

On Approximating the Rate Region for Source Coding with Coded Side Information

WeiHsin Gu and Michelle Effros
 Department of Electrical Engineering
 California Institute of Technology
 Pasadena, California, CA 91125, USA
 Email: {wgu, effros}@caltech.edu

Abstract—The achievable rate region for the problem of lossless source coding with coded side information was derived by Ahlswede and Körner in 1975. While the Ahlswede-Körner bound completely characterizes the achievable rate region when the source and side information are memoryless, calculating this bound for a given memoryless joint probability mass function on the source and side information requires an optimization over all possible auxiliary random variables meeting a given Markov condition and alphabet size constraint. This optimization turns out to be surprisingly difficult even for very simple distributions on the source and side information. We here propose a $(1 + \epsilon)$ -approximation algorithm for the given rate region. The proposed technique involves quantization of a space of conditional distributions followed by linear programming. The resulting algorithm guarantees performance within a multiplicative factor $(1 + \epsilon)$ of the optimal performance – even when that optimal performance is unknown.

I. INTRODUCTION

Consider the coded side information problem shown in Fig. 1. Source X and side information Y are drawn i.i.d. according to joint probability mass function $p(x, y)$ on finite alphabets \mathcal{X} and \mathcal{Y} , respectively. One encoder observes only source X ; the other observes only source Y ; there is no communication between the two encoders. The two encoders describe their observed sources to the decoder at rates R_X and R_Y , respectively. The decoder uses those descriptions to build a lossless reconstruction of X . Thus source Y acts as coded side information – it is useful in that describing Y in part or in whole may decrease the rate required to describe X but it is not itself required by the decoder.

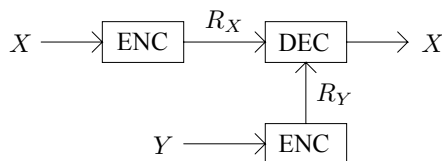


Fig. 1. The coded side information problem.

In [1], Ahlswede and Körner derive the achievable rate region for the coded side information problem. The Ahlswede-

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

Körner rate region is the set of points (R_X, R_Y) satisfying

$$\begin{aligned} R_X &\geq H(X|U) \\ R_Y &\geq I(Y;U) \end{aligned}$$

for some auxiliary random variable U from finite alphabet \mathcal{U} satisfying Markov condition

$$X \rightarrow Y \rightarrow U,$$

and cardinality bound

$$|\mathcal{U}| \leq |\mathcal{Y}| + 2.$$

For any rate pair (R_X, R_Y) on the lower convex hull of the achievable rate region, the cardinality bound on \mathcal{U} is improved to

$$|\mathcal{U}| \leq |\mathcal{Y}|$$

in [2].

Calculating the achievable rate region for any particular probability mass function $p(x, y)$ involves an optimization over the space of possible auxiliary random variables. Rate regions for a few special cases appear in [3] and [4].

We here consider the calculation of the given rate region for an arbitrary probability mass function $p(x, y)$. In particular, since our interest lies in the lower convex hull of the achievable rate region and the auxiliary random variable must satisfy the Markov condition $X \rightarrow Y \rightarrow U$, for each Lagrangian constant $\lambda > 0$ it suffices to derive the conditional probability mass function $p(u|y)$ on finite alphabet $\{1, \dots, |\mathcal{Y}|\}$ that minimizes the Lagrangian functional

$$H(X|U) + \lambda I(Y;U).$$

(Note that only the cardinality of \mathcal{U} affects the achievable rate pair.)

Unfortunately, the given optimization is non-trivial for a general probability mass function $p(x, y)$ on X and Y . As a result, rather than proposing a precise solution for a limited collection of probability mass functions, we propose a $(1 + \epsilon)$ -approximation algorithm applicable on an arbitrary probability mass function $p(x, y)$. For any $\lambda > 0$, let

$$J^*(\lambda) = \min_{\{p(u|y)\}} [H(X|U) + \lambda I(Y;U)],$$

where the given minimization is over all conditional probability mass functions $\{p(u|y)\}_{u \in \mathcal{U}, y \in \mathcal{Y}}$. Our algorithm finds an auxiliary random variable U_o on an alphabet \mathcal{U} of cardinality $|\mathcal{U}| \leq |\mathcal{Y}|$ such that $X \rightarrow Y \rightarrow U_o$ and

$$J^*(\lambda) \leq H(X|U_o) + \lambda I(Y; U_o) \leq (1 + \epsilon)J^*(\lambda).$$

The proposed technique involves first quantizing a space of possible conditional distributions and then running a linear program to optimize over the quantized space of conditional distributions. Following the description of our algorithm, we comment briefly on techniques by which the efficiency of our algorithm may be improved and generalizations of our approach to calculate rate regions for other source and channel coding problems.

II. $(1 + \epsilon)$ -APPROXIMATION ALGORITHM

In this section, we reformulate the optimization problem inherent in the solution of the Ahlswede-Körner region and propose an efficient method for approximating its solution. The reformulation involves translating the problem from an optimization of conditional probability mass function $\{p(u|y)\}$ to an optimization of conditional probability mass function $\{p(y|u)\}$ and marginal $\{p(u)\}$ subject to the constraint that $\sum_{u \in \mathcal{U}} p(y|u)p(u) = p(y)$ for all $y \in \mathcal{Y}$. We make this choice because given any finite alphabet \mathcal{U} and any collection of conditional probability mass functions $\{\{p(y|u)\}_{y \in \mathcal{Y}}\}_{u \in \mathcal{U}}$, the Lagrangian $H(X|U) + \lambda I(Y; U)$ is linear in $p(u)$. Thus, given any fixed, finite collection $\mathcal{P}_o = \{\{p(y|u)\}_{y \in \mathcal{Y}}\}_{u \in \mathcal{U}}$ of conditional distributions on \mathcal{Y} given u , the optimization

$$\begin{aligned} J(\mathcal{P}_o, \lambda) &= \min_{\{p(u)\}_{u \in \mathcal{U}}} [H(X|U) + \lambda I(Y; U)] \\ \text{subject to} & \sum_{u \in \mathcal{U}} p(u)p(y|u) = p(y) \text{ for each } y \in \mathcal{Y} \\ & \sum_{u \in \mathcal{U}} p(u) = 1 \\ & p(u) \geq 0 \text{ for each } u \in \mathcal{U} \end{aligned}$$

is a linear program [2].

Unfortunately, the optimal choice of \mathcal{P}_o is unknown a priori and the space

$$\mathcal{P}_Y = \left\{ \left\{ p(y|u) \right\}_{y \in \mathcal{Y}} : \begin{aligned} & p(y|u) \geq 0 \quad \forall y \in \mathcal{Y} \\ & \sum_{y \in \mathcal{Y}} p(y|u) = 1 \end{aligned} \right\},$$

of possibilities is uncountably infinite. As a result, the first step in our algorithm is to carefully quantize \mathcal{P}_Y . In particular, we choose an index set $\mathcal{I} = \{1, \dots, N\}$, and from the uncountably infinite space of conditional probability mass functions we choose a finite subset $\hat{\mathcal{P}}_Y = \{\{p(y|u_i)\}_{y \in \mathcal{Y}} :$

$i \in \mathcal{I}\} \subset \mathcal{P}_Y$ in a manner that guarantees that

$$\begin{aligned} J(\hat{\mathcal{P}}_Y, \lambda) &= \min_{\{p(u)\}_{u \in \mathcal{U}}} [H(X|U) + \lambda I(Y; U)] \\ \text{subject to} & \sum_{u \in \mathcal{U}} p(u)p(y|u) = p(y) \text{ for each } y \in \mathcal{Y} \\ & \sum_{u \in \mathcal{U}} p(u) = 1 \\ & p(u) \geq 0 \text{ for each } u \in \mathcal{U} \end{aligned}$$

satisfies the desired constraints

$$J^*(\lambda) \leq J(\hat{\mathcal{P}}_Y, \lambda) \leq (1 + \epsilon)J^*(\lambda).$$

While the number N of conditional distributions in our quantized collection $\hat{\mathcal{P}}_Y$ is larger than the desired cardinality bound $|\mathcal{U}| \leq |\mathcal{Y}|$, the linear program is guaranteed to place non-zero probability on at most $|\mathcal{Y}|$ of those conditional distributions by [2].

A. Quantizing the Space of Conditional Distributions

Our quantization of \mathcal{P}_Y to design $\hat{\mathcal{P}}_Y$ relies on the following lemma.

Lemma 1: Let $p(x)$ and $q(x)$ be two distributions on finite alphabet \mathcal{X} . Let $\eta \in [0, 1)$ be fixed. If $|p(x) - q(x)| \leq \eta p(x)$, for all $x \in \mathcal{X}$, then

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

Proof. Consider any fixed $x \in \mathcal{X}$. By the mean-value theorem, there is some $r_x \in [(1 - \eta)p(x), (1 + \eta)p(x)]$ such that

$$\begin{aligned} & \left| p(x) \ln \frac{1}{p(x)} - q(x) \ln \frac{1}{q(x)} \right| \\ &= |p(x) - q(x)| \left| \ln \frac{1}{r_x} - 1 \right| \\ &\leq \eta p(x) \max \left\{ \left| \ln \frac{1}{e(1 + \eta)p(x)} \right|, \left| \ln \frac{1}{e(1 - \eta)p(x)} \right| \right\} \\ &\leq \eta p(x) \ln \frac{1}{p(x)} + \eta p(x) + \eta p(x) \ln \frac{1}{1 - \eta} \\ &= \eta p(x) \ln \frac{1}{p(x)} + \eta p(x) \ln \frac{e}{1 - \eta}. \end{aligned}$$

So the result follows. \square

For any $0 < \eta, \delta < 1/e$, let

$$\mathcal{P}_Y(\eta, \delta) = \cup_{y_o \in \mathcal{Y}} \mathcal{P}_Y(y_o, \eta, \delta)$$

where for each $y_o \in \mathcal{Y}$

$$\begin{aligned} & \mathcal{P}_Y(y_o, \eta, \delta) \\ &= \left\{ \{p(y)\} \in \mathcal{P}_Y : p(y_o) = \max_{y \in \mathcal{Y}} p(y) \right. \\ & \quad \left. y \neq y_o \Rightarrow p(y) \in \{0\} \cup \left\{ \left(1 + \frac{\eta}{|\mathcal{Y}|}\right)^n : 0 \leq n \leq \frac{-\ln \delta}{\ln(1 + \eta/|\mathcal{Y}|)} \right\} \right\}. \end{aligned}$$

Then $\mathcal{P}_Y(\eta, \delta)$ is a finite subset of \mathcal{P}_Y . We choose an arbitrary order on this set, and index the elements by $\mathcal{I}(\eta, \delta) =$

$\{1, \dots, N(\eta, \delta)\}$, where $N(\eta, \delta) = |\mathcal{P}_Y(\eta, \delta)|$. Using these indices gives

$$\mathcal{P}_Y(\eta, \delta) = \{\{\hat{p}(y|u_i)\} : i \in \mathcal{I}(\eta, \delta)\}.$$

The given collection of distributions has the property that for any distribution $\{p(y|u)\} \in \mathcal{P}_Y$ on \mathcal{Y} there exists an index $i(u) \in \mathcal{I}(\eta, \delta)$ such that $|p(y|u) - \hat{p}(y|u_{i(u)})| < \eta p(y|u)$ for all $y \in \mathcal{Y}$ for which $p(y|u) \geq \delta$ while $\hat{p}(y|u_{i(u)}) = 0$ for all $y \in \mathcal{Y}$ for which $p(y|u) < \delta$.

Finally, let

$$\begin{aligned} \mathcal{P}_U(\eta, \delta) &= \{\{p(u_i)\}_{i \in \mathcal{I}(\eta, \delta)} : \\ &\sum_{i \in \mathcal{I}(\eta, \delta)} p(u_i) \hat{p}(y|u_i) = p(y) \quad \forall y \in \mathcal{Y} \\ &\sum_{i \in \mathcal{I}(\eta, \delta)} p(u_i) = 1 \\ &p(u_i) \geq 0 \quad \forall i \in \mathcal{I}(\eta, \delta)\} \end{aligned}$$

and

$$\hat{J}(\eta, \delta, \lambda) = \min_{\{p(u)\} \in \mathcal{P}_U} [H(X|\hat{U}) + \lambda I(Y; \hat{U})],$$

where $H(X|\hat{U})$ and $I(Y; \hat{U})$ are the conditional entropy and mutual information, respectively, using the conditional probability mass function $\mathcal{P}_Y(\eta, \delta)$.

We next show $\hat{J}(\eta, \delta, \lambda)$ is a good estimation of $J^*(\lambda)$ for an appropriate choice of (η, δ) . Let $\{p^*(u)\}$ be the distribution on \mathcal{P}_Y that achieves the optimal value $J^*(\lambda)$. Note that while \mathcal{P}_Y is uncountably infinite, $p^*(u) > 0$ on at most $|\mathcal{Y}|$ elements of \mathcal{P}_Y by [2]. Let

$$\mathcal{U}(\lambda) = \{u : p^*(u) > 0\}.$$

For each $u \in \mathcal{U}(\lambda)$,

$$(1 - \eta)p(y|u) - \delta \leq \hat{p}(y|u_{i(u)}) \leq (1 + \eta)p(y|u). \quad (1)$$

Based on Lemma 1,

$$\begin{aligned} &|H(\{p(y|u)\}_{y \in \mathcal{Y}}) - H(\{p(y|u_{i(u)})\}_{y \in \mathcal{Y}})| \\ &\leq \eta H(\{p(y|u)\}_{y \in \mathcal{Y}}) + \eta \log \frac{e}{1 - \eta} + |\mathcal{Y}| \delta \log \frac{1}{\delta}. \quad (2) \end{aligned}$$

In particular, by choosing δ such that $|\mathcal{Y}| \delta \log \frac{1}{\delta} < \eta \log \frac{1}{1 - \eta}$, (2) becomes

$$\begin{aligned} &|H(\{p(y|u)\}_{y \in \mathcal{Y}}) - H(\{p(y|u_{i(u)})\}_{y \in \mathcal{Y}})| \\ &\leq \eta H(\{p(y|u)\}_{y \in \mathcal{Y}}) + 2\eta \log \frac{e}{1 - \eta}. \quad (3) \end{aligned}$$

Let $\{\hat{p}(u_i)\}_{i \in \mathcal{I}(\eta, \delta)}$ be a distribution defined as follows. For each $u \in \mathcal{U}(\lambda)$ for which $\hat{p}(y|u_{i(u)}) < 1$ for all $y \in \mathcal{Y}$, let

$$\hat{p}(u_{i(u)}) = (1 - \eta)p^*(u),$$

and for each $y \in \mathcal{Y}$ if index $j(y)$ is chosen so that $\hat{p}(y|u_{j(y)}) = 1$, let

$$\hat{p}(u_{j(y)}) = p(y) - \sum_{i \in \mathcal{I}(\eta, \delta)} \hat{p}(y|u_{i(u)}) \hat{p}(u_i).$$

Finally, for each $i \notin \{i(u) : u \in \mathcal{U}(\lambda)\} \cup \{j(y) : y \in \mathcal{Y}\}$, let

$$\hat{p}(u_i) = 0.$$

Then $\hat{p} \in \mathcal{P}_U$ by construction. Moreover,

$$\begin{aligned} \hat{p}(u_{j(y)}) &\leq p(y) - (1 - \eta)^2 p(y) + (1 - \eta)\delta \\ &= (2\eta - \eta^2)p(y) + (1 - \eta)\delta, \quad (4) \end{aligned}$$

where the inequality follows from (1).

Let

$$H_\lambda^*(Y|U) = \sum_{u \in \mathcal{U}(\lambda)} p^*(u) H(\{p(y|u)\}_{y \in \mathcal{Y}})$$

and let

$$\hat{H}_{\lambda, \eta, \delta}(Y|U) = \sum_{u \in \mathcal{U}(\lambda) \cup_{y \in \mathcal{Y}} \{j(y)\}} \hat{p}(u_{i(u)}) H(\{\hat{p}(y|u_{i(u)})\}_{y \in \mathcal{Y}}).$$

Then we have

$$\begin{aligned} &H_\lambda^*(Y|U) - \hat{H}_{\lambda, \eta, \delta}(Y|U) \\ &= \sum_{u \in \mathcal{U}(\lambda)} p^*(u) H(\{p(y|u)\}_{y \in \mathcal{Y}}) \\ &\quad - \sum_{u \in \mathcal{U}(\lambda)} \hat{p}(u_{i(u)}) H(\{p(y|u_{i(u)})\}_{y \in \mathcal{Y}}) \\ &\geq \sum_{u \in \mathcal{U}(\lambda)} p^*(u) H(\{p(y|u)\}_{y \in \mathcal{Y}}) - (1 - \eta)p^*(u) \times \\ &\quad ((1 + \eta)H(\{p(y|u)\}_{y \in \mathcal{Y}}) + 2\eta \log \frac{e}{1 - \eta}) \\ &\geq \eta^2 H_\lambda^*(Y|U) - 2\eta(1 - \eta) \log \frac{e}{1 - \eta}. \end{aligned}$$

Similarly,

$$\begin{aligned} &H_\lambda^*(Y|U) - \hat{H}_{\lambda, \eta, \delta}(Y|U) \\ &\leq (2\eta - \eta^2) H_\lambda^*(Y|U) + 2\eta(1 - \eta) \log \frac{e}{1 - \eta}. \quad (5) \end{aligned}$$

Hence

$$\begin{aligned} &|H_\lambda^*(Y|U) - \hat{H}_{\lambda, \eta, \delta}(Y|U)| \\ &\leq (2\eta - \eta^2) H_\lambda^*(Y|U) + 2\eta(1 - \eta) \log \frac{e}{1 - \eta}. \end{aligned}$$

We next look at the term $H(X|U)$ in our estimation. Define for each $u \in \hat{U}(\lambda)$

$$\hat{p}(x|u_{i(u)}) := \sum_{y \in \mathcal{Y}} \hat{p}(y|u_{i(u)}) p(x|y)$$

and

$$p(x|u) := \sum_{y \in \mathcal{Y}} p(y|u) p(x|y).$$

Now for any $u \in \mathcal{U}(\lambda)$,

$$\begin{aligned} \hat{p}(x|u_{i(u)}) &= \sum_{y \in \mathcal{Y}} \hat{p}(y|u_{i(u)}) p(x|y) \\ &\leq \sum_{y \in \mathcal{Y}} (1 + \eta)p(y|u) p(x|y) \\ &= (1 + \eta)p(x|u). \end{aligned}$$

Therefore, for any $u \in \mathcal{U}(\lambda)$,

$$\begin{aligned}
H(\{\hat{p}(x|u_{i(u)})\}_{x \in \mathcal{X}}) &= \sum_{x \in \mathcal{X}} \hat{p}(x|u_{i(u)}) \log \frac{1}{\hat{p}(x|u_{i(u)})} \\
&\leq \sum_{x \in \mathcal{X}} \hat{p}(x|u_{i(u)}) \log \frac{1}{p(x|u)} \\
&\leq (1 + \eta) \sum_{x \in \mathcal{X}} p(x|u) \log \frac{1}{p(x|u)} \\
&\leq (1 + \eta) H(\{p(x|u)\}_{x \in \mathcal{X}}). \quad (6)
\end{aligned}$$

Define

$$H_{\lambda}^*(X|U) = \sum_{u \in \mathcal{U}(\lambda)} p^*(u) H(\{p(x|u)\}_{x \in \mathcal{X}})$$

and

$$\hat{H}_{\lambda, \eta, \delta}(X|U) = \sum_{u \in \mathcal{U}(\lambda)} \hat{p}(u_{i(u)}) H(\{\hat{p}(x|u_{i(u)})\}_{x \in \mathcal{X}}).$$

Then by (4) and (6)

$$\begin{aligned}
\hat{H}_{\lambda, \eta, \delta}(X|U) &= \sum_{u \in \mathcal{U}(\lambda)} \hat{p}(u_{i(u)}) H(\{p(x|u_{i(u)})\}_{x \in \mathcal{X}}) \\
&= \sum_{u \in \mathcal{U}(\lambda)} \hat{p}(u_{i(u)}) H(\{p(x|u_{i(u)})\}_{x \in \mathcal{X}}) \\
&\quad + \sum_{y \in \mathcal{Y}} \hat{p}(u_{j(y)}) H(X|Y) \\
&\leq (1 - \eta^2) H_{\lambda}^*(X|U) + (2\eta - \eta^2) \\
&\quad + |\mathcal{Y}|(1 - \eta)\delta H(X|Y).
\end{aligned}$$

Now since $\hat{H}_{\lambda, \eta, \delta}(X|U) \geq H(X|Y)$, there exists a constant $M > 0$ (independent of η and δ) such that

$$\hat{H}_{\lambda, \eta, \delta}(X|U) < (1 + M\eta) H_{\lambda}^*(X|U). \quad (7)$$

Now by definition, $J^*(\lambda) \leq \hat{J}(\eta, \delta, \lambda)$. By (5) and (7),

$$\begin{aligned}
\hat{J}(\eta, \delta, \lambda) - J^*(\lambda) &\leq \hat{H}_{\lambda, \eta, \delta}(X|U) - \lambda \hat{H}_{\lambda, \eta, \delta}(Y|U) \\
&\leq M\eta H_{\lambda}^*(X|U) + (2\eta - \eta^2) H_{\lambda}^*(Y|U) \\
&\quad + 2\eta(1 - \eta) \log \frac{e}{1 - \eta} \\
&= \eta(M H_{\lambda}^*(X|U) + (2 - \eta) H_{\lambda}^*(Y|U) \\
&\quad + 2(1 - \eta) \log \frac{e}{1 - \eta})
\end{aligned}$$

Since $J_{\lambda} > H(X|Y) > 0$, there exists $L > 0$ such that

$$\begin{aligned}
M H_{\lambda}^*(X|U) + (2 - \eta) H_{\lambda}^*(Y|U) + 2(1 - \eta) \log \frac{e}{1 - \eta} \\
\leq L J^*(\lambda)
\end{aligned}$$

Therefore,

$$J^*(\lambda) \leq \hat{J}(\eta, \delta, \lambda) \leq (1 + L\eta) J^*(\lambda).$$

By letting $\eta = \epsilon/L$, we get

$$J^*(\lambda) \leq \hat{J}(\eta, \delta, \lambda) \leq (1 + \epsilon) J^*(\lambda)$$

as desired.

B. Linear Program

The performance $\hat{J}(\eta, \delta, \lambda)$ is calculated using a distribution $\{\hat{p}(u_i)\}_{i \in \mathcal{I}(\eta, \delta)}$ that depends on the unknown optimal distribution $\{p^*(u)\}$. Notice, however, that this marginal on \mathcal{U} achieves performance at best as good as the solution to the following linear program

$$\sum_{i \in \mathcal{I}(\eta, \delta)} [H(X|U = u_i) p(u_i) + \lambda [H(Y) - H(Y|U = u_i) p(u_i)]]$$

$$\begin{aligned}
\text{subject to} \quad &\sum_{i \in \mathcal{I}(\eta, \delta)} p(u_i) = 1 \\
&\sum_{i \in \mathcal{I}(\eta, \delta)} p(u_i) \hat{p}(y|u_i) = p(y) \quad \forall y \in \mathcal{Y} \\
&p(u_i) \geq 0 \quad \forall i \in \mathcal{I}(\eta, \delta).
\end{aligned}$$

Thus

$$\begin{aligned}
J^*(\lambda) &\leq \hat{J}(\eta, \delta, \lambda) \\
&\leq [H_{\lambda}^*(X|U) + \lambda [H(Y) - H_{\lambda}^*(Y|U)]] \\
&\leq (1 + \epsilon) J^*(\lambda).
\end{aligned}$$

The result is a $(1 + \epsilon)$ -approximation algorithm for $J^*(\lambda)$.

III. CONCLUSIONS

The proposed approach yields a $(1 + \epsilon)$ -approximation algorithm for the Ahlswede-Körner rate region. The given solution may be sub-optimal, but it is provably good – meaning that it guarantees performance within a multiplicative factor $(1 + \epsilon)$ of the optimal. This guarantee is realized even in cases where the true rate region is unknown. The proposed technique combines quantization of the space of possible conditional distributions with linear programming. The given quantization scheme is entirely generic – meaning that it does not vary with the given distribution $p(x, y)$. Faster approximation algorithms might result from designing a quantization specific to that distribution. While this paper treats only the coded side information problem, we expect that the underlying strategy may generalize to other problems where calculation of known rate regions for source or channel coding requires a nontrivial optimization of the type seen in this example.

REFERENCES

- [1] 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.
- [2] W. Gu, R. Koetter, M. Effros, and T. Ho. On source coding with coded side information for a binary source with binary side information. In *Proceedings of the IEEE International Symposium on Information Theory*, Nice, France, June 2007. IEEE.
- [3] H. S. Witsenhausen and A. D. Wyner. A conditional entropy bound for a pair of discrete random variables. *IEEE Transactions on Information Theory*, 21:493–501, 1976.
- [4] D. Marco and M. Effros. A partial solution for lossless source coding with coded side information. In *Proceedings of the Information Theory Workshop*, Punta del Este, Uruguay, March 2006. IEEE.