Blahut–Arimoto Algorithms for Inner and Outer Bounds on Capacity Regions of Broadcast Channels [†]

Yanan Dou [1], Yanqing Liu [2], Xueyan Niu [3], Bo Bai [3], Wei Han [3] and Yanlin Geng [1,*]

1   State Key Laboratory of ISN, Xidian University, Xi'an 710071, China; yanandou@stu.xidian.edu.cn
2   IT Operation Center, Bank of China, Beijing 100094, China; lyq2017616@126.com
3   Theory Lab, Central Research Institute, 2012 Labs, Huawei Tech. Co., Ltd., Shatin, N.T., Hong Kong SAR, China; niuxueyan3@huawei.com (X.N.); baibo8@huawei.com (B.B.); harvey.hanwei@huawei.com (W.H.)
*   Correspondence: ylgeng@xidian.edu.cn
†   This paper is an extended version of our paper published in IEEE International Symposium on Information Theory (ISIT), 2022 (Espoo, Finland, 26 June–1 July) and 2023 (Taipei, Taiwan, 25–30 June).

Abstract: The celebrated Blahut–Arimoto algorithm computes the capacity of a discrete memoryless point-to-point channel by alternately maximizing the objective function of a maximization problem. This algorithm has been applied to degraded broadcast channels, in which the supporting hyperplanes of the capacity region are again cast as maximization problems. In this work, we consider general broadcast channels and extend this algorithm to compute inner and outer bounds on the capacity regions. Our main contributions are as follows: first, we show that the optimization problems are max–min problems and that the exchange of minimum and maximum holds; second, we design Blahut–Arimoto algorithms for the maximization part and gradient descent algorithms for the minimization part; third, we provide convergence analysis for both parts. Numerical experiments validate the effectiveness of our algorithms.

Keywords: Blahut–Arimoto algorithm; broadcast channel; capacity region; superposition coding inner bound; Marton's inner bound; UV outer bound

1. Introduction

In 1972, Cover [1] introduced the two-receiver discrete memoryless broadcast channel $p(y, z|x)$ to model a system of downlink communication in which X is the sender and (Y, Z) are the receivers. In the same paper, he proposed a coding scheme which resulted in the superposition coding inner bound (SCIB). It turns out that the SCIB is indeed the capacity region for two-receiver broadcast channels in which the receivers are comparable in the following partial orders: degraded [2], less noisy [3], and more capable [3]. However, for a general broadcast channel the single-letter capacity region remains open.

To characterize the capacity region of a broadcast channel, a standard approach is to show that one inner bound matches another outer bound. Currently, the best inner bound for general broadcast channels is Marton's inner bound (MIB) [4], while the UV outer bound (UVOB) [5] was the best outer bound until recently, when a better one called J version outer bound was proposed in [6].

The evaluation of inner and outer bounds is critical in the following aspects: (1) the evaluation of an inner bound usually results in an optimal input distribution which can help in the design of practical coding schemes; (2) the identification of the capacity region of a particular broadcast channel through the comparison of one inner bound and one outer bound relies on the evaluation of these two bounds; and (3) when claiming to establish a new bound, it is necessary to show that the new bound strictly improves on old ones through the evaluation of bounds on a particular channel.

Remark 1. *This is the full version of conference papers accepted by ISIT 2022 and 2023 [7,8].*

However, this evaluation is usually difficult due to its non-convexity [9]. To alleviate this issue, there exist a number of generic optimization algorithms, such as interior point [10], active set [10], and sequential quadratic programming [11]. However, efficient algorithms should use the domain knowledge of information theory as well; from this viewpoint, we consider the Blahut–Arimoto (BA) algorithm, which is specially customized for information theory.

The original BA algorithm was independently developed by Blahut [12] and Arimoto [13] to calculate the channel capacity $C = \max_{q(x)} I(X; Y)$ for a general point-to-point channel $p(y|x)$. The algorithm transforms the original maximization problem into an alternating maximization problem:

$$\max_{q(x)} \sum_{x,y} q(x)p(y|x) \ln \frac{q(x|y)}{q(x)}, \quad \rightarrow \quad \max_{q(x), Q(x|y)} \sum_{x,y} q(x)p(y|x) \ln \frac{Q(x|y)}{q(x)}$$

where the updating formulae are explicit within each iteration.

There have been numerous extensions of the BA algorithm to various scenarios in information theory. For example, [14] applied the BA algorithm to compute the sum rate of the multiple-access channel. Later, using the idea from the BA algorithm, the whole capacity region of the multiple-access channel was formulated in [15] as a rank-one constrained problem and solved by relaxation methods. It is beyond the scope of this paper to list all of these references. Instead, we discuss those papers closely related to computing the bounds on capacity regions of broadcast channels.

In [16], the authors considered the capacity region of a degraded broadcast channel $p(y, z|x)$, where receiver Z is a degraded version of Y.

In this scenario, the capacity region of the rate pairs (R_Y, R_Z) is known, and can be achieved by the simplified version of superposition coding. The supporting hyperplanes can be characterized as

$$\theta R_Y + (1 - \theta) R_Z = \max_{q(u,x)} \theta I(X; Y|U) + (1 - \theta) I(U; Z).$$

Using a similar idea to that of the BA algorithm, the authors designed an algorithm to alternatively maximize the objective function.

The method in [16] is directly applicable to less noisy broadcast channels, as the characterization of the capacity region is the same as that of the degraded case. However, this equivalence no longer holds for the more capable case, as this time the value of the supporting hyperplane $\theta R_Y + (1 - \theta) R_Z$ is characterized as a max–min optimization problem (e.g., see Equation (14)). As a mater of fact, the supporting hyperplanes of the above-mentioned bounds, that is, SCIB, MIB, and UVOB, are all of the max–min form. The main issue is that the minimization part is inside the maximization part, which prevents the application of the BA algorithm to the whole problem.

The algorithms for calculating inner bounds and outer bounds for general broadcast channels are very limited. The authors of [17] considered MIB (see Section 3.2) and designed a BA algorithm to compute the sum rate $R_Y + R_Z$ of the simplified version, where the auxiliary random variable $W = \varnothing$. The objective function,

$$\sum_{u,v,x,y,z} q(u,v)q(x|u,v)p(y,z|x) \ln \frac{Q(u|y)Q(v|z)}{q(u,v)}$$

is convex in $q(x|u, v)$, which means that the maximum input X is a function of (U, V). Noticing this, the authors performed optimization over all fixed mappings $q(x|u, v)$. However, discarding W can result in a strictly smaller sum rate [18], making it is necessary to consider the complete version of MIB.

In this paper, we seek to design BA algorithms for general broadcast channels in order to compute the following inner and outer bounds: SCIB, MIB, and UVOB. The key difference here is that the optimization problems are max–min problems, rather than only

containing a maximization part. In Table 1, we provide an intuitive comparison of related references.

Table 1. Comparison of typical scenarios related to the BA algorithm.

Channel	Reference	Objective	Form	Algorithm
point-to-point	[12,13]	capacity	max	BA
multiple access	[14]	sum-rate	max	BA
multiple access	[15]	inner/outer bounds	max	relaxation
degraded broadcast	[16]	capacity region	max	BA
general broadcast	this paper	inner/outer bounds	max–min	BA + gradient

The notation we use is as follows. p denotes a fixed (conditional) probability distribution such as $p(y, z|x)$, while q and Q are used for (conditional) probabilities that are changeable. Calligraphic letters such as $\mathcal{S}$ are used to denote sets. The use of square brackets in the function $f[g]$ means that f is specified by the variable g; $\bar{\theta}$ denotes $\bar{\theta} := 1 - \theta$; and unless otherwise specified, we use the natural logarithm. To make the mathematical expressions more concise, we use the following abbreviations of the Kullback–Leibler divergences:

$$D_Y(p||q) := \sum_y p(y) \ln \frac{p(y)}{q(y)},$$

$$D_{Y|x}(p||q) := \sum_y p(y|x) \ln \frac{p(y|x)}{q(y|x)},$$

$$D_{Y|X}(p||q) := \sum_x p(x) \cdot D_{Y|x}(p||q).$$

The organization is as follows. First, in Section 2 we introduce the necessary background on the BA algorithm and its extension in [16]. Then, in Section 3 we extend the BA algorithm to the evaluation of SCIB, MIB, and UVOB. Convergence analyses of these algorithms are presented in Section 4. Finally, in Section 5 we perform numerical experiments to validate the effectiveness and efficiency of our algorithms.

2. Mathematical Background of Blahut–Arimoto Algorithms

We first introduce the standard BA algorithm in Section 2.1, as we will rely on several of its properties later. Then, in Section 2.2 we discuss why the method in [16] cannot be applied to general broadcast channels.

2.1. Blahut–Arimoto Algorithm for Point-to-Point Channel

For a point-to-point channel $p(y|x)$, the capacity C is the maximum of the mutual information $C = \max_{q(x)} I(X;Y) = \max_{q(x)} C(q)$, where

$$C(q) = H(X) - H(X|Y) = \sum_{x,y} q(x)p(y|x) \ln \frac{q(x|y)}{q(x)}. \tag{1}$$

By replacing $q(x|y)$ with a free variable $Q(x|y)$, the BA algorithm performs alternating maximization $\max_{q,Q} C(q, Q)$, where

$$C(q, Q) = \sum_{x,y} q(x)p(y|x) \ln \frac{Q(x|y)}{q(x)}. \tag{2}$$

Notice here that we abuse the notation of $C(\cdot)$, which should not cause confusion in general.

The above objective function can be reformulated as follows:

$$C(q, Q) = \sum_x q(x)(d[Q](x) - \ln q(x)), \tag{3}$$

$$\text{where} \quad d[Q](x) = \sum_y p(y|x) \ln Q(x|y). \tag{4}$$

We call this the basic form. For different scenarios, BA algorithms mainly differ in the distribution $q(\cdot)$ and function $d[Q](\cdot)$.

The following theorem (see proof in Appendix A) provides the explicit formulae for the maximum Q given q (denoted as $Q[q]$) and maximum q given Q (denoted as $q[Q]$).

Theorem 1. *The following properties hold for the problem* $\max_{q,Q} C(q, Q)$.

1. *Given a fixed q, $C(q, Q)$ is concave in Q, and the maximum point $Q[q]$ is induced by the input and the channel,*

$$Q[q](x|y) = q(x|y) = \frac{q(x)p(y|x)}{\sum_{x'} q(x')p(y|x')}. \tag{5}$$

Further, the function values satisfy

$$C(q) = C(q, Q[q]) \geq C(q, Q). \tag{6}$$

2. *Given a fixed Q, $C(q, Q)$ is concave in q, and the maximum point $q[Q]$ is obtained by the Lagrangian,*

$$q[Q] = \frac{\exp\{d[Q](x)\}}{\sum_{x'} \exp\{d[Q](x')\}}. \tag{7}$$

Further, evaluation of the function value results in

$$C(q[Q], Q) = \ln \sum_{x'} \exp\{d[Q](x')\} = d[Q](x) - \ln q[Q](x), \; \forall x. \tag{8}$$

Starting from an initial $q_0(x) > 0$, the BA algorithm performs alternating maximization and produces a sequence of points

$$q_0 \rightarrow Q_1 \rightarrow q_1 \rightarrow Q_2 \rightarrow \ldots \rightarrow q_n \rightarrow Q_{n+1} \rightarrow \ldots$$

where

$$Q_n = Q[q_{n-1}],$$
$$q_n = q[Q_n]$$

according to Equations (5) and (7), respectively.

The criterion for stopping the iterations is based on the following result [12,13].

Proposition 1 (Proposition 1 in [13], Theorem 2 in [12]). *It is the case that q_* maximizes $C(q)$ if and only if the following holds for $Q[q_*]$ and some scalar D:*

$$d[Q[q_*]](x) - \ln q_*(x) \begin{cases} = D, & q_*(x) > 0, \\ \leq D, & q_*(x) = 0. \end{cases}$$

Remark 2. *It should be mentioned that in order to avoid infinity minus infinity for those $q_*(x) = 0$ in the above proposition, the following equivalent formulae can be used:*

$$d[Q[q]](x) - \ln q(x) = \sum_y p(y|x) \ln \frac{q(x|y)}{q(x)} = \sum_y p(y|x) \ln \frac{p(y|x)}{q(y)}. \tag{9}$$

This is not be an issue in the BA algorithm, as $q_n(x) > 0$ according to Equation (7).

Thus, at the end of the n-th step, if the difference

$$\max_x\{d[Q[q_{n-1}]](x) - \ln q_{n-1}(x)\} - C(q_n, Q_n)$$
$$= \max_x\{d[Q_n](x) - \ln q_{n-1}(x)\} - \ln \sum_{x'} \exp\{d[Q_n](x')\}$$

is small enough then the iteration is stopped.

Summarizing the above details, we arrive at the BA algorithm depicted in Algorithm 1. The convergence of the resulted sequence of $C(q_n, Q_n)$ is characterized in the following theorem (see proof in Appendix A).

Theorem 2 (Theorem 1 in [13], Theorem 3 in [12]). *If $p_0 > 0$, then the value $C(q_n, Q_n)$ converges monotonically from below to the capacity C.*

Algorithm 1: Computing channel capacity

Input: $p(y|x)$, maximum iterations N, threshold $\epsilon > 0$;
Initialization: $q_0(x) > 0$, $\epsilon_0 > \epsilon$, $n = 0$;
while *$n < N$ and $\epsilon_n > \epsilon$* **do**
  $n \leftarrow n + 1$;
  $Q_n = Q[q_{n-1}]$ using Equation (5);
  $q_n = q[Q_n]$ using Equation (7);
  $C(q_n, Q_n) = \ln \sum_x \exp\{d[Q_n](x)\}$ using Equations (8) and (4);
  $\epsilon_n = \max_x\{d[Q_n](x) - \ln q_{n-1}(x)\} - C(q_n, Q_n)$ using Equation (4);
end
Output: $q_n(x)$, $Q_n(x|y)$, $C(q_n, Q_n)$

2.2. Blahut–Arimoto Algorithm for Degraded Broadcast Channel

In [16], the authors considered the capacity region of the degraded broadcast channel. The original objective function for the value of the supporting hyperplane $\theta R_Y + \bar{\theta} R_Z$ (where $\theta < \frac{1}{2}$) is:

$$
\begin{aligned}
&F(q(u, x)) \\
&= \theta I(X; Y|U) + \bar{\theta} I(U; Z) \\
&= \bar{\theta}(I(X; Y|U) + I(U; Z)) - (\bar{\theta} - \theta)I(X; Y|U) \\
&= \bar{\theta}(H(U, X) - H(X|Y, U) - H(U|Z)) - (\bar{\theta} - \theta)(H(Y|U) - H(Y|X)).
\end{aligned}
$$

(10)

(11)

Similar to the BA algorithm, the new objective function is

$$F(q, Q) = \bar{\theta} \cdot \sum_{u,x} q(u, x)(d[Q](u, x) - \ln q(u, x)), \tag{12}$$

where

$$d[Q](u, x) = \sum_{y,z} p(y, z|x)\left(\ln Q(x|y, u)Q(u|z) + \frac{\bar{\theta} - \theta}{\bar{\theta}} \ln \frac{Q(y|u)}{p(y|x)}\right). \tag{13}$$

Then, the authors designed an extended BA algorithm that alternately maximizes $F(q, Q)$ and analysed the convergence.

However, the above method does not generalize to allow for evaluating the capacity bounds of general broadcast channels. The main reason can be summarized in just one sentence: the minimum of the expectation is, in general, greater than the expectation of the

minimum. Taking SCIB as an example, using the representation $\mathcal{A}(U, X)$ in Lemma 1, the supporting hyperplane is

$$
\begin{aligned}
\theta R_Y + \bar{\theta} R_Z &= \max_{q(u,x)} \theta I(X; Y|U) + \bar{\theta} \cdot \min\{I(U; Z), I(U; Y)\} \\
&= \max_{q(u,x)} \Big(\bar{\theta}(H(U, X) - H(X|Y, U) + \min\{-H(U|Z), -H(U|Y)\}) \\
&\quad - (\bar{\theta} - \theta)(H(Y|U) - H(Y|X)) \Big).
\end{aligned}
$$

A direct extension might try to reformulate the last expression in the form of Equation (12) by letting

$$
d[Q](u, x) = \sum_{y,z} p(y, z|x) \Big(\ln Q(x|y, u) + \min\{\ln Q(u|z), \ln Q(u|y)\} + \frac{\bar{\theta} - \theta}{\bar{\theta}} \ln \frac{Q(y|u)}{p(y|x)} \Big),
$$

however, this is not an equivalent reformulation, as

$$
\min\{-H(U|Z), -H(U|Y)\} \geq \sum_{u,x} q(u, x) \cdot \sum_{y,z} p(y, z|x) \cdot \min\{\ln Q(u|z), \ln Q(u|y)\}.
$$

In the following section, we first use the fact that $\min\{f, g\} = \min_{\alpha \in [0,1]}\{\alpha f + \bar{\alpha} g\}$ to parameterize the minimum, then show the max–min exchanges so that we can apply the BA algorithm in the maximization part.

3. Blahut–Arimoto Algorithms for Capacity Bounds of Broadcast Channels

In this section, we first introduce two inner bounds and one outer bound on the capacity region of the broadcast channel. We characterize their supporting hyperplanes as max–min problems and show that the maximum and minimum can be exchanged. Then, we design BA algorithms for the maximization parts and gradient descent algorithms for the minimization parts.

3.1. Superposition Coding Inner Bound

The superposition coding inner bound was proposed by Cover in [1], which corresponds to region $\mathcal{B}$ in the following lemma (see proof in Appendix B). This region actually has three equivalent characterizations.

Lemma 1 (folklore). *The following regions $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are equivalent characterizations of the superposition coding inner bound:*

$$
\mathcal{A} := \bigcup_{q(u,x)} \mathcal{A}(U, X) :=
\begin{cases}
R_Y \leq I(X; Y|U), \\
R_Z \leq \min\{I(U; Z), I(U; Y)\},
\end{cases}
$$

$$
\mathcal{B} := \bigcup_{q(u,x)} \mathcal{B}(U, X) :=
\begin{cases}
R_Y \leq I(X; Y|U), \\
R_Z \leq I(U; Z), \\
R_Y + R_Z \leq I(X; Y),
\end{cases}
$$

$$
\mathcal{C} := \bigcup_{q(u,x)} \mathcal{C}(U, X) :=
\begin{cases}
R_Z \leq I(U; Z), \\
R_Y + R_Z \leq \min\{I(X; Y|U) + I(U; Z), I(X; Y)\}.
\end{cases}
$$

To characterize the supporting hyperplane $\theta R_Y + \bar{\theta} R_Z = F$, we choose to use the representation $\mathcal{A}(U, X)$. It is clear that

$$
\begin{aligned}
F &= \max_{q(u,x)} \theta I(X;Y|U) + \bar{\theta} \cdot \min\{I(U;Z), I(U;Y)\} \\
&= \max_{q(u,x)} \theta I(X;Y|U) + \bar{\theta} \cdot \min_{\alpha \in [0,1]} \left(\alpha I(U;Z) + \bar{\alpha} I(U;Y) \right) \\
&= \max_{q(u,x)} \min_{\alpha \in [0,1]} \theta I(X;Y|U) + \bar{\theta}\alpha I(U;Z) + \bar{\theta}\bar{\alpha} I(U;Y).
\end{aligned}
\tag{14}
$$

Notice here that we cannot use the BA algorithm directly, as there is a minimum inside the maximum. If we are able to swap the orders of the maximum and minimum, then we can adopt the BA algorithm in the maximization part.

To show this kind of exchange, we first introduce a Terkelsen-type min–max result in the following lemma.

Lemma 2 (Corollary 2 in Appendix A of [19]). *Let Λ_d be the d-dimensional simplex, i.e., $\lambda_i \geq 0$ and $\sum_{i=1}^{d} \lambda_i = 1$, let $\mathcal{P}$ be a set of probability distributions $p(u)$, and let $T_i(p(u))$, $i = 1, .., d$ be a set of functions such that the set $\mathcal{T}$ defined by*

$$
\mathcal{T} = \{(a_1, ..., a_d) \in \mathbb{R}^d : a_i \leq T_i(p(u)) \text{ for some } p(u) \in \mathcal{P}\}
$$

is a convex set; then,

$$
\sup_{p(u) \in \mathcal{P}} \min_{\lambda \in \Lambda_d} \sum_i \lambda_i T_i(p(u)) = \min_{\lambda \in \Lambda_d} \sup_{p(u) \in \mathcal{P}} \sum_i \lambda_i T_i(p(u)).
$$

With this lemma, we are to establish the following theorem (proof in Appendix B).

Theorem 3. *The supporting hyperplane $\theta R_Y + \bar{\theta} R_Z = F$ of the superposition coding inner bound is as follows: if $\theta \in [\frac{1}{2}, 1]$, then $F = \max_{q(x)} \theta I(X;Y)$; otherwise, if $\theta \in [0, \frac{1}{2})$ then*

$$
F = \min \Big\{ \min_{\alpha \leq \frac{1-2\theta}{1-\theta}} \max_{q(x)} \bar{\theta}\bar{\alpha} I(X;Y) + \bar{\theta}\alpha I(X;Z),
$$

$$
\min_{\alpha > \frac{1-2\theta}{1-\theta}} \max_{q(u,x)} \theta I(X;Y|U) + \bar{\theta}\bar{\alpha} I(U;Y) + \bar{\theta}\alpha I(U;Z) \Big\}.
$$

Further, it suffices to consider the cardinality size: $|\mathcal{U}| \leq |\mathcal{X}|$.

For the maximization part and the nontrivial case where $\theta \in (0, \frac{1}{2})$, following the above theorem, two types of BA algorithms can be designed according to the value of α.

When $\alpha \in (\frac{1-2\theta}{1-\theta}, 1]$, the original objective function is

$$
\begin{aligned}
F(\alpha, q) &= \theta I(X;Y|U) + \bar{\theta}\bar{\alpha} I(U;Y) + \bar{\theta}\alpha I(U;Z) \tag{15} \\
&= \bar{\theta}(I(X;Y|U) + \bar{\alpha} I(U;Y) + \alpha I(U;Z)) - (\bar{\theta} - \theta) I(X;Y|U) \\
&= \bar{\theta}(H(U,X) - H(X|Y,U) - \bar{\alpha} H(U|Y) - \alpha H(U|Z)) \\
&\quad + (\bar{\theta} - \theta)(H(Y|X) - H(Y|U)) \tag{16} \\
&= \bar{\theta} \cdot \sum_{u,x,y,z} q(u,x) p(y,z|x) \Big(\ln \frac{q(x|y,u)}{q(u,x)} q(u|y)^{\bar{\alpha}} q(u|z)^{\alpha} + \frac{\bar{\theta} - \theta}{\bar{\theta}} \ln \frac{q(y|u)}{p(y|x)} \Big).
\end{aligned}
$$

By replacing the conditional qs with free variables Qs, we have the new objective function

$$
F(\alpha, q, Q) = \bar{\theta} \cdot \sum_{u,x} q(u,x) (d[Q](u,x) - \ln q(u,x)),
\tag{17}
$$

where

$$d[Q](u,x) = \sum_{y,z} p(y,z|x) \left(\ln Q(x|y,u)Q(u|y)^{\bar{\alpha}}Q(u|z)^{\alpha} + \frac{\bar{\theta}-\theta}{\bar{\theta}} \ln \frac{Q(y|u)}{p(y|x)} \right). \tag{18}$$

When $\alpha \in [0, \frac{1-2\theta}{1-\theta}]$, similar to Equation (4), the new objective function is

$$F(\alpha, q, Q) = \bar{\theta} \cdot \sum_x q(x)(d[Q](x) - \ln q(x)), \tag{19}$$

where

$$d[Q](x) = \sum_{y,z} p(y,z|x)(\bar{\alpha} \ln Q(x|y) + \alpha \ln Q(x|z)). \tag{20}$$

For the minimization part, it is possible to use the optimal q and Qs obtained in the maximization part to update α. Because the values of the optimal q and Qs may vary greatly when α changes, we propose changing α locally in the neighbourhood. A candidate approach is to use the gradient descent method, as follows:

$$\alpha_{k+1} = \alpha_k - \tau_k \cdot \frac{\partial}{\partial \alpha} F(\alpha_k, q), \tag{21}$$

where

$$\frac{\partial F(\alpha, q)}{\partial \alpha} = \begin{cases} \bar{\theta} \cdot (I(X;Z) - I(X;Y)), & \text{if} \quad 0 \le \alpha \le \frac{1-2\theta}{1-\theta}, \\ \bar{\theta} \cdot (I(U;Z) - I(U;Y)), & \text{if} \quad \frac{1-2\theta}{1-\theta} < \alpha \le 1. \end{cases} \tag{22}$$

If the change in α is sufficiently small, it can be assumed that the optimization with respect to α converges and then stop the iteration.

We summarize the above procedures in Algorithm 2. Note that the updating rules for the q and Qs depend on the interval in which the value of α falls.

Algorithm 2: Computing the superposition coding inner bound for $\theta \in (0, \frac{1}{2})$

Input: $p(y,z|x)$, maximum iterations K, N, thresholds η, $\epsilon > 0$, step size $\tau > 0$;
Initialization: $\alpha_0 \in (0,1)$, $q_0(u,x) > 0$, $\eta_\alpha > \eta$, $k = 0$;
while $k < K$ *and* $\eta_\alpha > \eta$ **do**
    initialize $\epsilon_q > \epsilon$, $n = 0$;
    while $n < N$ *and* $\epsilon_q > \epsilon$ **do**
        $n \leftarrow n + 1$;
        $Q_n = Q[q_{n-1}]$ using Equation (5) similarly;
        $q_n = q[Q_n]$ using Equation (7) similarly;
        $F(\alpha_k, q_n, Q_n) = \bar{\theta} \cdot \ln \sum_{u,x} \exp\{d[Q_n]\}$ using Equation (20) or (18);
        $\epsilon_q = \bar{\theta} \cdot \max\{d[Q_n] - \ln q_{n-1}\} - F(\alpha_k, q_n, Q_n)$;
    end
    $k \leftarrow k + 1$;
    calculate α_k using Equations (21) and (22);
    $\alpha_k \leftarrow \min\{1, \max\{0, \alpha_k\}\}$;
    $\eta_\alpha = |\alpha_k - \alpha_{k-1}|$;
    $q_0 \leftarrow q_n$;
end
Output: α_k, $q_n(u,x)$, Q_n, $F(\alpha_{k-1}, q_n, Q_n)$

Remark 3. *Given α_k, according to Equation (8),*

$$F(\alpha_k, q_n, Q_n) = \bar{\theta} \cdot \ln \sum_{u,x} \exp\{d[Q_n](u,x)\},$$

$$\implies \exp\{\frac{1}{\theta}F(\alpha_k, q_n, Q_n)\} = \sum_{u,x} \exp\{d[Q_n](u,x)\}.$$

According to Equation (18) *or* (20), $\sum_{u,x} \exp\{d[Q_n]\}$ *equals a sum of exponents, as it is a function of α_k. This is a simple function; thus, we might wonder whether we can minimize $F(\alpha_k, q_n, Q_n)$ to update α_k. It turns out that this kind of global updating rule can result in an oscillating effect, as can be observed from Figure 1 in [7]. The main reason for this is that $q[Q]$ depends locally on α; therefore, it is not suitable to update α globally.*

3.2. Marton's Inner Bound

Marton's inner bound [4] refers to the union over $q(u, v, w, x)$ (such that $(U, V, W) \to X \to (Y, Z)$ is Markov) of the non-negative rate pairs (R_Y, R_Z) satisfying

$$R_Y \le I(U, W; Y),$$
$$R_Z \le I(V, W; Z),$$
$$R_Y + R_Z \le \min\{I(W; Y), I(W; Z)\} + I(U; Y|W) + I(V; Z|W) - I(U; V|W).$$

For general broadcast channels, this is the most well known inner bound.

In the following, we characterize the supporting hyperplane $\theta R_Y + \bar{\theta}R_Z = M$ of MIB. Because the expressions in MIB have symmetry in Y and Z, without loss of generality we can assume that $\theta \le \bar{\theta}$, i.e., $\theta \in [0, \frac{1}{2}]$. According to [20], the supporting hyperplane is stated in the following lemma.

Lemma 3 (Equations (2) and (5) in [20]). *The supporting hyperplane $\theta R_Y + \bar{\theta}R_Z = M$ of Marton's inner bound, where $\theta \in [0, \frac{1}{2}]$, is*

$$M = \max_{q(u,v,w,x)} \min_{\alpha \in [0,1]} M(\alpha, q(u, v, w, x))$$
$$= \min_{\alpha \in [0,1]} \max_{q(u,v,w,x)} M(\alpha, q(u, v, w, x)),$$

where

$$M(\alpha, q) = (\bar{\theta} - \alpha\theta)I(W; Z) + \alpha\theta I(W; Y) + \bar{\theta}I(V; Z|W) + \theta I(U; Y|W) - \theta I(U; V|W). \quad (23)$$

Further, it suffices to consider the following cardinalities: $|\mathcal{U}|, |\mathcal{V}| \le |\mathcal{X}|$, and $|\mathcal{W}| \le |\mathcal{X}| + 4$.

To compute the value of this supporting hyperplane, we can reformulate $M(\alpha, q)$ as follows:

$$M(\alpha, q) = \bar{\theta}H(U, V, W, X) - (\bar{\theta} - \alpha\theta)H(W|Z) - \alpha\theta H(W|Y) - \bar{\theta}H(V|W, Z)$$
$$- \theta H(U|W, Y) - (\bar{\theta} - \theta)H(U|V, W) - \bar{\theta}H(X|U, V, W). \quad (24)$$

Then, the objective function $M(\alpha, q, Q)$ can be expressed as

$$M(\alpha, q, Q) = \bar{\theta} \sum_{u,v,w,x} q(u, v, w, x)(d[Q](u, v, w, x) - \ln q(u, v, w, x)), \quad (25)$$

where

$$d[Q] = \sum_{y,z} p(y,z|x) \Big(\frac{\bar{\theta} - \alpha\theta}{\bar{\theta}} \ln Q(w|z) + \frac{\alpha\theta}{\bar{\theta}} \ln Q(w|y) + \ln Q(v|w,z)$$

$$+ \frac{\theta}{\bar{\theta}} \ln Q(u|w,y) + \frac{\bar{\theta} - \theta}{\bar{\theta}} \ln Q(u|v,w) + \ln Q(x|u,v,w) \Big). \tag{26}$$

For minimization over α, similar to Section 3.1, we update α along the gradient

$$\alpha_{k+1} = \alpha_k - \tau_k \cdot \frac{\partial}{\partial \alpha} M(\alpha_k, q) = \alpha_k - \tau_k \cdot \theta \cdot (I(W;Y) - I(W;Z)). \tag{27}$$

Similar to Algorithm 2, we summarize the algorithm for MIB in Algorithm 3.

Algorithm 3: Computing Marton's inner bound for $\theta \in (0, \frac{1}{2}]$

Input: $p(y,z|x)$, maximum iterations K, N, thresholds $\eta, \epsilon > 0$, step size $\tau > 0$;
Initialization: $\alpha_0 \in (0,1)$, $q_0(u,v,w,x) > 0$, $\eta_\alpha > \eta$, $k = 0$;
while $k < K$ *and* $\eta_\alpha > \eta$ **do**
    initialize $\epsilon_q > \epsilon$, $n = 0$;
    while $n < N$ *and* $\epsilon_q > \epsilon$ **do**
        $n \leftarrow n + 1$;
        $Q_n = Q[q_{n-1}]$ using Equation (5) similarly;
        $q_n = q[Q_n]$ using Equation (7) similarly;
        $M(\alpha_k, q_n, Q_n) = \bar{\theta} \cdot \ln \sum_{u,v,w,x} \exp\{d[Q_n]\}$ using Equation (26);
        $\epsilon_q = \bar{\theta} \cdot \max\{d[Q_n] - \ln q_{n-1}\} - M(\alpha_k, q_n, Q_n)$;
    end
    $k \leftarrow k + 1$;
    calculate α_k using Equation (27);
    $\alpha_k \leftarrow \min\{1, \max\{0, \alpha_k\}\}$;
    $\eta_\alpha = |\alpha_k - \alpha_{k-1}|$;
    $q_0 \leftarrow q_n$;
end
Output: α_k, $q_n(u,v,w,x)$, Q_n, $M(\alpha_{k-1}, q_n, Q_n)$

3.3. UV Outer Bound

The UV outer bound [5] refers to the union over $q(u,v,x)$ of non-negative rate pairs (R_Y, R_Z) satisfying

$$R_Y \leq I(U;Y),$$
$$R_Z \leq I(V;Z),$$
$$R_Y + R_Z \leq I(U;Y) + I(X;Z|U),$$
$$R_Y + R_Z \leq I(V;Z) + I(X;Y|V).$$

For general broadcast channels, this was the best outer bound until [6] strictly improved upon it over an erasure Blackwell channel. The following theorem (proof in Appendix B) characterizes the supporting hyperplanes.

Theorem 4 (Claim 2 and Remark 1 in [21]). *The supporting hyperplane of the UV outer bound is*

$$\theta R_Y + \bar{\theta} R_Z = \min_{\alpha, \beta} \max_{q(u,v,x)} \bar{\theta}\alpha I(V;Z) + \bar{\theta}\bar{\alpha} I(X;Z|U) + \theta\beta I(U;Y) + \theta\bar{\beta} I(X;Y|V), \tag{28}$$

where $\alpha, \beta \in [0,1]$ satisfy $\bar{\theta}\alpha + \theta\beta \geq \max\{\theta, \bar{\theta}\}$. Further, it suffices to consider the cardinality sizes $|\mathcal{U}|, |\mathcal{V}| \leq |\mathcal{X}|$.

The original objective function can be reformulated as

$$
\begin{aligned}
G(\alpha, \beta, q) = {}& \bar{\theta}\alpha(I(X;Y|V) + I(V;Z)) - (\bar{\theta}\alpha - \theta\bar{\beta})I(X;Y|V) \\
& + \theta\beta(I(X;Z|U) + I(U;Y)) - (\theta\beta - \bar{\theta}\bar{\alpha})I(X;Z|U).
\end{aligned}
$$

The right-hand side contains two parts, both of which are similar to Equation (10), i.e., the objective function of the degraded broadcast channel. It seems workable to apply the BA algorithm twice, however, it should be noted that these two parts are coupled by the same $q(x)$.

Observe that the first part depends only on $q(v, x)$, while the other depends on $q(u, x)$. It suffices to consider the subset of distributions such that $q(u, v, x) = q(x)q(v|x)q(u|x)$. Thus, it is natural to decouple these two parts by fixing $q(x)$ and applying the BA algorithm separately to $q(v|x)$ and $q(u|x)$. After some manipulations, we have

$$
\begin{aligned}
& G(\alpha, \beta, q, Q) \\
&= (\bar{\theta}\alpha + \theta\beta)H(X) \\
&\quad + \bar{\theta}\alpha(H(V|X) - H(X|Y, V) - H(V|Z)) - (\bar{\theta}\alpha - \theta\bar{\beta})(H(Y|V) - H(Y|X)) \\
&\quad + \theta\beta(H(U|X) - H(X|Z, U) - H(U|Y)) - (\theta\beta - \bar{\theta}\bar{\alpha})(H(Z|U) - H(Z|X)) \quad (29) \\
&= -(\bar{\theta}\alpha + \theta\beta)\sum_x q(x)\ln q(x) \\
&\quad + \bar{\theta}\alpha \sum_{x,v} q(x)q(v|x)(d_1[Q](v, x) - \ln q(v|x)) \\
&\quad + \theta\beta \sum_{x,u} q(x)q(u|x)(d_2[Q](u, x) - \ln q(u|x)). \quad (30)
\end{aligned}
$$

The functions d_1 and d_2 in the above are

$$
d_1[Q](v, x) = \sum_{y,z} p(y, z|x)\left(\ln Q(x|y, v)Q(v|z) + \frac{\bar{\theta}\alpha - \theta\bar{\beta}}{\bar{\theta}\alpha}\ln \frac{Q(y|v)}{p(y|x)}\right), \quad (31)
$$

$$
d_2[Q](u, x) = \sum_{y,z} p(y, z|x)\left(\ln Q(x|z, u)Q(u|y) + \frac{\theta\beta - \bar{\theta}\bar{\alpha}}{\theta\beta}\ln \frac{Q(z|u)}{p(z|x)}\right). \quad (32)
$$

For fixed $q(x)q(v|x)q(u|x)$, according to Equation (5), the optimal Qs are induced $Q[q]$s. For fixed Qs, according to Equation (7), for each x we have

$$
q[Q](v|x) = \frac{\exp\{d_1[Q](v, x)\}}{\sum_{v'} \exp\{d_1[Q](v', x)\}}, \quad (33)
$$

$$
q[Q](u|x) = \frac{\exp\{d_2[Q](u, x)\}}{\sum_{u'} \exp\{d_2[Q](u', x)\}}. \quad (34)
$$

The value of the objective function is

$$
G(\alpha, \beta, q(x), q[Q](v|x), q[Q](u|x), Q) = (\bar{\theta}\alpha + \theta\beta)\sum_x q(x)(d[Q](x) - \ln q(x)), \quad (35)
$$

where

$$
d[Q](x) = \frac{\bar{\theta}\alpha}{\bar{\theta}\alpha + \theta\beta}\ln \sum_v \exp\{d_1[Q](v, x)\} + \frac{\theta\beta}{\bar{\theta}\alpha + \theta\beta}\ln \sum_u \exp\{d_2[Q](u, x)\}. \quad (36)
$$

Again, according to Equation (7) the optimal $q[Q](x)$ and corresponding function value are

$$
q[Q](x) = \frac{\exp\{d[Q](x)\}}{\sum_{x'} \exp\{d[Q](x')\}}, \quad (37)
$$

$$
G(\alpha, \beta, q[Q], Q) = (\bar{\theta}\alpha + \theta\beta)\ln \sum_x \exp\{d[Q](x)\}. \quad (38)
$$

For minimization over (α, β), similar to Section 3.1, we update (α, β) along the gradient:

$$\alpha_{k+1} = \alpha_k - \tau_k \cdot \frac{\partial}{\partial \alpha} G(\alpha_k, \beta_k, q) = \alpha_k - \tau_k \cdot \bar{\theta} \cdot (I(V;Z) - I(X;Z|U)), \quad (39)$$

$$\beta_{k+1} = \beta_k - \tau_k \cdot \frac{\partial}{\partial \beta} G(\alpha_k, \beta_k, q) = \beta_k - \tau_k \cdot \theta \cdot (I(U;Y) - I(X;Y|V)). \quad (40)$$

Here, it should be mentioned that (α, β) must satisfy the constraint $\bar{\theta}\alpha + \theta\beta \geq \max\{\theta, \bar{\theta}\}$. Thus, if the resulting $(\alpha_{k+1}, \beta_{k+1})$ violate this constraint, then we need to scale $\bar{\theta}\alpha_{k+1} + \theta\beta_{k+1}$ up to be (at least) equal to $\max\{\theta, \bar{\theta}\}$. One way to accomplish this is to use the equality to make β dependent on α, in which case the gradient descent update becomes $\alpha_{k+1} = \alpha_k - \tau_k \cdot d_k, \beta_{k+1} = \beta_k - \tau_k \cdot (-\bar{\theta}d_k/\theta)$, where

$$d_k = \frac{\partial}{\partial \alpha} G(\alpha_k, \beta_k, q) - \frac{\bar{\theta}}{\theta} \cdot \frac{\partial}{\partial \beta} G(\alpha_k, \beta_k, q).$$

Similar to Algorithm 2, we summarize the algorithm for UVOB in Algorithm 4.

Algorithm 4: Computing the UV outer bound

Input: $p(y, z|x)$, maximum iterations K, N, thresholds $\eta, \epsilon > 0$, step size $\tau > 0$;
Initialization: $\alpha_0, \beta_0 \in (0,1)$, $q_0(u, v, x) > 0$, $\bar{\theta}\alpha_0 + \theta\beta_0 \geq \max\{\theta, \bar{\theta}\}$, $\eta_\alpha, \eta_\beta > \eta$, $k = 0$;
while $k < K$ *and* $\max\{\eta_\alpha, \eta_\beta\} > \eta$ **do**
　　initialize $\epsilon_q > \epsilon$, $n = 0$;
　　while $n < N$ *and* $\epsilon_q > \epsilon$ **do**
　　　　$n \leftarrow n + 1$;
　　　　$Q_n = Q[q_{n-1}]$ using Equation (5) similarly;
　　　　$q_n = q[Q_n]$ using Equations (33), (34) and (37);
　　　　$G(\alpha_k, \beta_k, q_n, Q_n) = (\bar{\theta}\alpha + \theta\beta) \cdot \ln \sum_x \exp\{d[Q_n]\}$ using Equation (36);
　　　　$\epsilon_q = (\bar{\theta}\alpha + \theta\beta) \cdot \max\{d[Q_n](x) - \ln q_{n-1}(x)\} - G(\alpha_k, \beta_k, q_n, Q_n)$;
　　end
　　$k \leftarrow k + 1$;
　　calculate α_k and β_k using Equations (39) and (40);
　　$\alpha_k \leftarrow \min\{1, \max\{0, \alpha_k\}\}$;
　　$\beta_k \leftarrow \min\{1, \max\{0, \beta_k\}\}$;
　　if $\bar{\theta}\alpha_k + \theta\beta_k < \max\{\theta, \bar{\theta}\}$, scale up to equality;
　　$\eta_\alpha = |\alpha_k - \alpha_{k-1}|$;
　　$\eta_\beta = |\beta_k - \beta_{k-1}|$;
　　$q_0 \leftarrow q_n$;
end
Output: $\alpha_k, \beta_k, q_n(u, v, x), Q_n, G(\alpha_{k-1}, \beta_{k-1}, q_n, Q_n)$

---

4. Convergence Analysis

Here, we aim to show that certain convergence results hold if q_n lies in a proper convex set which contains the global maximizer q_*. For this purpose, we first introduce the first-order characterization of a concave function.

Lemma 4 (Lemma 3 in [22])**.** *Given a convex set $\mathcal{S}$, a differentiable function f is concave in $\mathcal{S}$ if and only if, for all $x, y \in \mathcal{S}$,*

$$f(y) - f(x) - (y - x)^T \nabla f|_x \leq 0. \quad (41)$$

Similar to [22], we use the superlevel set to construct the convex set $\mathcal{S}$. Let $\mathcal{S}_F(\alpha, k)$ be the superlevel set of the objective function $F(\alpha, q)$ of SCIB:

$$\mathcal{S}_F(\alpha, k) := \{q | F(\alpha, q) \geq k\}. \quad (42)$$

For a fixed k, it is possible for $\mathcal{S}_F(\alpha, k)$ to contain more than one connected set. For $q \in \mathcal{S}_F(\alpha, k)$, we denote the connected set that contains q as $\mathcal{T}_F(\alpha, k, q)$.

Similarly, for MIB and UVOB we define the corresponding (connected) superlevel sets: $\mathcal{S}_M(\alpha, k)$, $\mathcal{T}_M(\alpha, k, q)$, $\mathcal{S}_G(\alpha, \beta, k)$, $\mathcal{T}_G(\alpha, \beta, k, q)$. Note that k here should not be confused with the notation indicating the number of iterations in the algorithms.

4.1. Superposition Coding Inner Bound

According to Theorem 3, the expression of the objective function $F(\alpha, q)$ of SCIB depends on the value of α. Without loss of generality, we can consider the objective function depicted in Equation (16). An equivalent condition for $F(\alpha, q)$ to be concave is provided in the following lemma.

Lemma 5. *Given a convex set $\mathcal{S}$ with a distribution $q(u, x)$, then $F(\alpha, q)$ as depicted in Equation (16) is concave in $\mathcal{S}$ if and only if, for all $q_1, q_2 \in \mathcal{S}$, we have*

$$\bar{\theta}(-D_{UX} + D_{X|YU} + \bar{\alpha}D_{U|Y} + \alpha D_{U|Z}) + (\bar{\theta} - \theta)D_{Y|U} \leq 0,$$

where $D_{A|B}$ denotes $D_{A|B}(q_2||q_1)$.

The following lemma shows that q_{n+1} lies in the same connected superlevel set as that of q_n. The proof (see Appendix C) is similar to that for Lemma 4 in [16].

Lemma 6. *In Algorithm 2, if $q_n(u, x) \in \mathcal{S}_F(\alpha, k)$, then $q_{n+1} \in \mathcal{T}_F(\alpha, k, q_n)$.*

Fixing α and letting $q_*(u, x)$ be the maximizer, the following theorem states that the function values $F(\alpha, q_n, Q_n)$ converge. The proof (see Appendix C) is similar to that of Theorem 2.

Theorem 5. *If $q_*, q_0 \in \mathcal{T}_F(\alpha, k, \tilde{q})$ for some k and $\tilde{q}$, and if $F(\alpha, q)$ is concave in $\mathcal{T}_F(\alpha, k, \tilde{q})$, then the sequence $F(\alpha, q_n, Q_n)$ generated by Algorithm 2 converges monotonically from below to $F(\alpha, q_*)$.*

The following corollary is implied by the proof of Theorem 5.

Corollary 1. *If $q_*, q_0 \in \mathcal{T}_F(\alpha, k, \tilde{q})$ for some k and $\tilde{q}$, and if $F(\alpha, q)$ is concave in $\mathcal{T}_F(\alpha, k, \tilde{q})$, then*

$$F(\alpha, q_*) - F(\alpha, q_N, Q_N) \leq \frac{\bar{\theta}}{N} \cdot D_{UX}(q_*||q_0).$$

The above analyses deal with $F(\alpha, q_n, Q_n)$ for a fixed α. When α_m changes to α_{m+1}, the estimation for the one-step change in the function value is presented in the following proposition (see proof in Appendix C).

Proposition 2. *Given α_m, suppose that Algorithm 2 converges to the optimal variables $\tilde{q}_*$ and $\tilde{Q}_*$ such that $\tilde{q}_* = q[\tilde{Q}_*]$ and $\tilde{Q}_* = Q[\tilde{q}_*]$. Letting α_{m+1} be updated using Equation (22) and letting $q_0 = \tilde{q}_*$ be the initial point for the next round, we have*

$$F(\alpha_{m+1}, q_1, Q_1) \approx F(\alpha_m, \tilde{q}_*) - \frac{(\alpha_{m+1} - \alpha_m)^2}{\tau_m}.$$

4.2. Marton's Inner Bound

Next, we present the convergence results of the BA algorithm for MIB. The proofs are omitted, as they are similar to those for SCIB.

Lemma 7. *Given a convex set $\mathcal{S}$ with a distribution $q(u, v, w, x)$, $M(\alpha, q)$ as depicted in Equation (24) is concave in $\mathcal{S}$ if and only if, for all q_1, $q_2 \in \mathcal{S}$, we have*

$$
-\bar{\theta} D_{UVWX} + (\bar{\theta} - \alpha\theta) D_{W|Z} + \alpha\theta D_{W|Y} + \bar{\theta} D_{V|WZ}
$$
$$
+ \theta D_{U|WY} + (\bar{\theta} - \theta) D_{U|VW} + \bar{\theta} D_{X|UVW} \le 0,
$$

where $D_{A|B}$ denotes $D_{A|B}(q_2||q_1)$.

Lemma 8. *In Algorithm 3, if $q_n(u, v, w, x) \in \mathcal{S}_M(\alpha, k)$, then $q_{n+1} \in \mathcal{T}_M(\alpha, k, q_n)$.*

Fixing α and letting $q_*(u, v, w, x)$ be the maximizer, the following theorem states that the function values $M(\alpha, q_n, Q_n)$ converge.

Theorem 6. *If $q_*, q_0 \in \mathcal{T}_M(\alpha, k, \tilde{q})$ for some k and $\tilde{q}$ and if $M(\alpha, q)$ is concave in $\mathcal{T}_M(\alpha, k, \tilde{q})$, then the sequence $M(\alpha, q_n, Q_n)$ generated by Algorithm 3 converges monotonically from below to $M(\alpha, q_*)$.*

The following corollary is implied by the proof of Theorem 6.

Corollary 2. *If $q_*, q_0 \in \mathcal{T}_M(\alpha, k, \tilde{q})$ for some k and $\tilde{q}$, and if $M(\alpha, q)$ is concave in $\mathcal{T}_M(\alpha, k, \tilde{q})$, then*

$$
M(\alpha, q_*) - M(\alpha, q_N, Q_N) \le \frac{\bar{\theta}}{N} \cdot D_{UVWX}(q_*||q_0).
$$

The estimation for the one-step change in the function value for MIB is presented in the following proposition.

Proposition 3. *Given α_m, suppose that Algorithm 3 converges to the optimal variables $\tilde{q}_*$ and $\tilde{Q}_*$ such that $\tilde{q}_* = q[\tilde{Q}_*]$ and $\tilde{Q}_* = Q[\tilde{q}_*]$. Letting α_{m+1} be updated using Equation (27) and letting $q_0 = \tilde{q}_*$ be the initial point for the next round, we have*

$$
M(\alpha_{m+1}, q_1, Q_1) \approx M(\alpha_m, \tilde{q}_*) - \frac{(\alpha_{m+1} - \alpha_m)^2}{\tau_m}.
$$

4.3. UV Outer Bound

Now, we present the convergence results of of the BA algorithm for UVOB. The proofs are again omitted, as they are similar to those of SCIB.

Lemma 9. *Given a convex set $\mathcal{S}$ of distribution $q(u, v, x)$, $G(\alpha, \beta, q)$ as depicted in Equation (29) is concave in $\mathcal{S}$ if and only if, for all q_1, $q_2 \in \mathcal{S}$, we have*

$$
-(\bar{\theta}\alpha + \theta\beta) D_X + \bar{\theta}\alpha(-D_{V|X} + D_{X|YV} + D_{V|Z}) + (\bar{\theta}\alpha - \theta\bar{\beta}) D_{Y|V}
$$
$$
+ \theta\beta(-D_{U|X} + D_{X|ZU} + D_{U|Y}) + (\theta\beta - \bar{\theta}\bar{\alpha}) D_{Z|U} \le 0,
$$

where $D_{A|B}$ denotes $D_{A|B}(q_2||q_1)$.

Lemma 10. *In Algorithm 4, if $q_n(u, v, x) \in \mathcal{S}_G(\alpha, \beta, k)$, then $q_{n+1} \in \mathcal{T}_G(\alpha, \beta, k, q_n)$.*

Fixing (α, β) and letting $q_*(u, v, x)$ be the maximizer, the following theorem states that the function values $G(\alpha, \beta, q_n, Q_n)$ converge.

Theorem 7. *If $q_*, q_0 \in \mathcal{T}_G(\alpha, \beta, k, \tilde{q})$ for some k and $\tilde{q}$, and if $G(\alpha, \beta, q)$ is concave in $\mathcal{T}_G(\alpha, \beta, k, \tilde{q})$, then the sequence $G(\alpha, \beta, q_n, Q_n)$ generated by Algorithm 4 converges monotonically from below to $G(\alpha, \beta, q_*)$.*

The following corollary is implied by the proof of Theorem 7.

Corollary 3. *If $q_*, q_0 \in \mathcal{T}_G(\alpha, \beta, k, \tilde{q})$ for some k and $\tilde{q}$, and if $G(\alpha, \beta, q)$ is concave in $\mathcal{T}_G(\alpha, \beta, k, \tilde{q})$, then*

$$G(\alpha, \beta, q_*) - G(\alpha, \beta, q_N, Q_N) \leq \frac{\bar{\theta}\alpha + \theta\beta}{N} \cdot D_{UVX}(q_*||q_0).$$

The estimation for the one-step change in the function value for UVOB is presented in the following proposition.

Proposition 4. *Given (α_m, β_m), suppose that Algorithm 4 converges to the optimal variables $\tilde{q}_*$ and $\tilde{Q}_*$ such that $\tilde{q}_* = q[\tilde{Q}_*]$ and $\tilde{Q}_* = Q[\tilde{q}_*]$. Letting $(\alpha_{m+1}, \beta_{m+1})$ be updated using Equations (39) and (40) and letting $q_0 = \tilde{q}_*$ be the initial point for the next round, we have*

$$G(\alpha_{m+1}, \beta_{m+1}, q_1, Q_1) \approx G(\alpha_m, \beta_m, \tilde{q}_*) - \frac{(\alpha_{m+1} - \alpha_m)^2}{\tau_m} - \frac{(\beta_{m+1} - \beta_m)^2}{\tau_m}. \tag{43}$$

5. Numerical Results

We take the binary skew-symmetric broadcast channel $p(y, z|x)$ as the test channel. The conditional probability matrices are

$$P_{Y|X} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}, \quad P_{Z|X} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \end{bmatrix}.$$

This is perhaps the simplest broadcast channel for which the capacity region is still unknown.

This broadcast channel plays a very important role in research on capacity bounds. It was first studied in [23] to show that the time-sharing random variable is useful for the Cover–van der Meulen inner bound [24,25]. Later, [26–28] demonstrated that the sum rate of the UVOB for this broadcast channel is strictly larger than that of the MIB, showing for the first time that at least one of these two bounds are suboptimal.

Our algorithms are important in at least the following sense: supposing that it is not known whether the MIB matches the UVOB (or the other two bounds for a new scenario) and we want to check this; we can perform an exhaustive search on channel matrices of size 2×2 (or of higher dimensions) to check whether they match. According to the results shown below in Section 5.5, this does not take very much time compared with generic algorithms.

In the following, we apply the algorithms to compute the value of the supporting hyperplane $\theta R_Y + \bar{\theta} R_Z$, where $\theta = 0.4$. The initial values of α and β are $\alpha_0 = \beta_0 = 0.7$. This set of parameters is feasible for UVOB, as $\bar{\theta}\alpha + \theta\beta = 0.7 > \max\{\theta, \bar{\theta}\}$.

We demonstrate the algorithms in the following aspects: (1) the maximization part; (2) the minimization part; (3) the change from the maximum part to the minimization part; (4) the superlevel set; and (5) comparison with generic non-convex algorithms.

5.1. Maximization Part

In this part, we fix α and β to the initial values and let the BA algorithms iterate for $N = 200$ times. The results are presented in Figure 1. Because this is the maximization part, the function values increase as the iterations proceed. It is clear that the function values behave properly for fixed α and β.
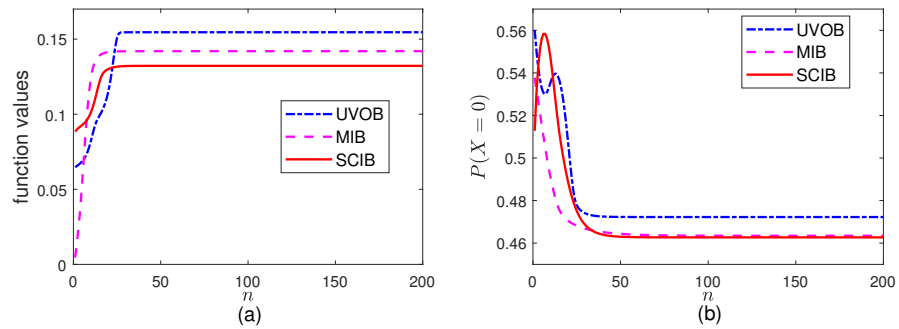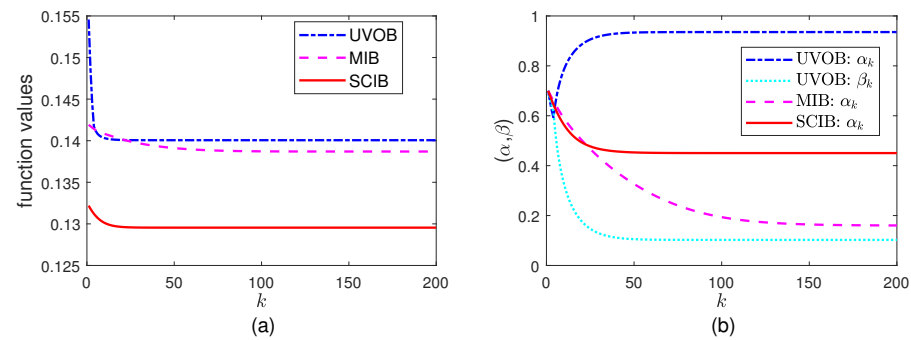
Figure 1. The maximization parts in the algorithms for BSSC with fixed values $\alpha_0 = \beta_0 = 0.7$: (**a**) the objective function values and (**b**) $P(X = 0)$.

5.2. Minimization Part

In this part, we start with the initial α_0 and β_0, then let the algorithms iterate for $K = 200$ times. The results for (α_k, β_k) are presented in Figure 2. Because this is the minimization part, the function values decrease as the iterations proceed. It is clear that α_k in SCIB and MIB gradually changes as k grows. For UVOB, it is necessary to ensure that $\bar{\theta}\alpha + \theta\beta \geq \max\{\theta, \bar{\theta}\}$. When the updated $(\alpha_{k+1}, \beta_{k+1})$ makes $\bar{\theta}\alpha + \theta\beta$ fall below this value, it becomes necessary to scale it back. This happens approximately starting from $k = 5$.

Figure 2. The minimization parts in the algorithms for BSSC with initial values $\alpha_0 = \beta_0 = 0.7$: (**a**) the objective function values and (**b**) (α_k, β_k).

5.3. Change from Maximization to Minimization

In this part, we consider UVOB and let K in the algorithm be $K = 100$. Figure 3 plots the following three values in Equation (43):

$$G(\alpha_k, \beta_k, \tilde{q}_*),$$
$$LHS := G(\alpha_{k+1}, \beta_{k+1}, q_1, Q_1),$$
$$RHS := G(\alpha_k, \beta_k, \tilde{q}_*) - \frac{(\alpha_{k+1} - \alpha_k)^2}{\tau_k} - \frac{(\beta_{k+1} - \beta_k)^2}{\tau_k}.$$

As the algorithm iterates, the estimate in Equation (43) becomes more and more accurate, as $\exp\{x\} \approx 1 + x$ and $\ln(1 + x) \approx x$ for small x.
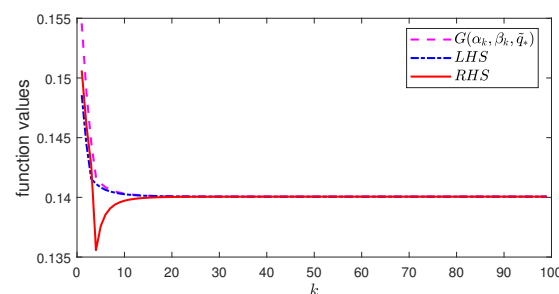

Figure 3. Function values of UVOB for BSSC with initial values $\alpha_0 = \beta_0 = 0.7$.

5.4. Superlevel Set

To visualize the convergence of q_n and its relation with the superlevel set, we take SCIB as an example and fix $q(u)$ such that $q(x|u)$ has two free variables. We reformulate the objective function of SCIB depicted in Equation (16) as follows:

$$\tilde{F}(\alpha, q(x|u), Q) = \bar{\theta}H(U) + \bar{\theta}(H(X|U) - H(X|Y, U) - \bar{\alpha}H(U|Y) - \alpha H(U|Z))$$
$$- (\bar{\theta} - \theta)(H(Y|U) - H(Y|X))$$
$$= \bar{\theta}H(U) + \bar{\theta}\sum_{x,u} q(u)q(x|u)(d[Q](u, x) - \ln q(x|u)),$$

where

$$d[Q](u, x) = \sum_{y,z} p(y, z|x)\left(\ln Q(x|y, u)Q(u|y)^{\bar{\alpha}}Q(u|z)^{\alpha} + \frac{\bar{\theta} - \theta}{\bar{\theta}}\ln\frac{Q(y|u)}{p(y|x)}\right).$$

In particular, we fix $\alpha_0 = 0.7$ and $P_U = (0.3, 0.7)$, then use the algorithm to find the values of $P(X = 0|U = 0)$ and $P(X = 0|U = 1)$. The results are shown in Figure 4. In this case, q_n for large enough n lies in the concave part of the superlevel set, meaning that the algorithm converges. Here, it should be mentioned that it is possible that the algorithm may not converge to the optimal point for some initial q_0s that do not lie in the concave part.
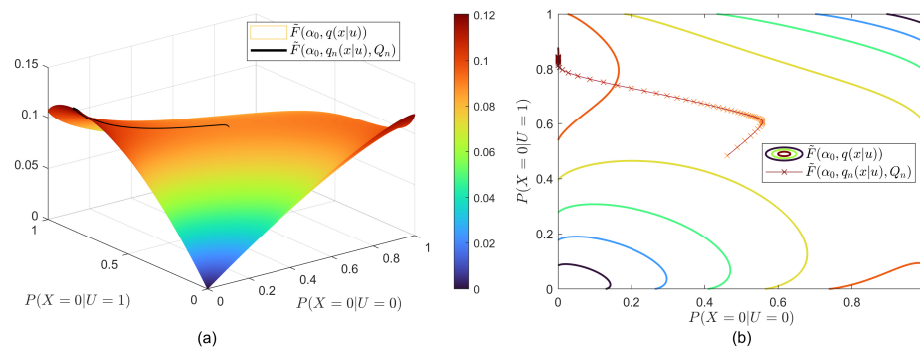

Figure 4. Function values of SCIB for BSSC with the initial value $\alpha_0 = 0.7$ and fixed probability vector $P_U = (0.3, 0.7)$: (**a**) 3D view and (**b**) contour view.

5.5. Comparison With Generic Non-Convex Algorithms

Here, we compare our algorithms with the following generic algorithms implemented using the "fmincon" MATLAB function: interior-point, active-set, and sequential quadratic programming (sqp). For simplicity, we only compare the sum rate of MIB, for which the optimal value is 0.2506717... nats (0.3616428... bits). The optimization problem for computing the sum rate is

$$\max_{q(u,v,w,x)} \min\{I(W; Y), I(W; Z)\} + I(U; Y|W) + I(V; Z|W) - I(U; V|W).$$

According to Lemma 3, the cardinality size is $|\mathcal{U}| \cdot |\mathcal{V}| \cdot |\mathcal{W}| \cdot |\mathcal{X}| = |\mathcal{X}|^3(|\mathcal{X}| + 4) = 48$.

Notice that we do not carry out a comparison with the method in [16], as it cannot be applied to cases where there is a minimum. For scenarios in which [16] can be used, our algorithms degenerate to the method in [16].

The initial point of $q(u, v, w, x)$ is randomly generated for all the algorithms. Table 2 lists the experimental results. For the first three algorithms, a randomly picked starting point usually does not provide a good enough result. Thus, we ran the first three algorithms multiple times until the best function value hit 0.2506 in order to test their effectiveness. It is clear from the table that only sqp can be considered comparable to our algorithms.

Table 2. Comparison with generic non-convex algorithms on BSSC.

Method	Time (Seconds)	Sum-Rate of MIB (Nats)
interior-point	513.82	0.25060...
active-set	2438.57	0.25061...
sqp	1.7621	0.25067...
this paper	0.0629	0.25067...

For further comparison with sqp, we randomly generated broadcast channels with cardinalities of $|\mathcal{X}| = 3, 4, 5, 6$, and $|\mathcal{Y}| = |\mathcal{Z}| = |\mathcal{X}|$. The corresponding dimensions are $|\mathcal{X}|^3(|\mathcal{X}| + 4) = 189, 512, 1125, 2160$. Because the optimal sum rate is not yet known, we ran sqp once to record the running time. The results in Table 3 suggest that our algorithms are highly scalable. This meets our expectation, as the updating formulae in Equations (5) and (7) are all explicit and can be computed rapidly.

Table 3. Comparison with sqp on random channels with alphabet sizes $|\mathcal{X}| = 3, 4, 5, 6$.

Method	Time (Seconds)				Sum-Rate of MIB (Nats)			
	$|\mathcal{X}| = 3$	$|\mathcal{X}| = 4$	$|\mathcal{X}| = 5$	$|\mathcal{X}| = 6$	$|\mathcal{X}| = 3$	$|\mathcal{X}| = 4$	$|\mathcal{X}| = 5$	$|\mathcal{X}| = 6$
sqp	2.6342	22.44	168.12	1065.51	0.1840	0.2348	0.1983	0.2351
this paper	0.0771	0.1031	0.1450	0.2086	0.1863	0.2375	0.2019	0.2423

6. Discussion and Conclusions

6.1. Initial Points of Algorithms

Taking MIB as an example, we next discuss how to choose the initial points. When there is no prior knowledge on the optimization problem, the initial point is usually generated randomly. In this paper, Theorem 6 and Lemma 7 provide some guidance on the choice of the initial point $p_0(u, v, w, x)$. A possibly workable method is to randomly generate an initial point and slightly perturb it to check whether these two points satisfy the inequality in Lemma 7. If the answer is no, then it is possible that the objective function is not concave in the neighbourhood of this point, and we continue to generate new initial points.

For the initial point α_0, because it lies in $[0, 1]$ it is affordable to perform a grid search, especially when $|\mathcal{X}|$ is small. For example, we can take 0.1 as the equal space and try each $\alpha_0 \in \{0, 0.1, 0.2, \ldots, 1\}$. This approach can to some extent help us avoid becoming stuck in local extreme points.

6.2. J Version Outer Bound

As mentioned earlier, the best general outer bound is the J version outer bound proposed in [6]. However, the evaluation of this outer bound turns out to be even harder, as there are additional constraints on the free variables and the auxiliary channel with the joint distribution

$$q(x)q(u, v, w|x)q(\tilde{u}, \tilde{v}, \tilde{w}|x)q(\hat{u}, \hat{v}, \hat{w}|x)p(y, z|x)T(j|x, y, z).$$

These constraints are presented in Equations (18a)–(18c) and (19a)–(19c) in [6]. Taking Equations (18a) and (19a) as an example,

$$(18a): \quad I(\tilde{W}; Z) - I(\tilde{W}; J) + I(\hat{W}; J) - I(\hat{W}; Y) = I(W; Z) - I(W; Y),$$
$$(19a): \quad 0 \leq I(X; Z|\tilde{U}, \tilde{W}) - I(X; J|\tilde{U}, \tilde{W}) \leq I(\tilde{V}; Z|\tilde{W}) - I(\tilde{V}; J|\tilde{W}).$$

Direct application of Equation (7) does not yield an updated $q[Q]$ guaranteed to satisfy these constraints; thus, the design of BA algorithms for the J version outer bound should carefully address this kind of problem. We leave this for future research.

Finally, to conclude our paper, the extension of the BA algorithm to inner and outer bounds for general broadcast channels encounters max–min problems. We have shown that the max–min order can be changed to min–max. Based on this observation, we have designed BA algorithms for the maximization parts and gradient descent algorithms for the minimization parts, then performed convergence analysis and numerical experiments to support our analysis. We have compared our algorithms to the following generic non-convex algorithms: interior-point, active-set, and sequential quadratic programming. The results show that our algorithms are both effective and efficient.

Author Contributions: Methodology and analysis, Y.G.; software and visualization, Y.D. and Y.L.; validation, X.N., B.B. and W.H.; writing—original draft preparation, Y.G.; writing—review and editing, X.N., B.B. and W.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: Yanqing Liu is employed by Bank of China. Xueyan Niu, Bo Bai, Wei Han are employed by Huawei Tech. Co., Ltd. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Proofs of Results in Section 1

Proof of Theorem 1. For the first property, Fox fixed q, the concavity is clear, as $\ln Q$ is concave in Q. The equality in Equation (6) is easy to show, as if we take $Q[q](x|y)$ to be $q(x|y)$ then the function $C(q, Q[q])$ in Equation (2) reduces to $C(q)$ in Equation (1). The maximum can be proved using the Kullback–Leibler divergence:

$$
\begin{aligned}
C(q, Q[q]) - C(q, Q) &= C(q) - C(q, Q) \\
&= \sum_{x,y} q(x) p(y|x) \ln \frac{q(x|y)}{Q(x|y)} \\
&= \sum_{x,y} q(y) q(x|y) \ln \frac{q(x|y)}{Q(x|y)} \\
&= D_{X|Y}(q||Q) \\
&\geq 0.
\end{aligned}
$$

For the second property, we can reformulate $C(q, Q)$ as

$$
C(q, Q) = H(X) - \sum_{x} q(x) \left(\sum_{y} p(y|x) \ln Q(x|y) \right).
$$

The concavity holds, as the first term $H(X)$ is concave and the second term is linear. To find the maximum $q[Q]$, because there is a constraint $\sum_{x} q(x) = 1$, we consider the derivative of the Lagrangian:

$$0 = \frac{\partial}{\partial q(x)} \mathcal{L}(\lambda, q, Q)$$

$$= \frac{\partial}{\partial q(x)} \sum_{x'} q(x')(d[Q](x') - \ln q(x')) - \lambda(1 - \sum_{x'} q(x'))$$

$$= d[Q](x) - \ln q(x) - 1 - \lambda.$$

This implies that $q(x) = \exp\{d[Q](x) - 1 - \lambda\}$ for all x. The common term $1 + \lambda$ can be eliminated by normalization, as depicted in Equation (7). We can then verify the function value as follows:

$$C(q[Q], Q) = \sum_x q[Q](x)(d[Q](x) - \ln q[Q](x))$$

$$\overset{(a)}{=} \sum_x q[Q](x)(\ln q[Q](x) + \ln \sum_{x'} \exp\{d[Q](x')\} - \ln q[Q](x))$$

$$= \ln \sum_{x'} \exp\{d[Q](x')\},$$

where (a) is due to Equation (7). The second equality is again per Equation (7). $\square$

Proof of Theorem 2. The basic idea is to show that the sum is bounded for decreasing and positive numbers $C - C(q_n, Q_n)$.

The monotonicity $C(q_{n+1}, Q_{n+1}) \geq C(q_n, Q_n)$ is clear, as we are performing alternating maximization; thus,

$$C(q_{n+1}, Q_{n+1}) = C(q[Q_{n+1}], Q_{n+1}) \geq C(q_n, Q_{n+1}) = C(q_n, Q[q_n]) \geq C(q_n, Q_n).$$

The positiveness of $C - C(q_n, Q_n)$ is because $C \geq C(q_n) \geq C(q_n, Q_n)$.

Let q_* achieve the maximum of $C(q)$; then,

$$C - C(q_n, Q_n) - \sum_x q_*(x) \ln \frac{q_n(x)}{q_{n-1}(x)}$$

$$= C(q_*, Q[q_*]) - C(q_n, Q_n) - \sum_x q_*(x) \ln \frac{q_n(x)}{q_{n-1}(x)}$$

$$\overset{(a)}{=} \sum_x q_*(x)(d[Q[q_*]](x) - \ln q_*(x) - d[Q_n](x) + \ln q_n(x) - \ln \frac{q_n(x)}{q_{n-1}(x)})$$

$$= \sum_x q_*(x)(d[Q[q_*]](x) - \ln q_*(x) - d[Q[q_{n-1}]](x) + \ln q_{n-1}(x))$$

$$\overset{(b)}{=} \sum_x p_*(x) \sum_y p(y|x)(\ln \frac{p(y|x)}{q_*(y)} - \ln \frac{p(y|x)}{q_{n-1}(y)}).$$

$$= \sum_{x,y} q_*(x) p(y|x) \ln \frac{q_{n-1}(y)}{q_*(y)}$$

$$= -D_Y(q_* \| q_{n-1})$$

$$\leq 0,$$

where (a) is due to Equation (8) and (b) holds according to Equation (9). Now, we can bound the sum as follows:

$$\sum_{n=1}^{N}(C - C(q_n, Q_n)) \leq \sum_{n=1}^{N}\sum_{x} q_*(x) \ln \frac{q_n(x)}{q_{n-1}(x)}$$

$$= \sum_{x} q_*(x) \ln \frac{q_N(x)}{q_0(x)}$$

$$= \sum_{x} q_*(x) \ln \frac{q_N(x)q_*(x)}{q_0(x)q_*(x)}$$

$$= D_X(q_*||q_0) - D_X(q_*||q_N)$$

$$\leq D_X(q_*||q_0).$$

The last term is finite, as $q_0 > 0$. This implies $C(q_n, Q_n) \to C$. $\square$

Appendix B. Proofs of Results in Section 3

Proof of Lemma 1. The equivalence $\mathcal{B} = \mathcal{C}$ is already stated without proof in Chapter 5.3 and Chapter 5.6 of [29]. We present the proof here for completeness.

It is clear that $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$. Thus, we only need to prove $\mathcal{C} \subseteq \mathcal{A}$. It suffices to show that the corner points of $\mathcal{C}(U, X)$ lie inside $\mathcal{A}$.

Because $\mathcal{C}(U, X)$ is a trapezoid with one corner point $(0, 0)$, there are at most three nontrivial corner points, as follows:

1. Lower right $(r_1, 0)$, where $r_1 = \min\{I(X;Y), I(X;Y|U) + I(U;Z)\}$. Because $r_1 \leq I(X;Y)$, we have $(r_1, 0) \in \mathcal{A}(\varnothing, X)$.
2. Upper left $(0, r_2)$, where $r_2 = \min\{I(U;Z), I(X;Y)\}$. Clearly, $(0, r_2) \in \mathcal{A}(X, X)$.
3. Upper right $(\min\{I(X;Y|U), I(X;Y) - I(U;Z)\}, I(U;Z))$; this corner point exists when $I(U;Z) \leq I(X;Y)$. We can consider two cases:

   (a) $I(U;Y) \geq I(U;Z)$: the corner point $(I(X;Y|U), I(U;Z))$ is inside $\mathcal{A}(U, X)$.
   (b) $I(U;Y) \leq I(U;Z)$: the corner point is $(I(X;Y) - I(U;Z), I(U;Z))$. It suffices to consider $I(U;Y) \leq I(U;Z) \leq I(X;Y)$, as otherwise this point does not exist. Now, let α be such that $(1 - \alpha)I(X;Y) + \alpha I(U;Y) = I(U;Z)$. We can construct a pair $(\hat{U}, \hat{X})$ such that this corner point is inside $\mathcal{A}(\hat{U}, \hat{X})$. Considering the random variables $Q \sim Bernoulli(\alpha)$, $\tilde{U} = X$ when $Q = 0$ and $\tilde{U} = U$ when $Q = 1$, we can let $\hat{U} = (\tilde{U}, Q)$, $\hat{X} = X$; then, $I(\hat{X};Y|\hat{U}) = 0 + \alpha I(X;Y|U) = I(X;Y) - I(U;Z)$. Now, we have $I(\hat{U};Z) \geq I(\tilde{U};Z|Q) \geq I(U;Z)$ and $I(\hat{U};Y) \geq I(\tilde{U};Y|Q) = I(U;Z)$; hence, $\min\{I(\hat{U};Z), I(\hat{U};Y)\} \geq I(U;Z)$.

The proof is finished. $\square$

Proof of Theorem 3. We can use $\mathcal{A}(U, X)$ in Lemma 1 to compute the supporting hyperplanes. When $\theta \in [\frac{1}{2}, 1]$, then $\theta R_Y + \bar{\theta} R_Z$ is bounded from above by

$$\max_{q(u,x)} \theta I(X;Y|U) + \bar{\theta} I(U;Y) = \max_{q(u,x)} \theta(I(X;Y) - I(U;Y)) + \bar{\theta} I(U;Y)$$

$$= \max_{q(u,x)} \theta I(X;Y) + (1 - 2\theta)I(U;Y)$$

$$\leq \max_{q(x)} \theta I(X;Y).$$

On the other hand, this upper bound is achieved by setting $U = \varnothing$ in $\mathcal{A}(U, X)$.

When $\theta \in [0, \frac{1}{2})$, we first have

$$F = \max_{p(u,x)} \min_{\alpha \in [0,1]} \theta I(X;Y|U) + \bar{\theta}\alpha I(U;Z) + \bar{\theta}\bar{\alpha} I(U;Y).$$

To show that the maximum and minimum can be exchanged, we use Lemma 2; in particular, letting $d = 2$, $\lambda_1 = \alpha$, and $\lambda_2 = \bar{\alpha}$, we have

$$T_1(q(u,x)) = \theta I(X;Y|U) + \bar{\theta} I(U;Z),$$
$$T_2(q(u,x)) = \theta I(X;Y|U) + \bar{\theta} I(U;Y).$$

Then, the objective function $F(\alpha, q)$ equals $\lambda_1 T_1 + \lambda_2 T_2$. It remains to prove that $\mathcal{T}$ is convex. Assuming that $(a_1, a_2) \in \mathcal{T}$ for some $q_a(u,x)$ and that $(b_1, b_2) \in \mathcal{T}$ for some $q_b(u,x)$ while letting $\tilde{U} = (U, Q)$, where $Q \sim Bernoulli(\beta)$ and $(U, X) \sim q_a(u,x)$ if $Q = 1$ and $(U, X) \sim q_b(u,x)$ if $Q = 0$, we then have

$$\begin{aligned} T_1(\beta q_a + \bar{\beta} q_b) &= T_1(q(\tilde{u}, x)) \\ &= \theta I(X;Y|U,Q) + \bar{\theta} I(U,Q;Z) \\ &\geq \theta I(X;Y|U,Q) + \bar{\theta} I(U;Z|Q) \\ &= \beta T_1(q_a) + \bar{\beta} T_1(q_b). \end{aligned}$$

Similar inequalities hold for T_2. This proves the convexity, and hence the exchange.

Now, we can show that the expression of F can be simplified for $\theta \in [0, \frac{1}{2})$ and $\alpha \leq \frac{1-2\theta}{1-\theta}$. Noting that $I(U;Y) = I(X;Y) - I(X;Y|U)$ and $I(U;Z) = I(X;Z) - I(X;Z|U)$, we have

$$\begin{aligned} &\theta I(X;Y|U) + \bar{\theta}\bar{\alpha} I(U;Y) + \bar{\theta}\alpha I(U;Z) \\ &= \bar{\theta}\bar{\alpha} I(X;Y) + \bar{\theta}\alpha I(X;Z) + (\theta - \bar{\theta}\bar{\alpha}) I(X;Y|U) - \bar{\theta}\alpha I(X;Z|U). \end{aligned}$$

When $\alpha \leq \frac{1-2\theta}{1-\theta}$, i.e., $\theta - \bar{\theta}\bar{\alpha} \leq 0$, the last two terms above have a non-positive sum. The maximum value equals zero, and can be achieved by taking $U = X$. This finishes the proof of the expression.

The cardinality can be proved as follows:

$$\begin{aligned} &\theta I(X;Y|U) + \bar{\theta}\bar{\alpha} I(U;Y) + \bar{\theta}\alpha I(U;Z) \\ &= \bar{\theta}\bar{\alpha} I(X;Y) + \bar{\theta}\alpha I(X;Z) + (\theta - \bar{\theta}\bar{\alpha}) I(X;Y|U) - \bar{\theta}\alpha I(X;Z|U) \\ &= f(q(x)) + \sum_u q(u) g(q(x|u)), \end{aligned}$$

where f and g are some continuous functions corresponding to the mutual information. Subject to fixed marginal $q(x)$, the maximum of $\sum_u q(u) g(q(x|u))$ over all feasible $q(u)$ and $q(x|u)$ is the upper concave envelope of the function g evaluated at $q(x)$. Notice that as the degree of freedom of the distribution $q(x)$ is $|\mathcal{X}| - 1$, it suffices to consider $|\mathcal{U}| \leq |\mathcal{X}| - 1 + 1 = |\mathcal{X}|$ for evaluating the envelope. $\square$

Proof of Theorem 4. For fixed $q(u,v,x)$, in the pentagon of the UVOB there are at most two corner points in the first quadrant, namely, the upper left and lower right ones. The line connecting these two points has slope -1. We need to compute the the supporting hyperplane value $G = \max_{(R_Y, R_Z) \in UVOB} \theta R_Y + \bar{\theta} R_Z$, where $\theta \in (0, 1)$.

For the case $\theta \in (0, \frac{1}{2}]$, note that the slope of the line $\theta R_Y + \bar{\theta} R_Z = G$ is $-\theta/\bar{\theta} \geq -1$; thus, it suffices to consider the upper left corner point. The expression of this point is different when $q(u,v,x)$ falls into one of the following two sets:

$$\begin{aligned} \mathcal{S}_1 &= \{q(u,v,x) : I(V;Z) \leq I(U;Y) + I(X;Z|U)\}, \\ \mathcal{S}_2 &= \{q(u,v,x) : I(V;Z) \geq I(U;Y) + I(X;Z|U)\}. \end{aligned}$$

When $q \in \mathcal{S}_1$, this corner point and the corresponding expression of $l_1(q) := \theta R_Y + \bar{\theta} R_Z$ are

$$\big(\min\{I(U;Y), I(U;Y) + I(X;Z|U) - I(V;Z), I(X;Y|V)\}, \, I(V;Z) \big),$$
$$l_1(q) = \theta \cdot \min\{I(U;Y), I(U;Y) + I(X;Z|U) - I(V;Z), I(X;Y|V)\} + \bar{\theta} I(V;Z).$$

Otherwise, when $q \in \mathcal{S}_2$, this corner point and corresponding expression of $l_2(q) := \theta R_Y + \bar{\theta} R_Z$ are

$$\left(0,\; I(U;Y) + I(X;Z|U)\right),$$
$$l_2(q) = \bar{\theta}\left(I(U;Y) + I(X;Z|U)\right).$$

Now, the supporting hyperplane value is

$$G = \max\{\max_{q \in \mathcal{S}_1} l_1(q), \quad \max_{q \in \mathcal{S}_2} l_2(q)\}.$$

We want to show that

$$\max_{q \in \mathcal{S}_2} l_2(q) = \max_{q \in \mathcal{S}_2} l_1(q).$$

For the left hand-side term, we have

$$\max_{q \in \mathcal{S}_2} l_2(q) = \max_{q \in \mathcal{S}_2} \bar{\theta}\left(I(U;Y) + I(X;Z|U)\right)$$
$$\leq \max_{q \in \mathcal{S}_2} \bar{\theta} I(V;Z)$$
$$\leq \max_{q \in \mathcal{S}_2} \bar{\theta} I(X;Z),$$

where the equalities hold with $U = \varnothing$, $V = X$ and this choice is in $\mathcal{S}_2$. For the other term,

$$\max_{q \in \mathcal{S}_2} l_1(q) = \max_{q \in \mathcal{S}_2} \theta(I(U;Y) + I(X;Z|U) - I(V;Z)) + \bar{\theta} I(V;Z)$$
$$\leq \max_{q \in \mathcal{S}_2} \bar{\theta} I(V;Z)$$
$$\leq \max_{q \in \mathcal{S}_2} \bar{\theta} I(X;Z),$$

where the equalities hold using the same settings as above. Hence, we have the supporting hyperplane value

$$G = \max\{\max_{q \in \mathcal{S}_1} l_1(q), \quad \max_{q \in \mathcal{S}_2} l_1(q)\} = \max_q l_1(q).$$

We can simplify this expression as follows:

$$\max_q l_1(q) = \max_q \min_{a,b \in [0,1]} \theta(1-a)I(U;Y) + \theta a(1-b)\left(I(U;Y) + I(X;Z|U) - I(V;Z)\right)$$
$$+ \theta a b I(X;Y|V) + \bar{\theta} I(V;Z)$$
$$= \max_q \min_{a,b \in [0,1]} \theta(1-ab)I(U;Y) + \theta a(1-b)I(X;Z|U)$$
$$+ (\bar{\theta} - \theta a(1-b))I(V;Z) + \theta a b I(X;Y|V).$$

Letting $\alpha = 1 - \theta a(1-b)/\bar{\theta}$, $\beta = 1 - ab$, we have $\alpha, \beta \in [0,1]$, $\bar{\theta}\alpha + \theta\beta = \bar{\theta} + \theta(1-a)$, and the supporting hyperplane value is

$$\theta R_Y + \bar{\theta} R_Z = \max_q \min_{\alpha,\beta \in [0,1]} \bar{\theta}\alpha I(V;Z) + \bar{\theta}\bar{\alpha} I(X;Z|U) + \theta\beta I(U;Y) + \theta\bar{\beta} I(X;Y|V).$$

Notice that the range of $\bar{\theta}\alpha + \theta\beta$ is $[\max\{\theta, \bar{\theta}\}, 1]$. Within this range, the reverse mapping from (α, β) to (a, b) is

$$a = \frac{1 - (\bar{\theta}\alpha + \theta\beta)}{\theta}, \quad b = \frac{\theta\bar{\beta}}{\theta\bar{\beta} + \bar{\theta}\bar{\alpha}}.$$

Notice that when $\theta \in [\frac{1}{2}, 1)$, we have $\bar{\theta} \in (0, \frac{1}{2}]$; we can use similar reasoning as above (by swapping Y and Z, R_Y and R_Z, U and V, θ and $\bar{\theta}$, and finally α and β) to obtain

$$\bar{\theta} R_Z + \theta R_Y = \max_q \min_{\alpha,\beta \in [0,1]} \theta\beta I(U;Y) + \theta\bar{\beta} I(X;Y|V) + \bar{\theta}\alpha I(V;Z) + \bar{\theta}\bar{\alpha} I(X;Z|U).$$

The constraint then becomes $\theta\beta + \bar{\theta}\alpha = \theta + \bar{\theta}(1-a)$, for which the range is again $[\max\{\bar{\theta},\theta\},1]$. Thus, the expression and the constraints are the same as for the case where $\theta \in (0,\frac{1}{2}]$. Putting these two cases together, we have the characterization of the supporting hyperplanes.

To exchange the max–min, we again use Lemma 2. The proof is similar to that of Theorem 3, and as such we omit the details, providing only the setting for $\theta \in (0,\frac{1}{2}]$, where $G = \max_q l_1(q): d = 3$ and the functions are

$$T_1(q) = \theta I(U;Y) + \bar{\theta}I(V;Z),$$
$$T_2(q) = \theta(I(U;Y) + I(X;Z|U) - I(V;Z)) + \bar{\theta}I(V;Z),$$
$$T_3(q) = \theta I(X;Y|V) + \bar{\theta}I(V;Z).$$

The proof of the cardinality bounds is similar to that of Theorem 3. □

Appendix C. Proofs of Results in Section 4

Proof of Lemma 5. Let A represent a generic random variable. The gradient of $H(A)$ can be calculated as follows:

$$\frac{\partial H(A)}{\partial q(u,x)} = \frac{\partial}{\partial q(u,x)}\left(-\sum_a q(a)\ln q(a)\right)$$
$$= -\sum_a \frac{\partial q(a)}{\partial q(u,x)}\ln q(a) - \sum_a q(a)\frac{1}{q(a)}\frac{\partial q(a)}{\partial q(u,x)}$$
$$= -\sum_a \frac{\partial q(a)}{\partial q(u,x)}(1 + \ln q(a)).$$

For the particular term $H(U|Y)$ in $F(\alpha,q)$, the gradient is

$$\frac{\partial(H(U,Y) - H(Y))}{\partial q(u,x)} = -\sum_{u',y}\frac{\partial q(u',y)}{\partial q(u,x)}(1 + \ln q(u',y)) + \sum_y \frac{\partial q(y)}{\partial q(u,x)}(1 + \ln q(y))$$
$$= -\sum_y p(y|x)(1 + \ln q(u,y)) + \sum_y p(y|x)(1 + \ln q(y))$$
$$= -\sum_y p(y|x)\ln q(u|y).$$

Now, we can calculate the first-order term in Equation (41):

$$-(q_2 - q_1)^T \nabla H_{q_1}(U|Y) = \sum_{u,x,y}(q_2(u,x) - q_1(u,x))p(y|x)\ln q_1(u|y)$$
$$= \sum_{u,y}q_2(u,y)\ln q_1(u|y) + H_{q_1}(U|Y).$$

Finally, the left-hand side of Equation (41) equals

$$H_{q_2}(U|Y) - H_{q_1}(U|Y) - (q_2 - q_1)^T\nabla H_{q_1}(U|Y) = H_{q_2}(U|Y) + \sum_{u,y}q_2(u,y)\ln q_1(u|y)$$
$$= -D_{U|Y}(q_2||q_1).$$

For the other terms in Equation (16), we can perform similar calculations to obtain the desired inequality. □

Proof of Lemma 6. Consider the superlevel set $\tilde{\mathcal{S}}_F(\alpha,k)$ of the function $F(\alpha,q,Q[q_n])$. According to Equation (6), $F(\alpha,q,Q[q_n]) \leq F(\alpha,q)$ for all q; thus, $\tilde{\mathcal{S}}_F(\alpha,k) \subseteq \mathcal{S}_F(\alpha,k)$. From Equation (6), $F(\alpha,q_n,Q[q_n]) = F(\alpha,q_n)$, we have $q_n \in \tilde{\mathcal{S}}_F(\alpha,k)$; further, because $F(\alpha,q,Q[q_n])$ is concave in q, $\tilde{\mathcal{S}}_F(\alpha,k)$ is a convex set, and as such is connected. This implies that $\tilde{\mathcal{S}}_F(\alpha,k) \subseteq \mathcal{T}_F(\alpha,k,q_n)$. Because q_{n+1} makes the function value $F(\alpha,q,Q[q_n])$ larger than that of q_n, it must lie in $\tilde{\mathcal{S}}_F(\alpha,k)$, and as such in $\mathcal{T}_F(\alpha,k,q_n)$. □

Proof of Theorem 5. The proof is similar to that of Theorem 2. We perform the following manipulations:

$$F(\alpha, q_*) - F(\alpha, q_n, Q_n) - \bar{\theta} \sum_{u,x} q_*(u,x) \ln \frac{q_n(u,x)}{q_{n-1}(u,x)}$$

$$= F(\alpha, q_*, Q[q_*]) - F(\alpha, q_n, Q_n) - \bar{\theta} \sum_{u,x} q_*(u,x) \ln \frac{q_n(u,x)}{q_{n-1}(u,x)}$$

$$\overset{(a)}{=} \bar{\theta} \sum_{u,x} q_*(u,x) \Big(d[Q[q_*]] - \ln q_* - d[Q_n] + \ln q_n - \ln \frac{q_n}{q_{n-1}} \Big)$$

$$= \bar{\theta} \sum_{u,x} q_*(u,x) \Big(d[Q[q_*]] - \ln q_* - d[Q[q_{n-1}]] + \ln q_{n-1} \Big)$$

$$\overset{(b)}{=} \bar{\theta} \big(-D_{UX}(q_* \| q_{n-1}) + D_{X|YU} + \bar{\alpha} D_{U|Y} + \alpha D_{U|Z} \big) + (\bar{\theta} - \theta) D_{Y|U}$$

$$\overset{(c)}{\leq} 0,$$

where (a) holds from Equation (8), (b) is due to

$$\sum_{u,x} q_* \big(d[Q[q_*]] - d[Q[q_{n-1}]] \big)$$

$$= \sum_{u,x,y,z} q_*(u,x) p(y,z|x) \Big(\ln \frac{q_*(x|y,u)}{q_{n-1}(x|y,u)} \Big(\frac{q_*(u|y)}{q_{n-1}(u|y)} \Big)^{\bar{\alpha}} \Big(\frac{q_*(u|z)}{q_{n-1}(u|z)} \Big)^{\alpha}$$

$$+ \frac{\bar{\theta} - \theta}{\bar{\theta}} \ln \frac{q_*(y|u)}{q_{n-1}(y|u)} \Big)$$

$$= D_{X|YU}(q_* \| q_{n-1}) + \bar{\alpha} D_{U|Y} + \alpha D_{U|Z} + \frac{\bar{\theta} - \theta}{\bar{\theta}} D_{Y|U},$$

and (c) is from Lemma 5. This implies that

$$\sum_{n=1}^{N} \Big(F(\alpha, q_*) - F(\alpha, q_n, Q_n) \Big) \leq \bar{\theta} \sum_{u,x} q_*(u,x) \ln \frac{q_N(u,x)}{q_0(u,x)}$$

$$= \bar{\theta} \sum_{u,x} q_*(u,x) \ln \frac{q_N(u,x) q_*(u,x)}{q_0(u,x) q_*(u,x)}$$

$$= \bar{\theta} \big(D_{UX}(q_* \| q_0) - D_{UX}(q_* \| q_N) \big)$$

$$\leq \bar{\theta} D_{UX}(q_* \| q_0).$$

The last term is finite and positive, as $q_0 > 0$. Finally, a sequence of positive terms has a finite sum, which implies that the terms converge to zero, i.e., $F(\alpha, q_n, Q_n) \to F(\alpha, q_*)$. □

Proof of Proposition 2. Let C_m and $\Delta_d(u,x)$ be

$$C_m = \sum_{u,x} \exp\{d_m[\tilde{Q}_*](u,x)\},$$

$$\Delta_d(u,x) = d_{m+1}[\tilde{Q}_*] - d_m[\tilde{Q}_*],$$

where $d_k[Q]$ is as in Equation (18) with $\alpha = \alpha_k$. According to Equations (7) and (8),

$$q_0 = \tilde{q}_* = \frac{\exp\{d_m[\tilde{Q}_*]\}}{\sum \exp\{d_m[\tilde{Q}_*]\}} = \frac{\exp\{d_m[\tilde{Q}_*]\}}{C_m},$$

$$F(\alpha_m, \tilde{q}_*) = \bar{\theta} \ln C_m.$$

Noting that $Q_1 = Q[q_0] = Q[\tilde{q}_*] = \tilde{Q}_*$, we estimate the difference as follows:

$$
\begin{aligned}
F(\alpha_{m+1}, q_1, Q_1) - F(\alpha_m, \tilde{q}_*) &\overset{(a)}{=} \bar{\theta} \ln \sum \exp\{d_{m+1}[\tilde{Q}_*]\} - \bar{\theta} \ln C_m \\
&= \bar{\theta} \ln \sum \exp\{d_m[\tilde{Q}_*]\} \exp\{\Delta_d\} - \bar{\theta} \ln C_m \\
&= \bar{\theta} \ln \sum C_m \tilde{q}_* \exp\{\Delta_d\} - \bar{\theta} \ln C_m \\
&= \bar{\theta} \ln \sum \tilde{q}_* \exp\{\Delta_d\} \\
&\approx \bar{\theta} \ln \sum \tilde{q}_*(1 + \Delta_d) \\
&= \bar{\theta} \ln(1 + \sum \tilde{q}_* \Delta_d) \\
&\approx \bar{\theta} \sum \tilde{q}_* \Delta_d,
\end{aligned}
$$

where (a) holds from Equation (8) and from the fact that $Q_1 = \tilde{Q}_*$.

According to the definition of $d[Q]$ in Equation (18),

$$
\Delta_d = (\alpha_{m+1} - \alpha_m) \sum_{y,z} p(y,z|x) \ln \frac{\tilde{Q}_*(u|z)}{\tilde{Q}_*(u|y)}.
$$

The expectation of this difference is

$$
\begin{aligned}
\sum_{u,x} \tilde{q}_* \Delta_d &= (\alpha_{m+1} - \alpha_m)(I(U;Z) - I(U;Y)) \\
&\overset{(b)}{=} (\alpha_{m+1} - \alpha_m)(\alpha_{m+1} - \alpha_m)\frac{1}{-\tau_m \bar{\theta}} \\
&= -\frac{(\alpha_{m+1} - \alpha_m)^2}{\tau_m \bar{\theta}},
\end{aligned}
$$

where (b) is due to Equation (22). The proof is finished. $\square$

References

1. Cover, T. Broadcast Channels. *IEEE Trans. Inform. Theory* **1972**, *18*, 2–14. [CrossRef]
2. Bergmans, P. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inform. Theory* **1973**, *19*, 197–207. [CrossRef]
3. Körner, J.; Marton, K. Comparison of two noisy channels. In *Topics in Information Theory*; Csiszár, I., Elias, P., Eds.; North-Holland: Amsterdam, The Netherlands, 1977; pp. 411–423.
4. Marton, K. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inform. Theory* **1979**, *25*, 306–311. [CrossRef]
5. Nair, C.; El Gamal, A. An outer bound to the capacity region of the broadcast channel. *IEEE Trans. Inform. Theory* **2007**, *53*, 350–355. [CrossRef]
6. Gohari, A.; Nair, C. Outer bounds for multiuser settings: The auxiliary receiver approach. *IEEE Trans. Inform. Theory* **2022**, *68*, 701–736. [CrossRef]
7. Liu, Y.; Geng, Y. Blahut-Arimoto Algorithms for Computing Capacity Bounds of Broadcast Channels. In Proceedings of the IEEE International Symposium on Information Theory, Espoo, Finland, 26 June–1 July 2022; pp. 1145–1150.
8. Liu, Y.; Dou, Y.; Geng, Y. Blahut-Arimoto Algorithm for Marton's Inner Bound. In Proceedings of the IEEE International Symposium on Information Theory, Taipei, Taiwan, 25–30 June 2023; pp. 2159–2164.
9. Calvo, E.; Palomar, D.P.; Fonollosa, J.R.; Vidal, J. The computation of the capacity region of the discrete degraded BC is a nonconvex DC problem. In Proceedings of the IEEE International Symposium on Information Theory, Toronto, ON, Canada, 6–11 July 2008; pp. 1721–1725.
10. Nocedal, J.; Wright, S.J. *Numerical Optimization*, 2nd ed.; Springer: New York, NY, USA, 2006.
11. Wilson, R.B. A simplicial Algorithm for Concave Programming. Ph.D. Thesis, Graduate School of Business Administration, Harvard University, Cambridge, MA, USA, 1963.
12. Blahut, R. Computation of channel capacity and rate distortion functions. *IEEE Trans. Inform. Theory* **1972**, *18*, 460–473. [CrossRef]
13. Arimoto, S. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inform. Theory* **1972**, *18*, 14–20. [CrossRef]
14. Rezaeian, M.; Grant, A. Computation of Total Capacity for Discrete Memoryless Multiple-Access Channels. *IEEE Trans. Inform. Theory* **2004**, *50*, 2779–2784. [CrossRef]

15. Calvo, E.; Palomar, D.P.; Fonollosa, J.R.; Vidal, J. On the Computation of the Capacity Region of the Discrete MAC. *IEEE Trans. Commun.* **2010**, *58*, 3512–3525. [CrossRef]

16. Yasui, K.; Matsushima, T. Toward Computing the Capacity Region of Degraded Broadcast Channel. In Proceedings of the IEEE International Symposium on Information Theory, Austin, TX, USA, 13–18 June 2010; pp. 570–574.

17. Dupuis, F.; Yu, W.; Willems, F.M.J. Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information. In Proceedings of the IEEE International Symposium on Information Theory, Chicago, IL, USA, 27 June–2 July 2004; p. 179. Available online: https://www.comm.utoronto.ca/~weiyu/ab_isit04.pdf (accessed on 8 January 2024).

18. Gohari, A.; El Gamal, A.; Anantharam, V. On Marton's Inner Bound for the General Broadcast Channel. *IEEE Trans. Inform. Theory* **2014**, *60*, 3748–3762. [CrossRef]

19. Geng, Y.; Gohari, A.; Nair, C.; Yu, Y. On Marton's Inner Bound and Its Optimality for Classes of Product Broadcast Channels. *IEEE Trans. Inform. Theory* **2014**, *60*, 22–41. [CrossRef]

20. Anantharam, V.; Gohari, A.; Nair, C. On the Evaluation of Marton's Inner Bound for Two-Receiver Broadcast Channels. *IEEE Trans. Inform. Theory* **2019**, *65*, 1361–1371. [CrossRef]

21. Geng, Y. Single-Letterization of Supporting Hyperplanes to Outer Bounds for Broadcast Channels. In Proceedings of the IEEE/CIC International Conference on Communications in China, Changchun, China, 11–13 August 2019; pp. 70–74.

22. Gowtham, K.R.; Thangaraj, A. Computation of secrecy capacity for more-capable channel pairs. In Proceedings of the IEEE International Symposium on Information Theory, Toronto, ON, Canada, 6–11 July 2008; pp. 529–533.

23. Hajek, B.E.; Pursley, M.B. Evaluation of an achievable rate region for the broadcast channel. *IEEE Trans. Inform. Theory* **1979**, *25*, 36–46. [CrossRef]

24. Cover, T.M. An achievable rate region for the broadcast channel. *IEEE Trans. Inform. Theory* **1975**, *21*, 399–404. [CrossRef]

25. van der Meulen, E.C. Random coding theorems for the general discrete memoryless broadcast channel. *IEEE Trans. Inform. Theory* **1975**, *21*, 180–190. [CrossRef]

26. Nair, C.; Wang, Z.V. On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels. In Proceedings of the Information Theory and Applications Workshop, San Diego, CA, USA, 27 January–1 February 2008; pp. 226–229.

27. Gohari, A.; Anantharam, V. Evaluation of Marton's inner bound for the general broadcast channel. In Proceedings of the IEEE International Symposium on Information Theory, Seoul, Repubic of Korea, 28 June–3 July 2009; pp. 2462–2466.

28. Jog, V.; Nair, C. An information inequality for the BSSC channel. In Proceedings of the Information Theory and Applications Workshop, La Jolla, CA, USA, 31 January–5 February 2010; pp. 1–8.

29. El Gamal, A.; Kim, Y.-H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.