z, \widehat{\psi}}$. Thus

$$\begin{aligned} &\widehat{J}(\lambda) - J^*(\lambda) \\ &\quad\leq [(2\eta - \eta^2)H(X|Z) - \eta^2 H(Y|Z)]_{P_Z^*, Q_{X|Z}} \\ &\qquad + (2\eta - \eta^2)H(Y|X) + 4(1-\eta)\tau + |\mathcal{Y}||\mathcal{Z}|\delta(1-\eta) \\ &\quad\leq \eta\left((2-\eta)(|\mathcal{X}| + |\mathcal{Y}|) + 8 + |\mathcal{Y}||\mathcal{Z}|\right) \\ &\quad\leq \eta(2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|). \end{aligned}$$

when $\delta < \eta < 1 - \frac{e}{4}$ and $(\max\{|\mathcal{X}|, |\mathcal{Y}|\})\,\delta \log \frac{1}{\delta} < \tau$. Define $L^*(\lambda) := \min_{0\leq D\leq D_{\max}} \left(R_{X|Y}(D) + \lambda D\right)$ where $R_{X|Y}(D)$ is the conditional rate-distortion function for X given Y. Then $J^*(\lambda) \geq L^*(\lambda)$ implies

$$\widehat{J}(\lambda) - J^*(\lambda) \leq \frac{\eta(2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|)}{L^*(\lambda)} J^*(\lambda).$$

We therefore wish to choose δ and η to satisfy

$$\delta < \eta = \frac{L^*(\lambda)}{2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|}\epsilon < 1 - \frac{e}{4}.$$

Define $f(x) := -x\log(x)$ for $x \in [0, 1/e)$. Function f is strictly increasing and therefore invertible. Setting

$$\eta = \min\left\{\frac{L^*(\lambda)}{2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|}\epsilon, 1 - \frac{e}{4}\right\} \quad \text{(A-1)}$$

$$\delta = \min\left\{\eta, f^{-1}\left(\min\left\{\frac{\eta}{|\mathcal{X}|}, \frac{\eta}{|\mathcal{Y}|}\right\}\right)\right\} \quad \text{(A-2)}$$

yields $(\max\{|\mathcal{X}|, |\mathcal{Y}|\})\,\delta \log \frac{1}{\delta} < \tau$ as desired and guarantees a $(1+\epsilon)$-approximation.

The interior-point solver for k variables runs in time $O(k^4)$ [13]. Since $k = |\mathcal{Z}|$ for our linear program, our algorithm runs in time $O(N(\delta, \eta, |\mathcal{X}|)^4)$. Applying the given choice of δ and η, our algorithm runs in time $O(\epsilon^{-4(|\mathcal{X}|+1)})$ as ϵ approaches 0.

B. Proof of Theorem 2

Let $J^*(\lambda_1, \lambda_2, \lambda_3) = \min J(\lambda_1, \lambda_2, \lambda_3)$ be the optimal value of $J(\lambda_1, \lambda_2, \lambda_3)$ for the lossy coded side information region, and let $\widehat{J}(\lambda_1, \lambda_2, \lambda_3)$ be the value computed by the algorithm proposed in Section IV. Then $\widehat{J}(\lambda_1, \lambda_2, \lambda_3) \geq J^*(\lambda_1, \lambda_2, \lambda_3)$. We next find (η, δ) such that $\widehat{J}(\lambda_1, \lambda_2, \lambda_3) \leq (1+\epsilon)J^*(\lambda_1, \lambda_2, \lambda_3)$.

Let $\mathcal{Z}_1' = \mathcal{Z}_1 \cup \{z_{x_1}\}_{x_1\in\mathcal{X}_1}$ and set

$$\widehat{Q}_{X_1|Z_1}(t|z_{x_1}) = \begin{cases} 1, & \text{if } t = x_1 \\ 0, & \text{otherwise} \end{cases}.$$

Let $P_{Z_1}^* Q_{Z_2|X_2}^*$ be a distribution on (Z_1, X_2) that achieves $J^*(\lambda_1, \lambda_2, \lambda_3)$. Define

$$\widehat{P}_{Z_1}(z) = \begin{cases} (1-\eta')P_{Z_1}^*(z) & \forall\, z \in \mathcal{Z}, \\ P_{X_1}(x_1) - \sum_z \widehat{Q}_{X_1|Z}(x_1|z)\widehat{P}_{Z_1}(z) & \forall\, x_1 \in \mathcal{X}_1. \end{cases}$$

Let $\tau' := \eta' \log \frac{e}{1-\eta'}$ and $\delta' := (|\mathcal{X}_1|+1)\delta\eta' = 3\eta$. Choose $\delta > 0$ such that $|\mathcal{X}_1||\mathcal{Z}_2|\delta' \log \frac{1}{\delta'} \le \tau'$. By Lemma 1, for all $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, and $z_1 \in \mathcal{Z}_1$

$$
\begin{aligned}
&|H(\widehat{Q}_{X_1|Z_1=z_1}) - H(Q_{X_1|Z_1=z_1})| \\
&\quad \le \eta' H(Q_{X_1|Z_1=z_1}) + 2\tau' \\
&|H(\widehat{Q}_{X_1,Z_2|Z_1=z_1}) - H(Q_{X_1,Z_2|Z_1=z_1})| \\
&\quad \le \eta' H(Q_{X_1,Z_2|Z_1=z_1}) + 2\tau' \\
&|H(\widehat{Q}_{Z_2|Z_1=z_1}) - H(Q_{Z_2|Z_1=z_1})| \\
&\quad \le \eta' H(Q_{Z_2|Z_1=z_1}) + 2\tau' \\
&|H(\widehat{Q}_{Z_2|X_2=x_2}) - H(Q_{Z_2|X_2=x_2})| \\
&\quad \le \eta' H(Q_{Z_2|X_2=x_2}) + 2\tau' \\
&|H(\widehat{Q}_{Z_2}) - H(Q_{Z_2})| \le \eta' H(Q_{Z_2}) + 2\tau'
\end{aligned}
$$

where $\widehat{Q}_{X_1,Z_2|Z_1}$, $\widehat{Q}_{Z_2|Z_1}$, and $\widehat{Q}_{Z_2}$ derive from $(\widehat{Q}_{X_1|Z_1}, \widehat{Q}_{Z_2|X_2})$. Let ψ^* be the optimal reproduction function for X_1 for $(P^*_{Z_1}, Q^*_{Z_2|X_2})$. Extend the function $\psi^*(z_1,z_2)$ to $\psi^*(z_{x_1},z_2) = x_1$ for all $x_1 \in \mathcal{X}_1$. Now

$$
\begin{aligned}
I(X_1;Z_1|Z_2) &\\
&= H(X_1|Z_1) - H(X_1|Z_1,Z_2) \\
&= H(X_1|Z_1) - H(X_1,Z_2|Z_1) - H(Z_2|Z_1) \\
I(X_2;Z_2) &= H(Z_2) - H(Z_2|X_2),
\end{aligned}
$$

and hence we have

$$
\begin{aligned}
\widehat{H}(X_1|Z_1) &\le (1-\eta'^2)H^*(X_1|Z_1) + 2(1-\eta')\tau' \\
\widehat{H}(X_1,Z_2|Z_1) &\ge (1-\eta')^2 H^*(X_1,Z_2|Z_1) - 2(1-\eta')\tau' \\
\widehat{H}(Z_2|Z_1) &\ge (1-\eta')^2 H^*(Z_2|Z_1) - 2(1-\eta')\tau' \\
\widehat{H}(Z_2|X_2) &\ge (1-\eta')^2 H^*(Z_2|X_2) - 2(1-\eta')\tau' \\
&Ed(X_1,\psi^*(Z_1,Z_2))_{\widehat{P}_{Z_1},\widehat{Q}^*_{Z_2|X_2}} \\
&\quad = (1-\eta')Ed(X_1,\psi^*(Z_1,Z_2))_{P^*_{Z_1},Q^*_{Z_2|X_2}}.
\end{aligned}
$$

Since $\widehat{J}(\lambda_1,\lambda_2,\lambda_3) \le \widehat{J}(\lambda_1,\lambda_2,\lambda_3)|_{\widehat{P}_{Z_1},\{\widehat{Q}_{X_1|Z_1}\},\widehat{Q}^*_{Z_2|X_2}}$, by taking $\eta' < 1 - \frac{e}{4}$, we have

$$
\begin{aligned}
&\widehat{J}(\lambda_1,\lambda_2,\lambda_3) - J^*(\lambda_1,\lambda_2,\lambda_3) \\
&\quad \le \lambda_1\left(2\eta'\left(|\mathcal{X}_1||\mathcal{Z}_1|\right) + |\mathcal{Z}_2|\right)) + \lambda_2(\eta'|\mathcal{Z}_2| + 2\eta'|\mathcal{Z}_2|) \\
&\qquad +(6\lambda_1 + 4\lambda_2)(1-\eta')\tau' \\
&\quad \le \eta'\left(2\lambda_1(|\mathcal{X}_1||\mathcal{Z}_1| + |\mathcal{Z}_1|) + 3\lambda_2|\mathcal{Z}_2| + (12\lambda_1 + 8\lambda_2)\right).
\end{aligned}
$$

Define

$$L(\lambda_1,\lambda_2,\lambda_3) := \min_D \left[\min\{\lambda_1,\lambda_2\}R_{X_1}(D) + \lambda_3 D\right].$$

Set

$$\eta = \min\left\{\frac{4-e}{12}, \frac{\epsilon L(\lambda_1,\lambda_2,\lambda_3)}{T(\lambda_1,\lambda_2,\lambda_3)}\right\} \quad \text{(A-3)}$$

$$\delta = \frac{1}{|\mathcal{X}_1|+1} f^{-1}\left(\frac{\eta}{3|\mathcal{X}_1||\mathcal{Z}_2|}\right), \quad \text{(A-4)}$$

where

$$
\begin{aligned}
&T(\lambda_1,\lambda_2,\lambda_3) \\
&\quad := 6\lambda_1(|\mathcal{X}_1||\mathcal{Z}_1| + |\mathcal{Z}_1|) + 9\lambda_2|\mathcal{Z}_2| + (36\lambda_1 + 24\lambda_2).
\end{aligned}
$$

Then the pair (η,δ) satisfies the following inequalities

$$
\begin{aligned}
\delta' &= (|\mathcal{X}_1|+1)\delta \\
\eta' &= 3\eta < 1 - \frac{e}{4} \\
&|\mathcal{X}_1||\mathcal{Z}_2|\delta' \log\frac{1}{\delta'} \le \tau' \\
\eta' &< \frac{\epsilon L(\lambda_1,\lambda_2,\lambda_3)}{2\lambda_1(|\mathcal{X}_1||\mathcal{Z}_1| + |\mathcal{Z}_1|) + 3\lambda_2|\mathcal{Z}_2| + (12\lambda_1 + 8\lambda_2)},
\end{aligned}
$$

which gives

$$J^*(\lambda_1,\lambda_2,\lambda_3) \le \widehat{J}(\lambda_1,\lambda_2,\lambda_3) \le (1+\epsilon)J^*(\lambda_1,\lambda_2,\lambda_3).$$

In this algorithm, there are $N(\delta,\eta,|\mathcal{X}_1|)$ variables in each of the linear programs in the inner loop, and there are $N(\delta,\eta,|\mathcal{X}_2|)^{|\mathcal{X}_2|}$ quantized conditional probabilities $\widehat{Q}_{Z_2|X_2}$ in the outer loop. Applying the given choice of δ and η, our algorithm runs in time $O(\epsilon^{-4(|\mathcal{X}_1|+|\mathcal{X}_2^2|+1)})$ as ϵ approaches 0.

References

[1] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, July 1973.

[2] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.

[3] S. Jana and R. Blahut. Achievable region for multiterminal source coding with lossless decoding in all sources except one. In *Proceedings of the IEEE Information Theory Workshop*, pages 460–465, Lake Tahoe, CA, September 2007.

[4] R.E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473, July 1972.

[5] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, IT-18:14–20, January 1972.

[6] I. Csiszár. On the computation of the rate-distortion function. *IEEE Transactions on Information Theory*, IT-20:122–124, January 1974.

[7] W. Gu and M. Effros. Approximation on computing rate regions for coded side information problem. In *Proceedings of the IEEE Information Theory Workshop*, Tahoe, CA, September 2007.

[8] I. Csiszár and J. Köner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.:Academic, 1981.

[9] T. Berger, K. B. Housewright, J. K. Omura, S. Tung, and J. Wolfowitz. An upper bound on the rate distortion function for source coding with partial side information at the decoder. *IEEE Transactions on Information Theory*, IT-25(6):664–666, November 1979.

[10] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-22(1):1–10, January 1976.

[11] T. Berger. Optimum quantizers and permutation codes. *IEEE Transactions on Information Theory*, 18(6):759–765, November 1972.

[12] S. Jana and R. Blahut. Partial side information problem: equivalence of two inner bounds. In *Proceedings of the Conference on Information Sciences and Systems*, Princeton, NJ, March 2008.

[13] S. Nash and A. Sofer. *Linear and Nonlinear Programming*. McGraw-Hill, 1996.