The Computation of the Capacity Region of the Discrete Degraded BC is a Nonconvex DC Problem

Eduard Calvo[*,0], Daniel P. Palomar[†], Javier R. Fonollosa[*,0], and Josep Vidal[*,0]
e-mail: {eduard,fono,pepe}@gps.tsc.upc.edu, palomar@ust.hk
[*]SPCOM Group, Dpt. Signal Theory and Communications, Technical University of Catalonia (UPC), Spain.
[†]Dpt. of Electronic and Computer Engineering, Honk Kong University of Science and Technology (HKUST), Hong Kong.

***Abstract*— While the capacity region of the discrete memoryless broadcast channel is in general unknown, it admits a computable single-letter characterization when it is degraded. In this case, we pose its computation as an optimization problem and analyze its structure. We show that the computation of the capacity region of the two-user discrete memoryless degraded broadcast channel can be characterized as a difference of convex optimization problem, a non-convex problem in general. For this problem, which cannot be solved optimally in polynomial time, we obtain necessary conditions for optimality which substantially reduce the set of potential capacity-achieving candidate distributions. As an application of this result, the capacity region of the BEC-BSC degraded broadcast channel is derived by maximizing the achievable rates over this set of reduced dimensionality.**

I. Introduction

The two-user discrete memoryless broadcast channel (DMBC) [1] models the situation in which one sender wishes to send information to two receivers. Although a single-letter characterization of the capacity region of this channel is not known in general, some achievable rate regions and outer bounds have been derived by Cover and van der Meulen [2], [3], Marton [4], and Nair and El Gamal [5].

Achievable regions and outer bounds are usually expressed in terms of auxiliary random variables. Whenever the cardinalities of the support set of all the random variables involved in the region are bounded, we say that region is computable. In this respect, Nair's outer bound and Cover-van der Meulen's achievable region in case of binary inputs [6] are computable. However, apart from showing the feasibility of the evaluation of the regions, little work has been done towards finding efficient methods for their computation. We shall try to bridge this gap in the specific setup where the DMBC is degraded and there is no common information to be transmitted simultaneously to both receivers. Assuming these constraints, the capacity region of the channel is known [7]–[9] and computable.

The contributions of this paper are the following. By posing the computation of the capacity region of the two-user degraded DMBC (dDMBC) as a constrained optimization problem we show that it can be characterized as a difference of convex (DC) optimization problem [10]. Since this kind of problems are not convex in general, there is no efficient algorithm available to compute the capacity region in practice. Additionally, the lack of convexity of this problem makes the Karush-Kuhn-Tucker (KKT) conditions [11] not sufficient, not even necessary (some regularity conditions are required), for claiming optimality of an input distribution. Regarding this, we are able to show that in this problem they are necessary optimality conditions and, therefore, by choosing the best among those input distributions satisfying the KKT conditions the capacity region can be optimally computed. Moreover, since the dimensionality of the set of distributions satisfying the KKT conditions is in general substantially lower than the original feasible set, the complexity involved in computing the capacity region is greatly reduced. The contributions of this paper are finally applied to compute the capacity region of the BEC-BSC dDMBC, reducing the search dimension from 4 to 2.

This paper is organized as follows. Section II introduces the problem of the computation of the capacity region of the DMBC and reformulates it as a DC optimization problem. Section III describes the necessary optimality conditions and how they can be applied to obtain capacity-achieving distributions. Then, Section IV provides the capacity region of the BEC-BSC dDMBC by using the results of Section III. Finally, Section V concludes the paper.

[0]This work has been partially funded by the European Commission, the Spanish Ministry of Education and Science, the Catalan Government and FEDER funds under contracts TEC2006-06481, TEC2004-04526, 27187 SURFACE, and 2005SGR-00639, and grant FPU-AP-2004-3549.

II. The capacity region as a DC optimization problem

A. The problem of the capacity region

The capacity region $\mathcal{C}$ of the two-user dDMBC $X \rightarrow Y_1 \rightarrow Y_2$ is the convex hull of the set of rate pairs (R_1, R_2) satisfying

$$0 \leq R_1 \leq I(X; Y_1|U) \quad (1)$$

$$0 \leq R_2 \leq I(U; Y_2) \quad (2)$$

for some choice of the distribution $P_{UXY_1Y_2} = P_{UX}P_{Y_1|X}P_{Y_2|X}$ on $\mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$, where P_{UX} is the joint probability distribution of the auxiliary random variable U and the transmitted codeword X, and $P_{Y_1|X}, P_{Y_2|X}$ are the given conditional distributions that depend on the nature of the channel. The region $\mathcal{C}$ is computable since it suffices to consider $|\mathcal{U}| = \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$ in the evaluation of (1)-(2). Applying the property that $\mathcal{C}$ is convex by time-sharing arguments, the supporting hyperplane theorem

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

[11, Sec. 2.5.2] allows us to parameterize its computation for some $\theta \in [0,1]$ as

$$\begin{aligned} \underset{\{R_1,R_2,P_{UX}\}}{\text{maximize}} \quad & \theta R_1 + (1-\theta)R_2 & (3)\\ \text{subject to} \quad & 0 \le R_1 \le I(X;Y_1|U) & (4)\\ & 0 \le R_2 \le I(U;Y_2) & (5)\\ & P_{UX}(u,x) \ge 0 \quad \forall (u,x) \in \mathcal{U}\times\mathcal{X} & (6)\\ & \sum_{u,x} P_{UX}(u,x) = 1, & (7) \end{aligned}$$

where the right hand side of (4)-(5) amount to (8)-(9) at the bottom of this page. Note that the solutions to (3)-(7), denoted by $(R_1^\star(\theta), R_2^\star(\theta), P_{UX}^\star(\theta))$, generally depend on θ.

B. A DC optimization problem

For each θ, the optimal rates $(R_1^\star(\theta), R_2^\star(\theta))$ (which belong to the boundary of $\mathcal{C}$) satisfy the right hand side inequalities of (4)-(5) with equality. This allows us to rephrase (3)-(7) as

$$\begin{aligned} \underset{P_{UX}}{\text{maximize}} \quad & \theta I(X;Y_1|U) + (1-\theta) I(U;Y_2) & (10)\\ \text{subject to} \quad & P_{UX}(u,x) \ge 0 \quad \forall (u,x) \in \mathcal{U}\times\mathcal{X} & (11)\\ & \sum_{u,x} P_{UX}(u,x) = 1, & (12) \end{aligned}$$

where R_1, R_2 have been removed from the optimization since they can be computed evaluating (8)-(9) using $P_{UX}^\star(\theta)$, the solution to (10)-(12).

Lemma 1: $I(X;Y_1|U)$ is concave in P_{UX}, whereas $I(U;Y_2)$ is a difference of concave functions of P_{UX}.

Proof: See Appendix A. ■

The computation of $\mathcal{C}$ (10)-(12) amounts hence to the maximization of the difference of two concave functions of P_{UX} (10) over the probability simplex. This falls within the class of DC problems[1], a wide class of non-convex optimization problems which can only be solved in cases where an underlying structure can be exploited. In general, their non-convexity makes them intractable and only brute-force or random search methods seem to be available. However, they provide no quantification on their incurred suboptimality.

III. Optimality Conditions

Pursuing the solution of the computation of $\mathcal{C}$, we explore both local and global optimality conditions in order to either characterize completely the solutions to (10)-(12) or determine their structure in order to reduce the dimensionality of the search space. To that end, consider the following necessary optimality condition.

Lemma 2: A necessary condition for global optimality (and sufficient condition for local optimality) of the joint probability distribution P_{UX} for any fixed $\theta \in [0,1]$ is

$$\theta D(P_{Y_1|X=x}||P_{Y_1|U=u}) + (1-\theta)\mathbb{E}_{Y_2|X=x}\left\{\log\frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\} \begin{cases} = R(\theta) & \text{if } P_{UX}(u,x) > 0\\ \le R(\theta) & \text{if } P_{UX}(u,x) = 0 \end{cases} \quad (13)$$

for all $(u,x) \in \mathcal{U}\times\mathcal{X}$ and some non-negative real constant $R(\theta)$. Any distribution P_{UX} satisfying (13) yields an objective value in (10) of $\theta I(X;Y_1|U) + (1-\theta)I(U;Y_2) = R(\theta)$.

Proof: See Appendix B. ■

Corollary 1: A condition for global optimality of the joint probability distribution P_{UX} for any fixed $\theta \in [0,1]$ is to satisfy Lemma 2 with

$$R(\theta) = C^\star(\theta) \triangleq \max_{(R_1,R_2)\in\mathcal{C}} \theta R_1 + (1-\theta)R_2, \quad (14)$$

where $C^\star(\theta)$ denotes the optimal value of (10)-(12) for the given θ.

Proof: The best among all the distributions satisfying a necessary condition for optimality must be optimal. ■

IV. The BEC-BSC Degraded Broadcast Channel

We shall consider here the application of the necessary conditions of Lemma 2 to the BEC-BSC dDMBC, whose channel transition probabilities are shown in a diagram in Figure 1. This channel is such that receivers one and two see a Binary Erasure Channel (BEC) and a Binary Symmetric Channel (BSC) respectively. It models the situation in which the sender intends to convey independent information to one receiver equipped with erasure correction capabilities and another not equipped so. The second receiver copies the output of the first receiver except when there is an erasure (denoted by e), in which case the output is chosen uniformly. Hence, if ϵ denotes the erasure probability of the BEC seen by the first receiver, the error probability of the equivalent BSC seen by the second receiver is $\epsilon/2$.

Proposition 1: The capacity region of the BEC-BSC degraded broadcast channel, $\mathcal{C}^{\mathrm{BEC-BSC}}$, is given by the convex hull of the points $\big((1-\epsilon)\log 2, 0\big)$, $(0, (1-h(0.5\epsilon))\log 2)$, and the set of rate pairs $(R_1(\theta), R_2(\theta))$, $0 < \theta < \frac{1}{2}\frac{1-\epsilon}{(1-0.5\epsilon)^2}$,

[1]It can equivalently be mapped to the minimization of the difference of two convex functions by minimizing the opposite of the objective in (10).

$$I(X;Y_1|U) = \sum_{u,x,y_1} P_{UX}(u,x)P_{Y_1|X}(y_1|x)\log\frac{\left(\sum_{x'}P_{UX}(u,x')\right)P_{Y_1|X}(y_1|x)}{\sum_{x'}P_{UX}(u,x')P_{Y_1|X}(y_1|x')} \quad (8)$$

$$I(U;Y_2) = \sum_{u,y_2}\Big(\sum_x P_{UX}(u,x)P_{Y_2|X}(y_2|x)\Big)\log\frac{\sum_{x'}P_{UX}(u,x')P_{Y_2|X}(y_2|x')}{\left(\sum_{x'}P_{UX}(u,x')\right)\left(\sum_{u',x'}P_{UX}(u',x')P_{Y_2|X}(y_2|x')\right)} \quad (9)$$

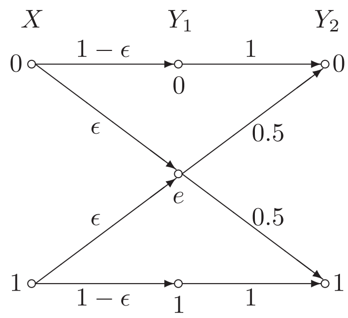

Fig. 1. The BEC-BSC degraded broadcast channel

achieved by the distribution

$$P_{UX}(0,0;\theta) = \frac{\gamma(\theta)-\beta(\theta)}{(1-\alpha(\theta)\beta(\theta))(1+\gamma(\theta))} \quad (15)$$

$$P_{UX}(1,1;\theta) = \frac{1-\alpha(\theta)\gamma(\theta)}{(1-\alpha(\theta)\beta(\theta))(1+\gamma(\theta))} \quad (16)$$

$$P_{UX}(0,1) = \alpha(\theta)P_{UX}(0,0;\theta) \quad (17)$$

$$P_{UX}(1,0) = \beta(\theta)P_{UX}(1,1;\theta), \quad (18)$$

where $h(x) \triangleq -x\log x - (1-x)\log(1-x)$ is the binary entropy function.

The parameters $\alpha(\theta), \beta(\theta)$, and $\gamma(\theta)$ are obtained as described next. Given $\alpha(\theta) > 0$,

$$\gamma(\theta) = \frac{0.5\epsilon\exp\big(g(\alpha(\theta);\theta,\epsilon)/(1-\theta)\big) - (1-0.5\epsilon)}{0.5\epsilon - (1-0.5\epsilon)\exp\big(g(\alpha(\theta);\theta,\epsilon)/(1-\theta)\big)}, \quad (19)$$

where

$$g(\alpha;\theta,\epsilon) = (1-\theta)\log\frac{(1-0.5\epsilon)\alpha+0.5\epsilon}{1-0.5\epsilon+0.5\epsilon\alpha} - \theta\log\alpha, \quad (20)$$

and $\beta(\theta) > 0$ is such that $g(\beta(\theta);\theta,\epsilon) = -g(\alpha(\theta);\theta,\epsilon)$ (there are at most three such possible values of $\beta(\theta)$ for each given $\alpha(\theta)$). Finally, $\alpha(\theta)$ is determined from the following unidimensional maximization

$$\alpha(\theta) = \arg \max_{\substack{P_{UX}:\alpha>0,\\ |g(\alpha;\theta,\epsilon)|<(1-\theta)\log(2/\epsilon-1)}} \theta I(X;Y_1|U) + (1-\theta)I(U;Y_2), \quad (21)$$

where it has been explicitly denoted that $P_{UX}(\theta)$ in (15)-(18) and hence $I(X;Y_1|U), I(U;Y_2)$ exclusively depend on α.

Proof: See Appendix C. ∎

The capacity region of the BEC-BSC degraded broadcast is shown in Figure 2. Note that the application of Lemma 2 and Corollary 1 allows us to compute $\mathcal{C}^{\text{BEC}-\text{BSC}}$ performing the maximization of a one-dimensional function instead of maximizing the achievable rates over the probability simplex containing the joint distributions on $\mathcal{U}\times\mathcal{X}$, for which three degrees of freedom would need to be explored for each θ.

V. Conclusions

We have characterized the computation of the capacity region of the two-user dDMBC as a DC problem. Since this class of problems cannot be solved optimally in general, we focused on obtaining local and global optimality conditions. In this respect, we found that satisfaction of the KKT conditions

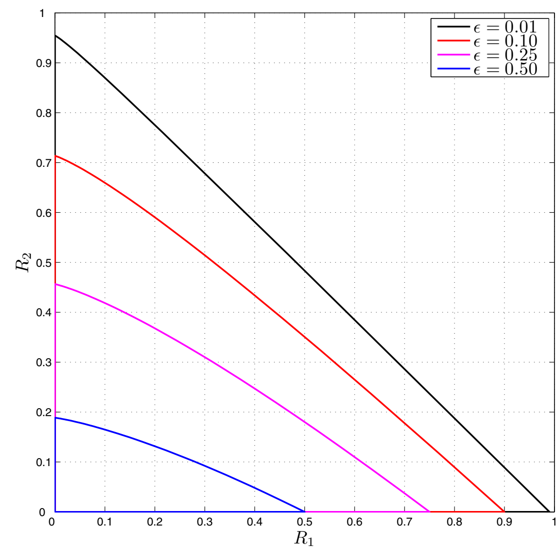

Fig. 2. The capacity region $\mathcal{C}^{\text{BEC}-\text{BSC}}$ of the BEC-BSC degraded broadcast channel, in [bit/ch. use], for different values of ϵ.

of this problem is necessary and sufficient for an input distribution to achieve a local maximum. Moreover, among all the distributions satisfying the KKT, those achieving the largest objective value are capacity-achieving (i.e., globally optimal). One immediate application of these results is to find capacity-achieving distributions from the maximization of the achievable rates over a candidate set of potentially much lower dimensionality than the original feasible set. In particular, we have applied this method to obtain the capacity region of the BEC-BSC dDMBC.

Appendix A: Proof of Lemma 1

The mutual information ruling the rate of the first link can be decomposed as[2].

$$I(X;Y_1|U) = \sum_{u,x,y_1} P_{UX}(u,x)P_{Y_1|X}(y_1|x)\Big(\log P_{Y_1|X}(y_1|x) + \log\frac{P_U(u)}{\sum_{x'}P_{UX}(u,x')P_{Y_1|X}(y_1|x')}\Big)$$

$$= \sum_x\Big(\sum_u P_{UX}(u,x)\Big)H(Y_1|X=x) - D(P_{UY_1}||Q_{UY_1}), \quad (22)$$

where

$$Q_{UY_1}(u,y_1) = P_U(u) \quad \forall(u,y_1)\in\mathcal{U}\times\mathcal{Y}_1 \quad (23)$$

is a dummy function which is not a probability distribution. While the first term is linear in P_{UX} and hence concave, the second is concave in (P_{UY_1}, Q_{UY_1}) thanks to the convexity of the divergence, which is based on the log-sum inequality, regardless of the fact that its two arguments may or may not be probability distributions[3]. Since (P_{UY_1}, Q_{UY_1}) are linear in

[2]When required, we denote the marginals of P_{UX} by P_U and P_X to shorten the notation.

[3]Actually, in this case $D(\cdot||\cdot)$ is not a proper divergence since its arguments are not probability distributions, but this is totally irrelevant for the convexity characterization of the problem.

P_{UX}, it follows that $I(X;Y_1|U)$ is concave in P_{UX}.

Regarding the other mutual information term,

$$\begin{aligned} I(U;Y_2) &= \sum_u \Big(\sum_{x,y_2} P_{UX}(u,x)P_{Y_2|X}(y_2|x)\Big)\log\frac{1}{P_U(u)} \\ &+ \sum_{u,y_2}\Big(\sum_x P_{UX}(u,x)P_{Y_2|X}(y_2|x)\Big)\times \\ &\times \log\frac{\sum_{x'} P_{UX}(u,x')P_{Y_2|X}(y_2|x')}{\sum_{x'} P_X(x')P_{Y_2|X}(y_2|x')} \\ &= H(U) - (-D(P_{UY_2}||Q_{UY_2})), \end{aligned} \quad (24)$$

where

$$Q_{UY_2}(u,y_2) = \sum_x P_X(x)P_{Y_2|X}(y_2|x) \ \forall (u,y_2) \in \mathcal{U}\times\mathcal{Y}_2 \quad (25)$$

is another dummy function not satisfying the properties of a probability distribution. By the concavity of the entropy function, the convexity of the divergence, and the fact that $(P_U, P_{UY_2}, Q_{UY_2})$ are linear in P_{UX} it follows that $I(U;Y_2)$ is a difference of concave functions of P_{UX}.

Appendix B: Proof of Lemma 2

We will show that satisfaction of the KKT conditions, rephrased as in (13) is a necessary and sufficient condition for ensuring that P_{UX} achieves a local maximum of (10)-(12) and therefore (13) is a necessary optimality condition. From [12, Sec. 3.4], we know that the KKT conditions of (10)-(12), are satisfied by any P_{UX} achieving a local maximum (hence proving the necessity part). The corresponding Lagrangian is

$$\begin{aligned} \mathcal{L}(P_{UX};\mathbf{\Phi}_{UX},\eta) &= \theta I(X;Y_1|U) + (1-\theta)I(U;Y_2) \\ &+ \sum_{u,x}\Phi_{UX}(u,x)P_{UX}(u,x) + \eta\Big(\sum_{u,x}P_{UX}(u,x)-1\Big), \end{aligned} \quad (26)$$

while the derivatives of mutual information are

$$\frac{\partial I(X;Y_1|U)}{\partial P_{UX}(u,x)} = D(P_{Y_1|X=x}||P_{Y_1|U=u}) \quad (27)$$

$$\frac{\partial I(U;Y_2)}{\partial P_{UX}(u,x)} = \mathbb{E}_{Y_2|X=x}\left\{\log\frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\} - \log(e). \quad (28)$$

Setting the derivative of the Lagrangian with respect to $P_{UX}(u,x)$ equal to zero it follows that at any local maximum the following holds

$$\begin{aligned} \theta D(P_{Y_1|X=x}||P_{Y_1|U=u}) + (1-\theta)\mathbb{E}_{Y_2|X=x}\left\{\log\frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\} \\ = (1-\theta)\log(e) - \eta - \Phi_{UX}(u,x). \end{aligned} \quad (29)$$

Expression (29) can be rephrased as (13) using complementary slackness ($\Phi_{UX}(u,x)P_{UX}(u,x) = 0$), dual feasibility ($\Phi_{UX}(u,x) \geq 0$), and noticing that an alternative formulation of the mutual informations involved is

$$I(X;Y_1|U) = \sum_{u,x} P_{UX}(u,x)D(P_{Y_1|X=x}||P_{Y_1|U=u}) \quad (30)$$

$$I(U;Y_2) = \sum_{u,x} P_{UX}(u,x)\mathbb{E}_{Y_2|X=x}\left\{\log\frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\}. \quad (31)$$

To show that any P_{UX} satisfying the KKT conditions is indeed a local maximum of (10)-(12), let $\partial R_{P_{UX}}(u,x) \triangleq \theta\frac{\partial I(X;Y_1|U)}{\partial P_{UX}(u,x)} + (1-\theta)\frac{\partial I(U;Y_2)}{\partial P_{UX}(u,x)}$ denote a linear combination of (27)-(28) evaluated using P_{UX}. If

$$\sum_{u,x}\partial R_{P_{UX}}(u,x)(Q_{UX}(u,x) - P_{UX}(u,x)) \leq 0 \quad (32)$$

holds for any arbitrary distribution $Q_{UX}(u,x)$ it follows that P_{UX} is a local maximum of (10)-(12) [12, Sec. 2.1]. Since any P_{UX} satisfying the KKT conditions has an associated $\partial R_{P_{UX}}(u,x) = -\eta - \Phi_{UX}(u,x)$ and satisfies complementary slackness ($\Phi_{UX}(u,x)P_{UX}(u,x) = 0$), (32) becomes

$$-\sum_{u,x}\Phi_{UX}(u,x)Q_{UX}(u,x) \leq 0, \quad (33)$$

which indeed holds for any Q_{UX} thanks to dual feasibility.

Appendix C: Proof of Proposition 1

Let us simplify the notation by using the equivalence $P_{UX}(u,x) = P_{ux}$ and imposing $\mathcal{U} = \{0,1\}$ ($|\mathcal{U}| = 2$ suffices). In this case the distributions involved amount to

$$P_{Y_1|X} = \begin{bmatrix} 1-\epsilon & 0 \\ \epsilon & \epsilon \\ 0 & 1-\epsilon \end{bmatrix}, \ P_{Y_2|X} = \begin{bmatrix} 1-0.5\epsilon & 0.5\epsilon \\ 0.5\epsilon & 1-0.5\epsilon \end{bmatrix} \quad (34)$$

for the channel transition matrices,

$$P_{Y_1|U} = \begin{bmatrix} \frac{(1-\epsilon)P_{00}}{P_{00}+P_{01}} & \frac{(1-\epsilon)P_{10}}{P_{10}+P_{11}} \\ \epsilon & \epsilon \\ \frac{(1-\epsilon)P_{01}}{P_{00}+P_{01}} & \frac{(1-\epsilon)P_{11}}{P_{10}+P_{11}} \end{bmatrix} \quad (35)$$

$$P_{Y_2|U} = \begin{bmatrix} \frac{(1-0.5\epsilon)P_{00}+0.5\epsilon P_{01}}{P_{00}+P_{01}} & \frac{(1-0.5\epsilon)P_{10}+0.5\epsilon P_{11}}{P_{10}+P_{11}} \\ \frac{0.5\epsilon P_{00}+(1-0.5\epsilon)P_{01}}{P_{00}+P_{01}} & \frac{0.5\epsilon P_{10}+(1-0.5\epsilon)P_{11}}{P_{10}+P_{11}} \end{bmatrix} \quad (36)$$

for the output distributions conditioned on U, and

$$P_{Y_2} = \begin{bmatrix} (1-0.5\epsilon)(P_{00}+P_{10}) + 0.5\epsilon(P_{01}+P_{11}) \\ 0.5\epsilon(P_{00}+P_{10}) + (1-0.5\epsilon)(P_{01}+P_{11}) \end{bmatrix} \quad (37)$$

for the output distribution of the second receiver. In (34)-(37) we have used the convention that the columns of the matrices represent the elements of the output alphabets ($\mathcal{Y}_1$ or $\mathcal{Y}_2$) and the rows correspond to the natural ordering of the inputs (U or X). Expressions (34)-(37) can be used to formulate the conditions of Lemma 2 for the BEC-BSC dDMBC, which are shown in (38)-(41) at the bottom of the next page, where $\ell_{ij} \geq 0$ and $\ell_{ij} = 0$ if $P_{ij} > 0$. In order to find an optimal distribution satisfying (38)-(41) we need to hypothesize on the number of entries of P_{ij} that are equal to zero. To that end, consider the following situations:

- *Hypothesis 1 - P_{UX} has three zero entries*. This implies $H(U) = H(X) = 0$ and, consequently, $I(X;Y_1|U) = I(U;Y_2) = 0$, which is clearly suboptimal.
- *Hypothesis 2 - P_{UX} has two zero entries*. This class of distributions comprises the cases: i) $X = U$ and $X = \overline{U}$, where all the capacity-achieving distributions achieve $(0, (1-\theta)(1-h(0.5\epsilon))$, ii) $U = 0$ and $U = 1$, where all the capacity achieving distributions achieve $(\theta(1-\epsilon)\log 2, 0)$, and iii) $X = 0$ and $X = 1$, which imply $I(X;Y_1|U) = I(U;Y_2) = 0$.

• *Hypothesis 3 - P_{UX} has one zero entry*. Consider w.l.o.g $P_{00} = 0$ and $P_{ij} > 0 \ \forall (i,j) \neq (0,0)$. This implies $\ell_{ij} = 0 \ \forall (i,j) \neq (0,0)$ and $\ell_{00} \geq 0$. The conditions (38)-(41) cannot be satisfied simultaneously because the left hand side of (38) equals $+\infty$, while $R(\theta) \leq C^\star(\theta)$ is bounded and $\ell_{00} \geq 0$. Therefore, distributions with one zero entry are never optimal.

We subsequently focus on distributions P_{UX} with strictly positive entries, which imply $\ell_{ij} = 0 \ \forall i,j \in \{0,1\}$. Let us describe such distributions by

$$P_{UX} = \begin{bmatrix} p_\alpha & \alpha p_\alpha \\ \beta p_\beta & p_\beta \end{bmatrix}, \tag{42}$$

where $p_\alpha, p_\beta, \alpha, \beta > 0$ and $(1+\alpha)p_\alpha + (1+\beta)p_\beta = 1$. Setting the left hand sides of (38) and (40) equal to each other and using (42) it follows

$$(1-\theta)\log \frac{(1-0.5\epsilon)+0.5\epsilon\alpha}{0.5\epsilon+(1-0.5\epsilon)\alpha}\frac{0.5\epsilon(p_\alpha+\beta p_\beta)+(1-0.5\epsilon)(\alpha p_\alpha+p_\beta)}{(1-0.5\epsilon)(p_\alpha+\beta p_\beta)+0.5\epsilon(\alpha p_\alpha+p_\beta)} + \theta\log\alpha = 0. \tag{43}$$

Proceeding similarly with the left hand sides of (39) and (41) we arrive at

$$(1-\theta)\log \frac{(1-0.5\epsilon)\beta+0.5\epsilon}{0.5\epsilon\beta+(1-0.5\epsilon)}\frac{0.5\epsilon(p_\alpha+\beta p_\beta)+(1-0.5\epsilon)(\alpha p_\alpha+p_\beta)}{(1-0.5\epsilon)(p_\alpha+\beta p_\beta)+0.5\epsilon(\alpha p_\alpha+p_\beta)} - \theta\log\beta = 0. \tag{44}$$

Since both (43) and (44) must hold, we can equal their left hand sides, which imposes $g(\alpha;\theta,\epsilon) = -g(\beta;\theta,\epsilon)$, where g is defined in (20). On the other hand, considering that $g(1/\alpha;\theta,\epsilon) = -g(\alpha;\theta,\epsilon)$ it follows that $\beta = 1/\alpha'$ with α' such that $g(\alpha';\theta,\epsilon) = g(\alpha;\theta,\epsilon)$. Rewriting (43) as

$$(1-\theta)\log \frac{0.5\epsilon(p_\alpha+\beta p_\beta) + (1-0.5\epsilon)(\alpha p_\alpha+p_\beta)}{(1-0.5\epsilon)(p_\alpha+\beta p_\beta)+0.5\epsilon(\alpha p_\alpha+p_\beta)} = g(\alpha;\theta,\epsilon), \tag{45}$$

it follows that

$$\frac{p_\alpha+\beta p_\beta}{\alpha p_\alpha+\beta} = \frac{0.5\epsilon\exp\big(g(\alpha;\theta,\epsilon)/(1-\theta)\big) - (1-0.5\epsilon)}{0.5\epsilon - (1-0.5\epsilon)\exp\big(g(\alpha;\theta,\epsilon)/(1-\theta)\big)} \triangleq \gamma. \tag{46}$$

Since γ is the ratio of two strictly positive probabilities, we impose $\gamma > 0$ in (46) to obtain that

$$\gamma > 0 \Leftrightarrow |g(\alpha;\theta,\epsilon)| < (1-\theta)\log(2/\epsilon - 1). \tag{47}$$

An equivalent rephrasing of (46) is $p_\alpha/p_\beta = (\gamma-\beta)/(1-\alpha\gamma)$, which induces a distribution of the form (15)-(18). Analyzing the monotony of $g(\alpha;\theta,\epsilon)$ over the interval of interest $\alpha \in (0,+\infty)$ we obtain that

$$\text{sign}\left\{\frac{\partial g(\alpha;\theta,\epsilon)}{\partial\alpha}\right\} = \text{sign}\left\{-\alpha^2 + \frac{2}{\epsilon}\left[\frac{1-\epsilon}{(1-0.5\epsilon)\theta} - 2(1-0.5\epsilon)\right]\alpha - 1\right\}, \tag{48}$$

which shows that for $\theta \geq \frac{1-\epsilon}{2(1-0.5\epsilon)^2}$ the function $g(\alpha;\theta,\epsilon)$ is strictly decreasing and hence $\alpha' = \alpha$, $\beta = 1/\alpha$. It can be checked that (44)-(45) imply $\alpha = \beta = 1$ which causes $R_2 = 0$ and the optimal rate pair to be $((1-\epsilon)\log 2, 0)$.

When $0 < \theta < \frac{1-\epsilon}{2(1-0.5\epsilon)^2}$, the function $g(\alpha;\theta,\epsilon)$ has one local minimum and one local maximum, which bounds the number of different values of α' such that $g(\alpha';\theta,\epsilon) = g(\alpha;\theta,\epsilon)$ to a maximum of three. The maximization of the objective value $\theta I(X;Y_1|U) + (1-\theta)I(U;Y_2)$ over the distributions of the class (15)-(18) satisfying (47) yields the best distribution satisfying Lemma 2 with strictly positive probabilities, and its associated rate pair $(R_1(\theta), R_2(\theta))$. The convex hull of these rate pairs together with the extreme points $((1-\epsilon)\log 2, 0)$, $(0, (1-h(0.5\epsilon))\log 2)$ is hence $\mathcal{C}^{\text{BEC}-\text{BSC}}$ since there is no other rate pair achieved by a distribution satisfying Lemma 2.

REFERENCES

[1] T. M. Cover, "Broadcast channels", *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2-14, Jan. 1972.

[2] E. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel", *IEEE. Trans. Inform. Theory*, vol. IT-21, pp. 180-190, March 1975.

[3] T. M. Cover, "An achievable rate region for the broadcast channel", *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 399-404, July 1975.

[4] K. Marton, "A coding theorem for the discrete memoryless broadcast channel", *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 306-311, May 1979.

[5] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel", *IEEE Trans. Inform. Theory*, vol. IT-53, pp. 350-355, Jan. 2007.

[6] B. E. Hajek and M. B. Pursley, "Evaluation of an achievable rate region for the broadcast channel", *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 36-46, Jan. 1979.

[7] P. Bergmans, "Random coding theorem for broadcast channels with degraded components", *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197-207, Mar. 1973.

[8] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II", *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 772-777, Nov. 1973.

[9] R. G. Gallager, "Capacity and coding for degraded broadcast channels", in *Problemy Peredaci Informaccii*, vol. 10, no. 3, pp.3-14, Oct. 1974.

[10] R. Horst and N. V. Thoai, "DC programming: an overview", in *Jrnl. of Optim. Theory and Applications*, vol. 103, no. 1, pp. 1-43, Oct. 1999.

[11] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[12] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, 1995.

$$\theta(1-\epsilon)\log\frac{P_{00}+P_{01}}{P_{00}} + (1-\theta)\Big[(1-0.5\epsilon)\log\frac{(1-0.5\epsilon)P_{00}+0.5\epsilon P_{01}}{(P_{00}+P_{01})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} + 0.5\epsilon\log\frac{0.5\epsilon P_{00}+(1-0.5\epsilon)P_{01}}{(P_{00}+P_{01})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta)-\ell_{00} \tag{38}$$

$$\theta(1-\epsilon)\log\frac{P_{10}+P_{11}}{P_{10}} + (1-\theta)\Big[(1-0.5\epsilon)\log\frac{(1-0.5\epsilon)P_{10}+0.5\epsilon P_{11}}{(P_{10}+P_{11})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} + 0.5\epsilon\log\frac{0.5\epsilon P_{10}+(1-0.5\epsilon)P_{11}}{(P_{10}+P_{11})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta)-\ell_{10} \tag{39}$$

$$\theta(1-\epsilon)\log\frac{P_{00}+P_{01}}{P_{01}} + (1-\theta)\Big[0.5\epsilon\log\frac{(1-0.5\epsilon)P_{00}+0.5\epsilon P_{01}}{(P_{00}+P_{01})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} + (1-0.5\epsilon)\log\frac{0.5\epsilon P_{00}+(1-0.5\epsilon)P_{01}}{(P_{00}+P_{01})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta)-\ell_{01} \tag{40}$$

$$\theta(1-\epsilon)\log\frac{P_{10}+P_{11}}{P_{11}} + (1-\theta)\Big[0.5\epsilon\log\frac{(1-0.5\epsilon)P_{10}+0.5\epsilon P_{11}}{(P_{10}+P_{11})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} + (1-0.5\epsilon)\log\frac{0.5\epsilon P_{10}+(1-0.5\epsilon)P_{11}}{(P_{10}+P_{11})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta)-\ell_{11} \tag{41}$$