Abstract—The significance of the rate transfer argument used in a class of two-stage source and channel coding problems is discussed. A very simple necessary and sufficient condition is derived as to when the argument helps expand the achievable region. Special attention is paid to several example scenarios such as successive refinement with or without side information, and degraded broadcast channels. Finally, for reliable separate source-channel coding, whether a test based on comparing marginal source and channel rate regions is more restrictive than comparing cumulative rate regions, obtained after using the rate transfer argument, is discussed. It turns out that if rate transfer does not matter in either the source or the channel coding problem, the two tests are equivalent.

I. INTRODUCTION

In many two-stage source and channel coding scenarios, one usually derives sufficient conditions for achievability initially in terms of marginal rates, i.e., rates at each stage. Then, the simple, yet powerful, rate transfer argument is used to expand the achievability region to a possibly larger one expressed in terms of cumulative rates, i.e., total rates “seen” at each stage.

In source coding, the argument is that any subset of the bits in the second stage description can actually be moved to the first stage. The first stage receiver simply ignores those extra bits, and the second stage receiver still has access to them through the first stage description.

Similarly, in channel coding, the argument is that any subset of the bits decoded by both receivers can be reserved for the use of the second receiver only, as a part of its private message. The first receiver simply ignores those readily available bits, and outputs as the common message only the remaining bits.

Since cumulative rates arise more naturally in converse results, the rate transfer argument can prove quite useful in fully characterizing the achievable rate region. For example, in [5], the single-letter characterization was completed using the rate transfer argument for the successive refinement problem. As shown by an example in [9], this was a necessary step, i.e., one cannot characterize the achievable region in terms of marginal rates in general.

On the other hand, it is known through a converse result that marginal rates suffice to characterize the capacity region for degraded broadcast channels [1]. One then wonders whether there is a general rule to decide when the cumulative rate region obtained using the rate transfer argument is equivalent to the original marginal rate region. Equivalence of the marginal and cumulative rate regions is desirable as the former is simpler. This work was motivated by this question, which turns out to have an affirmative answer.

Even when marginal and cumulative rate regions are equivalent for the source (channel) coding problem, it is not clear whether the rate transfer argument matters when the code is to be combined with a channel (source) code for which the cumulative rate region is larger. For example, if a successive refinement source code is to be used with a degraded broadcast channel code, does it suffice to characterize compatibility in terms of marginal rates? One can always use cumulative rates to be on the safe side, as was done in [8] and [4] for a similar combination. However, expressions in terms of marginal rates is still preferable due to their simplicity. Fortunately, the answer to this question is also affirmative, as we show in the sequel.

II. PRELIMINARIES AND BACKGROUND

A. Two-Stage Source Coding Scenarios

For a general class of 2-stage source coding scenarios, we say that an achievable rate region is a marginal rate region if it has a single-letter characterization of the form

\[ \mathcal{R}_m = \{ (R_1, R_2) : \exists \mathbf{X} \in \mathcal{D} \text{ s.t. } I_1(\mathbf{X}) \leq R_1, I_2(\mathbf{X}) \leq R_2 \} \]

where \( \mathbf{X} \) is a random vector, \( \mathcal{D} \) is a region in the probability simplex of \( \mathbf{X} \), and \( I_1 \) and \( I_2 \) are information measures intrinsic to the problem. On the other hand, if the single-letter characterization is of the form

\[ \mathcal{R}_c = \{ (R_1, R_2) : \exists \mathbf{X} \in \mathcal{D} \text{ s.t. } I_1(\mathbf{X}) \leq R_1, \]
\[ I_1(\mathbf{X}) + I_2(\mathbf{X}) \leq R_1 + R_2 \}

we say that it is a cumulative rate region.

In proving achievability theorems, one usually first proves that the region \( \mathcal{R}_m \) is achievable, and then uses the rate transfer argument mentioned above to expand the achievability region to the potentially larger\(^1\) \( \mathcal{R}_c \). The rate argument is utilized as follows. If \( (R_1, R_2) \in \mathcal{R}_c \), i.e., there exists an \( \mathbf{X} \) satisfying

\[ I_1(\mathbf{X}) \leq R_1 \tag{1} \]
\[ I_1(\mathbf{X}) + I_2(\mathbf{X}) \leq R_1 + R_2 \tag{2} \]

\(^1\)Throughout the paper, whenever we compare \( \mathcal{R}_m \) and \( \mathcal{R}_c \) for a given scenario, it is implicit that they are expressed in terms of the same \( \mathbf{X}, \mathcal{D}, I_1, I_2 \).
then letting \( \Delta R = R_1 - I_1(X) \), we have
\[
I_1(X) = R_1 - \Delta R, \\
I_2(X) \leq R_2 + \Delta R,
\]
so that \( (R_1 - \Delta R, R_2 + \Delta R) \in \mathcal{R}_m \). This, in turn, implies achievability of \((R_1, R_2)\) by simply transferring \( \Delta R \) bits to the first stage. Finally, that \( \mathcal{R}_c \) could be a larger region than \( \mathcal{R}_m \) is illustrated in Figure 1.

**Example S1:** Consider the classical successive refinement problem, where a memoryless source \( X \) with distribution \( P_X^\cdot \) is to be coded in two stages with possibly different distortion measures \( d_1, d_2 \) and prescribed distortion levels \( (D_1, D_2) \). The single-letter achievable rate region given in [3], [5] is a cumulative rate region \( \mathcal{R}_c \) with \( X = (X_1, X_2) \),
\[
\mathcal{D} = \left\{ X : P_{X_1X_2}(x, \hat{x}_1, \hat{x}_2) = P^*_X(x)P_{X_1X_2|x}(\hat{x}_1|X)P_{X_2|x}(\hat{x}_2|x), \\
\quad \mathbb{E}\{d_1(X, \hat{X}_1)\} \leq D_1, \mathbb{E}\{d_2(X, \hat{X}_2)\} \leq D_2 \right\},
\]
and
\[
I_1(X) = I(X; \hat{X}_1), \\
I_2(X) = I(X; \hat{X}_2|\hat{X}_1).
\]

**Example S2:** Let a source \( X \) be coded in two stages so that it can be decoded losslessly at the first stage decoder with the help of side information \( Y_1 \), and at the second stage decoder with the help of \( Y_2 \). One can immediately derive as a converse result that \( H(X|Y_1) \leq R_1 \) and \( H(X|Y_2) \leq R_1 + R_2 \) must necessarily be satisfied for achievability. On the other hand, the Slepian-Wolf result [6] can be easily extended to this scenario by using nested binning. More specifically, assuming that \( H(X|Y_1) \leq H(X|Y_2) \), one can create a first layer of \( 2^{nH(X|Y_2)} \) bins, and inside each bin, create a second layer of \( 2^{n(H(X|Y_1) - H(X|Y_2))} \) bins, where \( n \) is the block length. Letting the first and second stage descriptions be the bin indices in the first and second layers, respectively, and then using the rate transfer argument, we obtain a cumulative rate region \( \mathcal{R}_c \) that matches the converse. In our notation, \( X = (X_1, Y_1, Y_2) \),
\[
\mathcal{D} = \left\{ X : \sum_{y_2} P_{XY_1Y_2}(x, y_1, y_2) = P^*_X(x, y_1), \\
\quad \sum_{y_1} P_{XY_1Y_2}(x, y_1, y_2) = P^*_Y(x, y_2) \right\},
\]
and
\[
I_1(X) = H(X|Y_1), \\
I_2(X) = H(X|Y_2) - H(X|Y_1),
\]
where \( P^*_X, P^*_Y \) denote the joint distributions of source and side information available at each decoder. In fact, since the measures \( I_1 \) and \( I_2 \) depend solely on \( P^*_X \), \( P^*_Y \), the set \( \mathcal{D} \) is degenerate in the sense that all \( X \in \mathcal{D} \) yield the same \( I_1(X) \) and \( I_2(X) \).

**Example S3:** Consider the Wyner-Ziv [11] extension of Example S1 introduced in [7], where the first and second stage decoders have access to side information \( Y_1 \) and \( Y_2 \), respectively, satisfying \( X - Y_2 - Y_1 \). One characterization given in [7] is a cumulative rate region \( \mathcal{R}_c \) with \( X = (X_1, Y_1, Y_2, Z_1, Z_2) \),
\[
\mathcal{D} = \left\{ X : P_{XY_1Y_2Z_1Z_2}(x, y_1, y_2, z_1, z_2) = P^*_X(x)P_{Y_1,Y_2|X}(y_1|X)P_{Z_1,X}(z_1|x)P_{Z_2|Z_1,X}(z_2|x), \\
\quad \exists \hat{X}_1 = f_1(Z_1, Y_1), \hat{X}_2 = f_2(Z_2, Y_2) \text{ s.t.} \\
\quad \mathbb{E}\{d_1(X, \hat{X}_1)\} \leq D_1, \mathbb{E}\{d_2(X, \hat{X}_2)\} \leq D_2 \right\},
\]
and
\[
I_1(X) = I(X; Z_1|Y_1), \\
I_2(X) = I(X; Z_2|Z_1, Y_2),
\]
where \( P^*_X \) is the joint distribution of the source and the side information. Here, the sizes of the alphabets \( Z_1 \) and \( Z_2 \) need not be larger than \( |X| + 2 \) and \( (|X| + 1)^2 \), respectively. An alternative characterization was also given in [7] as a marginal rate region. However, it involves three auxiliary random variables instead of two, and is thus more complicated.

**B. Two-Stage Channel Coding Scenarios**

Similar to source coding, in two-stage channel coding scenarios, we say that an achievable rate region is a marginal capacity region if it has a single-letter characterization of the form
\[
\mathcal{C}_m = \left\{ (R_1, R_2) : \exists U \in \mathcal{P} \text{ s.t.} J_1(U) \geq R_1, J_2(U) \geq R_2 \right\}
\]
and that it is a cumulative capacity region if the single-letter characterization is of the form
\[
\mathcal{C}_c = \left\{ (R_1, R_2) : \exists U \in \mathcal{P} \text{ s.t.} J_1(U) \geq R_1, \\
\quad J_1(U) + J_2(U) \geq R_1 + R_2 \right\}
\]
where \( \mathcal{P} \) is a region in the probability simplex of \( U \), and \( J_1 \) and \( J_2 \) are information measures. Once achievability of \( \mathcal{C}_m \) is proven, achievability of the potentially larger \( \mathcal{C}_c \) follows from the rate transfer argument as in (1)-(4) with the exception that inequalities change direction, and \( \Delta R \leq 0 \). That \( \mathcal{C}_c \) could be a larger region than \( \mathcal{C}_m \) is illustrated in Figure 2.

**Example C1:** Let a degraded broadcast channel with channel input \( U \) and outputs \( V_1 \) and \( V_2 \) satisfying \( U - V_2 - V_1 \), a single-letter cost function \( q \), and a cost level \( Q \) be given. The single-letter achievable region is given usually as a
A. When Does Rate Transfer Matter?

It is obvious that given a breakdown of the total cost $Q$ into individual costs, i.e., $Q = Q_1 + Q_2$, each channel must be used to its maximum capacity $C_i(Q_i)$. The tradeoff therefore stems from solely the choice of $Q_1$ and $Q_2$.

III. RESULTS

A. When Does Rate Transfer Matter?

Let us first define

$$R_{\text{min}} = \min_{(R_1, R_2) \in R_m} [R_1 + R_2]$$

and

$$C_{\text{max}} = \max_{(R_1, R_2) \in C_m} [R_1 + R_2]$$

for the generic source and channel coding problems, respectively.

The following is our main result.

**Theorem 1:** If $R_m$ is a convex region, then $R_m = R_c$ if and only if

$$(R_{\text{min}}, 0) \in R_m.$$  

(6)

**Remark 1:** The convexity requirement is not too restrictive, because even if $R_m$ is not convex, it can be convexified using one more random variable for time sharing.

**Remark 2:** Intuitively, due to the shape of the underlying rate regions shown in Figure 1, whether $R_m = R_c$ boils down to whether the slope of the boundary of $R_m$ exceeds $-1$ or not, and in turn, whether $(R_{\text{min}}, 0) \in R_m$ or not. See Figure 3 for a pictorial interpretation.

**Proof:** We begin with the “only if” part. Assume $(R_{\text{min}}, 0) \notin R_m$. Let $X^*$ achieve the minimum in (5) with

$$R^*_1 = I_1(X^*)$$

$$R^*_2 = I_2(X^*)$$

with $R_1^* + R_2^* = R_{\text{min}}$. Clearly, $(R_1^*, R_2^*) \in R_m$, and thus $(R_1^* + \Delta R, R_2^* - \Delta R) \in R_c$ for any $0 \leq \Delta R \leq R_{\text{min}}$. Choosing $R_{\text{min}} = R_2^*$ yields $(R_{\text{min}}, 0) \in R_c$, and therefore $R_m \neq R_c$.

Now the “if” part. Assume $(R_{\text{min}}, 0) \in R_m$. Let $(R_1, R_2) \in R_c$, implying the existence of $X$ satisfying (1) and (2), and hence (3) and (4), implying $(R_1 - \Delta R, R_2 - \Delta R) \in R_m$. Since $R_m$ is convex, this, in turn, means that for any $0 \leq \lambda \leq 1$,

$$\lambda(R_1 - \Delta R) + (1 - \lambda) R_{\text{min}}, \lambda(R_2 + \Delta R) \in R_m.$$  

(7)

Choosing

$$\lambda = \frac{R_2}{R_2 + \Delta R}$$

(7) becomes

$$\left( \frac{R_2(R_1 - \Delta R) + \Delta R \cdot R_{\text{min}}, R_2}{R_2 + \Delta R} \right) \in R_m.$$  

(8)

But since, by definition, $R_{\text{min}} \leq R_1 + R_2$, (8) implies

$$\left( \frac{R_2(R_1 - \Delta R) + \Delta R \cdot (R_1 + R_2)}{R_2 + \Delta R}, R_2 \right) \in R_m$$

or, simplifying,

$$(R_1, R_2) \in R_m.$$  

(9)

We then proved $R_c \subseteq R_m$, implying $R_c = R_m$. □

The following theorem can be proven using the same steps as in the proof of Theorem 1.

**Theorem 2:** If $C_m$ is a convex region, then $C_m = C_c$ if and only if

$$(0, C_{\text{max}}) \in C_m.$$  

(10)
We then have the following lemma.

**Lemma 2:** If \( \hat{X}_1 = \hat{X}_2 \), \( d_1 = d_2 = d \), and \( D_1 \geq D_2 \), it is always true that \( R_m = R_c \).

**Proof:** We first show that \( \hat{X}_1^* = \hat{X}_2^* \). Since that implies \( X - \hat{X}_1^* - \hat{X}_2^* \), the result follows from Lemma 1. Towards that end, first note that

\[
R_{\text{min}} \geq R(D_2)
\]

where \( R(D) \) is the rate-distortion function of the source \( X \). Let \( \hat{X}_2 \) achieve \( R(D_2) \). In the minimization problem \( (9) \), \( \hat{X}_1 = \hat{X}_2 \) is a feasible choice, because

\[
\mathbb{E}\{d_1(X, \hat{X}_1)\} = \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2 \leq D_1
\]

\[
\mathbb{E}\{d_2(X, \hat{X}_2)\} \leq \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2.
\]

This, in turn, shows that

\[
R(D_2) \leq R_{\text{min}} \leq I(X; \hat{X}_1, \hat{X}_2) = I(X; \hat{X}_2) = R(D_2).
\]

Since both inequalities above must be satisfied with equality, it follows that \( \hat{X}_1^* = \hat{X}_2^* = \hat{X}_2 \), finishing the proof.

**Example S2:** It can be intuitively seen that \( R_m \neq R_c \), because \( R_m \) consists of a single rectangular region as shown in Figure 1. This can be verified by Theorem 1. More specifically,

\[
R_{\text{min}} = H(X|Y_2)
\]

and \( (R_{\text{min}}, 0) \notin R_m \).

**Example S3:** We have

\[
R_{\text{min}} = \min \left\{ \frac{R_2}{R_1} \right\} \tilde{X}_1 = \tilde{X}_2 = \tilde{X}_2^*, \text{ as the minimum rate in (9) is in fact }
\]

\[
\mathbb{E}\{d_2(X, \hat{X}_2)\} = \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2,
\]

\[
\mathbb{E}\{d_2(X, \hat{X}_2)\} \leq \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2.
\]

This, in turn, shows that

\[
R(D_2) \leq R_{\text{min}} \leq I(X; \hat{X}_1, \hat{X}_2) = I(X; \hat{X}_2) = R(D_2).
\]

Since both inequalities above must be satisfied with equality, it follows that \( \hat{X}_1^* = \hat{X}_2^* = \hat{X}_2 \), finishing the proof.

**Example S2:** It can be intuitively seen that \( R_m \neq R_c \), because \( R_m \) consists of a single rectangular region as shown in Figure 1. This can be verified by Theorem 1. More specifically,

\[
R_{\text{min}} = H(X|Y_2)
\]

and \( (R_{\text{min}}, 0) \notin R_m \).

**Example S3:** We have

\[
R_{\text{min}} = \min \left\{ \frac{R_2}{R_1} \right\} \tilde{X}_1 = \tilde{X}_2 = \tilde{X}_2^*, \text{ as the minimum rate in (9) is in fact }
\]

\[
\mathbb{E}\{d_2(X, \hat{X}_2)\} = \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2,
\]

\[
\mathbb{E}\{d_2(X, \hat{X}_2)\} \leq \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2.
\]

This, in turn, shows that

\[
R(D_2) \leq R_{\text{min}} \leq I(X; \hat{X}_1, \hat{X}_2) = I(X; \hat{X}_2) = R(D_2).
\]

Since both inequalities above must be satisfied with equality, it follows that \( \hat{X}_1^* = \hat{X}_2^* = \hat{X}_2 \), finishing the proof.

**Example S2:** It can be intuitively seen that \( R_m \neq R_c \), because \( R_m \) consists of a single rectangular region as shown in Figure 1. This can be verified by Theorem 1. More specifically,

\[
R_{\text{min}} = H(X|Y_2)
\]

and \( (R_{\text{min}}, 0) \notin R_m \).

**Example S3:** We have

\[
R_{\text{min}} = \min \left\{ \frac{R_2}{R_1} \right\} \tilde{X}_1 = \tilde{X}_2 = \tilde{X}_2^*, \text{ as the minimum rate in (9) is in fact }
\]

\[
\mathbb{E}\{d_2(X, \hat{X}_2)\} = \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2,
\]

\[
\mathbb{E}\{d_2(X, \hat{X}_2)\} \leq \mathbb{E}\{d(X, \hat{X}_2)\} \leq D_2.
\]

This, in turn, shows that

\[
R(D_2) \leq R_{\text{min}} \leq I(X; \hat{X}_1, \hat{X}_2) = I(X; \hat{X}_2) = R(D_2).
\]
where $C_2(Q)$ is the capacity-cost function for the better channel. Denote by $\hat{U}$ the achiever of $C_2(Q)$ and by $(T^*, U^*)$ the achiever of $C_{\text{max}}$. From (13), it is easy to see that $C_{\text{max}} = C_2(Q)$ with $U^* = \hat{U}$ and any dummy $T^*$. Since

$$I(T^*; V_1) = 0 \quad I(U^*; V_2[T^*]) = I(\hat{U}; V_2) = C_{\text{max}},$$

or in other words, $(0, C_{\text{max}}) \in C_m, C_m = C_c$ follows from Theorem 2.

**Example C2:** As discussed above, the maximum total capacity is given as

$$C_{\text{max}} = \max_{Q_1, Q_2: Q_1 + Q_2 \leq Q} C_1(Q_1) + C_2(Q_2)$$

where $C_i(Q)$ is the capacity of the $i$th channel. Assuming without loss of generality that $C_i(0) = 0$, whether $(0, C_{\text{max}}) \in C_m$ then translates to whether $Q_1 = 0, Q_2 = Q$ achieves the maximum in (14).

Now observe that $C_{\text{max}}$ is the maximum achievable capacity when given power is to be allocated to two parallel channels. For example, when the channels are Gaussian and $q(U) = U^2$, it is well-known that the optimal power allocation problem has the water-filling solution (see, e.g., [1])

$$Q_i = (\nu - N_i)^+$$

where $N_i$ is the variance of the noise on channel $i$, and $\nu$ satisfies

$$(\nu - N_1)^+ + (\nu - N_2)^+ = Q.$$ But that means $Q_1 = 0, Q_2 = Q$ is not optimal for Gaussian channels unless $N_1 > N_2$ and $Q \leq N_1 - N_2$. Therefore, in general,

$C_m \neq C_c.$

**C. Combination of Source and Channel Codes**

We now consider combination of two-stage source and channel codes. We focus on the case with equal bandwidth, but the result derived here can be easily generalized to the case with bandwidth expansion or compression.

The question we tackle here is whether $R_c \cap C_m \neq \emptyset$ is a more restrictive compatibility test than $R_c \cap C_c \neq \emptyset$ for reliable separate source-channel coding. As it turns out, there is a very simple sufficient condition for the two tests to be equivalent.

**Theorem 3:** If either $R_c \cap C_m \neq \emptyset$ or $C_m = C_c$, then

$$R_c \cap C_m \neq \emptyset \iff R_c \cap C_c \neq \emptyset.$$  

**Proof:** We consider only the case $R_m = R_c$ because the proof follows using similar steps if $C_m = C_c$ instead. Since $R_m \cap C_m \neq \emptyset$ always implies $R_c \cap C_c \neq \emptyset$, we only need to prove the other direction, i.e., (considering also $R_m = R_c$) that

$$R_c \cap C_c \neq \emptyset \implies R_c \cap C_m \neq \emptyset.$$  

Let $(R_1, R_2) \in R_c \cap C_c$, implying in particular the existence of $U$ satisfying

$$R_1 \leq J_1(U)$$  

$$(R_1 + R_2) \leq J_1(U) + J_2(U).$$  

Now, if $R_2 \leq J_2(U)$, then $(R_1, R_2) \in C_m$ and the result follows. On the other hand if $R_2 > J_2(U)$, then let $\Delta R = R_2 - J_2(U)$ and rewrite (16) as

$$R_1 + \Delta R \leq J_1(U)$$

implying that

$$(R_1 + \Delta R, R_2 - \Delta R) \in C_m.$$ But it follows from the rate transfer argument that

$$(R_1 + \Delta R, R_2 - \Delta R) \in R_c$$

also. Thus, $R_c \cap C_m \neq \emptyset$.

This result implies in particular that if the encoded bits in any of the two-stage source coding problems discussed here is to be transmitted over degraded broadcast channels, $R_m \cap C_m \neq \emptyset$ is a necessary and sufficient condition for reliable communication. For example, in the combination of Wyner-Ziv successive refinement with degraded broadcast channels, the conditions

$$I(X; Z_1 | Y_1) \leq I(T; V_1)$$  

$$I(X; Z_2 | Z_1, Y_2) \leq I(U; V_2 | T)$$  

are necessary and sufficient. Oblivious to this result, [4] used

$$I(X; Z_1 | Y_1) + I(X; Z_2 | Z_1, Y_2) \leq I(T; V_1) + I(U; V_2 | T)$$

instead of (17).

Similarly, in the combination of Examples S1 and C2, the necessary and sufficient conditions can be written as

$$I(X; \hat{X}_1) \leq C_1(Q_1)$$  

$$I(X; \hat{X}_2 | \hat{X}_1) \leq C_2(Q_2)$$

whenever $\hat{X}_1 = \hat{X}_2, d_1 = d_2$, and $D_1 \geq D_2$.

**References**


