Source Coding Theory for Multiterminal Communication Systems with a Remote Source

Hiro Suzuki YAMAMOTO* and Kohji ITOH**, Regular Members

UDC 519.728.2

SUMMARY The source coding problems are studied on the Slepian-Wolf-type system with a remote source (Fig. 1) and the Wyner-Ziv-type system with a remote source (Fig. 4). For the former, inner and outer bounds are obtained on the admissible rate region to attain a prescribed distortion tolerance. For the latter, the rate-distortion function is derived. As examples, a Gaussian remote source and a binary remote source are analyzed.

1. Introduction

In most cases the source coding problem assumes that a source output can be directly encoded for transmission over a channel. In some practical situations, however, encoded noisy transmission may intervene between a source and an encoder. For example, in telephone networks, any sophisticated encoding and decoding equipments could not be located at terminals while central offices could be equipped with encoders and decoders of considerable complexity. Also, there are some cases where information to be transmitted is measured data corrupted by measurement errors or a source output is quantized for digital processing. The source whose output may be distorted prior to encoding is called remote source (1). The source coding problem relating to the communication system comprising a single encoder connected to a remote source, a single channel and a single decoder has been studied in the literature (1)–(9). On the other hand, a practical communication system is often multiterminal. So the source coding problems for certain multiterminal communication systems also have been studied by various authors (4)–(7).

In this paper we consider source coding problems for two types of multiterminal communication systems with a remote source. Section 2 deals with the system in which a single remote source is connected to two separate encoders, their outputs being supplied to a decoder via individual channels (Fig. 1). The system resembles the one due to Slepian and Wolf (5) differing in that the decoder estimates the output of the original source and not the inputs to the encoders. Inner and outer bounds are obtained on the region of admissible coding rates to attain a prescribed distortion tolerance in this system. The case of Gaussian remote source is analyzed as an example. In Section 3, the rate-distortion function is obtained for the Wyner-Ziv (6)–(9) type system consisting of one encoder connected to a remote source and one decoder with side information. This is a special case of the Slepian-Wolf-type system in Section 2. As examples, the rate-distortion functions for a binary and a Gaussian remote sources are determined.

2. Slepian-Wolf-Type System with a Remote Source

Let us consider first the system depicted in Fig. 1, where \(X_k\) and \(Y_{1k}, Y_{2k}\) are sequences of random variables representing the source output and the noisy channel outputs, respectively. Let the source and the noisy channel be memoryless: that is, let \((X_k, Y_{1k}, Y_{2k})\), \(k = 1, 2, \ldots\) be generated by repeated independent drawings of a triplet of random variables \((X, Y_1, Y_2)\) which take values in \(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2\), respectively. Encoder \(i\), \(i = 1, 2\) can receive only the sequence \(Y_{1k}\) and encodes it at respective rate, \(R_i\) bits per input symbol, for transmission to the common decoder, which in its turn emits a reproduction sequence \(\hat{X}_k\) corresponding to the source output sequence \(X_k\). \(\hat{X}_k\) takes values in \(\mathcal{X}\), \(\mathcal{Y}_1\), \(\mathcal{Y}_2\), and \(\mathcal{X}\) are either discrete sets, the reals or arbitrary measurable spaces.

Encoding and decoding are done in blocks of length \(n\), and the fidelity criterion is given by

\[ E = \frac{1}{n} \sum_{k=1}^{n} D(X_k, \hat{X}_k) \] (1)

Fig. 1  Slepian-Wolf-type system with a remote source.
where $D : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ is a given distortion function and $E$ denotes expectation. A code $(n, M_1, M_2, d)$ is defined by three mappings $F_{E_1}, F_{E_2}$ and $F_D$, which correspond to encoder 1, 2 and the decoder, respectively:

\begin{align}
F_{E_1} : \mathcal{Y}^n \to I_{M_1} & \quad (2a) \\
F_{E_2} : \mathcal{Y}^n \to I_{M_2} & \quad (2b) \\
F_D : I_{M_1} \times I_{M_2} \to \mathcal{X}^n & \quad (2c)
\end{align}

where $I_{M_1} = \{0, 1, \ldots, M_1 - 1\}$. The parameter $d$ is defined by

\[ D \triangleq \frac{E}{n} \sum_{k=1}^{n} D(X_k, \hat{X}_k) \quad (3) \]

where $\hat{X}^n = F_D(F_{E_1}(Y^n_1), F_{E_2}(Y^n_2))$, bold face letters representing vectors with $n$ components.

A pair $(R_1, R_2)$ is said to be $d$-admissible if, for any given $\epsilon > 0$ and $n$ sufficiently large, there exists a code $(n, M_1, M_2, d)$ such that

\[ M_1 \geq e^{n(R_1 + \epsilon)}, \quad i = 1, 2 \quad (4) \]

and

\[ A \leq d + \epsilon \quad (5) \]

We define $\mathcal{R}^*(d)$ as the set of $d$-admissible rate-pairs.

Our main problem is to determine the region $\mathcal{R}^*(d)$. However, this problem is very difficult to solve exactly and bounds on $\mathcal{R}^*(d)$ are derived in the following.

Let $\hat{Y}_1$ and $\hat{Y}_2$ be random variables distributed jointly with $(X, Y_1, Y_2)$ and take values in certain sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively. Define subset $\mathcal{R}(\hat{Y}_1, \hat{Y}_2)$ in two dimensional Euclidean space by

\[ \mathcal{R}(\hat{Y}_1, \hat{Y}_2) = \{ (R_1, R_2) \in \mathbb{R}^2 \mid R_1 \geq I(Y_1; \hat{Y}_1 | \hat{Y}_2), \]

\[ R_2 \geq I(Y_2; \hat{Y}_2 | \hat{Y}_1), \]

\[ R_1 + R_2 \geq I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2) \} \quad (6) \]

Let $\mathcal{R}(d)$ be the set of pairs of random variables $\hat{Y}_1, \hat{Y}_2$ that satisfy properties (i) and (ii) below, and let $\mathcal{R}(d)$ be the set of those which satisfy properties (ii) and (iii).

(i) $I(\hat{Y}_1 \mid X, Y_1, Y_2) = 0$, $(i, j) = (1, 2)$.

(ii) $I(Y_1; Y_2 | \hat{Y}_2) = 0$, $(i, j) = (1, 2), (2, 1)$.

(iii) There exists a function $f : \hat{Y}_1 \times \hat{Y}_2 \to \hat{X}$ such that

\[ E[D(X, f(\hat{Y}_1, \hat{Y}_2))] \leq d \quad (9) \]

Now define two regions, $\mathcal{R}^*(d)$ and $\mathcal{R}^*(d)$, as follows.

\[ \mathcal{R}^*(d) = \bigcup_{\hat{Y}_1, \hat{Y}_2 \in \mathcal{R}(d)} \mathcal{R}(\hat{Y}_1, \hat{Y}_2) \]

\[ \mathcal{R}^*(d) = \bigcup_{\hat{Y}_1, \hat{Y}_2 \in \mathcal{R}(d)} \mathcal{R}(\hat{Y}_1, \hat{Y}_2) \]

where $\bigcup$ denotes set closure. Then the following theorem holds.

[Theorem 1]
Assume that a source satisfies the following conditions\(\dagger\)

1. If $X, Y_1, Y_2$ and $\hat{X}$ are finite discrete sets, it holds that for $\forall x \in X$ and $\forall \hat{x} \in \hat{X}$

\[ D(x, \hat{x}) \leq \epsilon \]

2. Otherwise\(\ddagger\), it holds that for all $\hat{x} \in \hat{X}$

\[ E[D(X, \hat{x})] \leq \epsilon \]

Furthermore, for the random variable $\hat{X}$ satisfying $E[D(X, \hat{X})] \leq \epsilon$ and for arbitrary $\epsilon > 0$, there exists a finite subsets $\{\hat{x}_j\}_{j=1}^\infty \subset \hat{X}$ and a quantization mapping $f_Q : \hat{X} \to \{\hat{x}_j\}_{j=1}^\infty$ such that

\[ E[D(X, f_Q(\hat{X}))] \leq (1 + \epsilon) E[D(X, \hat{X})] \]

Then,

\[ \mathcal{R}^*(d) \subseteq \mathcal{R}(d) \subseteq \mathcal{R}(d) \]

Theorem 1 can be proved in the similar way as Houseworth\(\dagger\) or Tung\(\ddagger\) did. However, we omit it because it is too lengthy. (The proof is given in Ref. (12)).

As an example, let us consider a Gaussian remote source shown in Fig. 2, where $X, W_1, W_2$ are independent Gaussian random variables, and let the distortion measure be $D(x, \hat{x}) = (x - \hat{x})^2$. Even under these specifications, it has been impossible to delimit the set $\mathcal{D}(d)$. Accordingly, we make assumption here that $(Y_1, Y_2, \hat{Y}_1, \hat{Y}_2)$ are jointly Gaussian. $\mathcal{R}(d)$ obtained by this contraction of $\mathcal{D}(d)$ is an inner-bound anyhow, though it may be slightly narrower\(\ddagger\) than the exact one. Owing to the Gaussian assumption and the condition (i),

\[ \hat{Y}_i = Y_i + V_i, \quad i = 1, 2 \]

where $V_i$ is the Gaussian random variable $N(0, \sigma^2_i)$ independent with $X, W_1$ and $W_2$, and let the function

\(\dagger\) It is quite natural to assume that $D(x, \hat{x}) \geq 0$ and for each $x \in X$ there exists at least one $\hat{x}$ such that $D(x, \hat{x}) = 0.$

\(\ddagger\) As shown in Ref. (9), if $X$ and $\hat{X}$ are the reals, $D(x, \hat{x}) = |x - \hat{x}|^\gamma$, $\gamma > 0$ and $E[X]^\gamma < \infty$, then condition (2) is satisfied.

\(\ddagger\) There is an indication in Ref. (13) that the jointly Gaussian assumption is insufficient.
Fig. 2 Gaussian remote source.

The minimum mean square estimator of \( X \) given \( \hat{Y}_1 \) and \( \hat{Y}_2 \) is known to be

\[
\hat{X} = \frac{1}{2} \log(2 \pi e \sigma^2_p). \tag{17}
\]

By using this relation, we can show that the region \( \mathcal{R}^I(d) \) defined by Eq. (11) is represented as follows.

\[
\mathcal{R}^I(d) = \Big\{ (R_1, R_2) : R_1 \geq r_1, R_2 \geq r_2, R_1 + R_2 \leq r \Big\} \tag{18}
\]

\[
r_1 = \frac{1}{2} \log \frac{k}{\sigma^2_1(1 + \sigma^2_2 + \sigma^2_3)} \tag{19}
\]

\[
r_2 = \frac{1}{2} \log \frac{k}{\sigma^2_2(1 + \sigma^2_1 + \sigma^2_3)} \tag{20}
\]

\[
r = \frac{1}{2} \log \frac{k}{\sigma^2_1 \sigma^2_2} \tag{21}
\]

\[
(a_1^2 + b_1^2) + (a_2^2 + b_2^2) + (a_1^2 + b_1^2)(a_2^2 + b_2^2)
\]

\[
= \frac{(a_1^2 + b_1^2)(a_2^2 + b_2^2)}{d^2} \tag{22}
\]

where \( b_i, i = 1, 2 \) are parameters and the following normalizations are performed.

\[
d_n = \frac{d}{\sigma_X}, \quad a_i = \frac{a_i}{\sigma_X}, \quad i = 1, 2 \tag{23}
\]

are performed.

Figure 3 illustrates \( \mathcal{R}^I(d_n) \). For the purpose of comparison, also drawn in the figure by dotted lines are the regions for the case where the two encoders can receive both \( Y_1 \) and \( Y_2 \), and the broken lines indicate for the case where the independent encoders make time-shared use of their respective input. The former are outerbounds on \( \mathcal{R}^I(d_n) \)'s and the latter are innerbounds. The boundaries of \( \mathcal{R}^I(d_n) \)'s obtained are seen to lie between the boundaries of the corresponding bounds.

3. Wyner-Ziv-Type System with a Remote Source

If the decoder in Fig. 1 can receive the sequence \( \{ \hat{Y}_i \} \) directly or equivalently, in case of finite \( \mathcal{L}_i \), \( R_i \geq H(Y_i) \), the system reduces to the Wyner-Ziv-type system with a remote source, which is illustrated in Fig. 4.

Let us denote the rate-distortion function for this system by \( R^*_{ws}(d) \), and define \( R_{ws}(d) \) as follows.

\[
R^*_{ws}(d) = \inf_{\hat{Y}_2 \in \mathcal{R}^*_{ws}(d)} I(Y_2; \hat{Y}_2 | Y_1) \tag{24}
\]

where \( \mathcal{R}^*_{ws}(d) \) is the set of the random variables \( \hat{Y}_2 \) jointly distributed with \( (X, Y_1, \hat{Y}_2) \) satisfying the conditions.

\[
I(Y_2; X | Y_1) = 0 \tag{25}
\]

There exists a function \( f^*_{ws} : \hat{Y}_1 \times \hat{Y}_2 \to \hat{X} \) such that

\[
E[D(X, f^*_{ws}(\hat{Y}_1, \hat{Y}_2))] \leq \delta \tag{26}
\]

Then, the following theorem is obtained from Theorem 1 by noting that \( R^*_{ws}(d) = \inf_{R_2} R_2 \) and \( \mathcal{R}^*(d) = \mathcal{R}^*(d) \) if \( \hat{Y}_1 = Y_1 \).

[Theorem 2]

If a source satisfies conditions (1) and (2) in section 2, then

\[
R^*_{ws}(d) = R^*_{ws}(\hat{Y}_1 \times \hat{Y}_2) \tag{27}
\]

As the first example, consider again the Gaussian remote source, Fig. 2. By letting \( \delta^2 \to 0 \) in Eqs. (20) and (21), we obtain

\[
r_2 = \frac{1}{2} \log \frac{1}{\sigma_X^2(1 + \sigma_X^2)} \tag{28}
\]

\[
a_1^2 + a_2^2 + a_1^2( a_2^2 + b_2^2) = a_1^2 \frac{(a_1^2 + b_2^2)}{d_n} \tag{29}
\]

By eliminating \( a_1^2 \) from these equations, we obtain

\[
R^*_{ws}(d_n) = \left\{ \frac{1}{2} \log \frac{a_1^2}{a_1^2(1 + a_1^2)} \left(1 + \frac{1}{a_1^2} + \frac{1}{a_1^2} \right) \left( d_n - \frac{1}{1 + \frac{1}{a_1^2} + \frac{1}{a_1^2}} \right), \right\}
\]

\[
1 \leq d_n \leq \frac{1}{1 + \frac{1}{a_1^2} + \frac{1}{a_1^2}} \tag{30}
\]

† When \( \hat{Y}_1 = Y_1 \), both \( \mathcal{R}^I(d) \) and \( \mathcal{R}^I(d) \) equal to \( \mathcal{R}^*_{ws}(d) \).
In this example, even if the encoder could know \( Y_1 \), the rate-distortion function would not decrease just by the same reason as in the jointly Gaussian example for the Wyner-Ziv system (Section 3). The reason can be explained by means of Figs. 5(a) and (b) which represent the optimal test channels for each situations, respectively, where \( \alpha \) is an arbitrary constant and \( f' \) is the minimum mean square estimator of \( X \) given \( Y_1 \) and \( \hat{Y}_3 \). Two configurations can be transformed to each other and attain the same values of the mean square error and the conditional mutual information as follows.

\[
d = E[(X - \hat{X})^2] = \frac{1}{d_X} + \frac{c_1^2}{d_{W_1}} + \frac{c_2^2}{d_{W_2}} + \frac{c_3^2}{d_{V_2}}
\]

\[
I(Y_2; \hat{Y}_3 | Y_1) = I(Y_3; \hat{Y}_3 | Y_1) = \frac{1}{2} \log \frac{\sigma_{V_2}^2}{\sigma_{V_2}^2}
\]

As the second example, let us consider a binary remote source, Fig. 6, where \( x = \{ a_1 = Y_1, a_2 = x = \{ 0, 1 \} \).
and the noisy channel consists of two binary symmetric channels. The input probability is $Q_X(0)=Q_X(1)=\frac{1}{2}$ and the bit error probabilities are $P_1$ and $P_2$. Let the distortion measure be the Hamming metric.

$$D(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

Then, the following theorem is obtained.

[Theorem 3]

The rate-distortion function for the Wyner-Ziv-type system with the binary remote source is given as follows.

(i) Average distortion such that $d < \min(P_1, P_2)$ is not attainable.

(ii) For $d \geq \min(P_1, P_2)$,

$$R^*_{w_2}(d) = \begin{cases} g(d), & P_2 \leq d \leq P_1 \\ 0, & P_1 \leq d \end{cases}$$

and the following notations are employed.

$$x * y = x(1-y) + (1-x)y$$

$$\lambda(x) = -x \log x - (1-x) \log (1-x)$$

The proof is given in the appendix.

$R^*_{w_2}(d)$ versus $d$ for the Gaussian and binary remote sources are depicted in Figs. 7 and 8, respectively.

If $Y_2$ equals $X$ in Fig. 4, the system reduces to the Wyner-Ziv system. This situation is attained by letting $a_1^2 = 0$ in Eq. (30) or in $P_2 = 0$ in Eq. (34), and $R^*_{w_2}(d)$ is observed to coincide with the rate-distortion function for the Wyner-Ziv system. If, on the other hand, $Y_1$ is independent of $X$ and $Y_2$ in Fig. 4, the system reduces to the Shannon system with a remote source. This situation is achieved by letting $a_1^2 = \infty$ in Eq. (30) or $P_1 = \frac{1}{2}$ in Eq. (34), and it is easy to see that $R^*_{w_2}(d)$ becomes equal to the rate-distortion function for the Shannon system with a remote source.

4. Concluding Remarks

We have discussed the source coding problems for the Slepian-Wolf-type system with a remote source and the Wyner-Ziv-type system with a remote source. For the former system the upper and lower bounds were obtained on the admissible rate region, and for the latter system the rate-distortion function was derived. The result can be easily extended to a more general multiterminal system having several remote sources, encoders and decoders.

In some situations, inputs to the separate encoders, $Y_1$ and $Y_2$, may be outputs of correlated sources and, at the decoder, it is wanted to reproduce some
function of $Y_1$ and $Y_2$, $X \triangleq F(Y_1, Y_2)$, within a prescribed distortion tolerance. For example, $X = Y_1 - Y_2$ for continuous amplitude sources, or $X = Y_1 + Y_2$ (mod 2) for binary sources. It is easily noticed that theorems 1 and 2 hold for such situations.

Acknowledgement

The authors wish to thank Dr. T.S. Han for providing the literatures and for his helpful suggestion.

References


Appendix: Proof of Theorem 3

It is not difficult to see that average distortion $d$ less than $\min (P_1, P_2)$ cannot be attained even if we estimate $X$ from $Y_1$ and $Y_2$ according to the maximum a posteriori probability criterion. So that $d < \min (P_1, P_2)$ cannot be realized at the decoder.
no matter what code is used. On the other hand, the decoder can obviously 
estimate \( X \) with \( d = p_1 \) from \( Y_1 \) only. Thus \( R^{*} \left( d \right) = 0 \) for \( d \geq p_1 \).

The most important part of theorem 3, Eq. (34), can be proved in a similar way as in Ref. (8), Section II except the following points. In proving the inequality \( R^{*} \left( d \right) \leq g \left( d \right) \), let \( \hat{X} = f_{u_x} \left( y_1, y_2 \right) = y_2 \), and \( \hat{X} = f_{u_x} \left( y_1, y_2 \right) = y_1 \) instead of letting \( \hat{X} = f \left( Y, Z \right) = Z \) and \( \hat{X} = f \left( Y, Z \right) = Y \) at paragraphs a) and b) in Ref. (8), Section II, respectively.

In order to prove the inequality \( R^{*} \left( d \right) \geq g \left( d \right) \), define

\[
\mathcal{A} \triangleq \left\{ \hat{Y}_2 : f_{u_y} \left( 0, \hat{Y}_2 \right) = f_{u_y} \left( 1, \hat{Y}_2 \right) \right\}
\]

\[
\mathcal{A}' \triangleq \left\{ \hat{Y}_2 : f_{u_y} \left( 0, \hat{Y}_2 \right) \neq f_{u_y} \left( 1, \hat{Y}_2 \right) \right\}
\]

\[
\theta \triangleq p_x \left[ \hat{Y}_2 \in \mathcal{A} \right]
\]

\[
\lambda_{\hat{Y}_2} \triangleq \frac{p_x \left[ \hat{Y}_2 = \hat{Y}_2 \right]}{p_r \left[ \hat{Y}_2 \in \mathcal{A} \right]}
\]

\[
d_{\hat{Y}_2} \triangleq E \left[ D \left( X, \hat{X} \right) | \hat{Y}_2 = \hat{Y}_2 \right]
\]

\[
d' \triangleq \theta \sum_{\hat{Y}_2 \in \mathcal{A}} \lambda_{\hat{Y}_2} d_{\hat{Y}_2} + \left( 1 - \theta \right) p_1
\]

then we can show

\[
d' \leq d \quad \tag{A\cdot7}
\]

\[
I \left( Y_2 ; \hat{Y}_2 | Y_1 \right) \geq \theta \sum_{\hat{Y}_2 \in \mathcal{A}} \lambda_{\hat{Y}_2} \left[ H \left( Y_1 | \hat{Y}_2 = \hat{Y}_2 \right) - H \left( Y_2 | \hat{Y}_2 = \hat{Y}_2 \right) \right] \quad \tag{A\cdot8}
\]

in the similar way as in Ref. (8), Eqs. (39) and (40).

Furthermore, define

\[
\tau \left( \hat{Y}_2 \right) \triangleq f_{u_y} \left( 0, \hat{Y}_2 \right) = f_{u_y} \left( 1, \hat{Y}_2 \right) \quad \text{for} \quad \hat{Y}_2 \in \mathcal{A}
\]

\[
\alpha_{\hat{Y}_2} \triangleq p_x \left( Y_2 = \hat{Y}_2 | \hat{Y}_2 \right) \quad \text{for} \quad \hat{Y}_2 \in \mathcal{A}
\]

\[
G \left( u \right) \triangleq h \left( P_1 \ast P_2 \ast u \right) - h \left( u \right)
\]

then we obtain

\[
H \left( Y_2 | \hat{Y}_2 = \hat{Y}_2 \right) = \tau \left( \alpha_{\hat{Y}_2} \right)
\]

\[
H \left( Y_1 | \hat{Y}_2 = \hat{Y}_2 \right) = \tau \left( \alpha_{\hat{Y}_2} \ast P_1 \right)
\]

\[
d_{\hat{Y}_2} = \alpha_{\hat{Y}_2} \ast P_2
\]

\[
\text{From Eqs. (A\cdot3), (A\cdot4), (A\cdot8), (A\cdot11)\text{--}(A\cdot13)\text{ and convexity of } G \left( \cdot , \text{Lemma A} \right)\text{,}}
\]

\[
I \left( Y_2 ; \hat{Y}_2 | Y_1 \right) \geq \theta \sum_{\hat{Y}_2 \in \mathcal{A}} \lambda_{\hat{Y}_2} G \left( \alpha_{\hat{Y}_2} \right)
\]

\[
\geq \theta G \left( \sum_{\hat{Y}_2 \in \mathcal{A}} \lambda_{\hat{Y}_2} \alpha_{\hat{Y}_2} \right)
\]

\[
= \theta \left[ \lambda \left( P_1 \ast P_2 \ast \beta = \lambda \left( \beta \right) \right] \quad \tag{A\cdot15}
\]

where

\[
\beta \triangleq \sum_{\hat{Y}_2 \in \mathcal{A}} \lambda_{\hat{Y}_2} \alpha_{\hat{Y}_2}
\]

On the other hand, from Eqs. (A\cdot6) and (A\cdot4),

\[
d' = \theta \sum_{\hat{Y}_2 \in \mathcal{A}} \lambda_{\hat{Y}_2} \left( \alpha_{\hat{Y}_2} \ast P_2 \right) + \left( 1 - \theta \right) p_1
\]

\[
= \theta \left( \left( \sum_{\hat{Y}_2 \in \mathcal{A}} \lambda_{\hat{Y}_2} \alpha_{\hat{Y}_2} \right) \ast P_2 + \left( 1 - \theta \right) p_1 \right)
\]

\[
= \theta \left( \beta \ast P_2 \right) + \left( 1 - \theta \right) p_1
\]

\[
\text{where the second equality follows from Eq. (A\cdot20).}
\]

Thus, Eqs. (A\cdot15), (A\cdot17) and (34)\text{--}(36) yield

\[
I \left( Y_2 ; \hat{Y}_2 | Y_1 \right) \geq g \left( d' \right) \quad \tag{A\cdot18}
\]

Now, because of (A\cdot7) and since \( g \left( d \right) \) is nonincreasing in \( d \), we have

\[
I \left( Y_2 ; \hat{Y}_2 | Y_1 \right) \geq g \left( d \right) \quad \tag{A\cdot19}
\]

Note:

Let \( \sum_{k} a_k = 1 \). Then,

\[
\sum_{k} a_k \left( x_k \ast y \right)
\]

\[
= \sum_{k} a_k \left( x_k \left( 1 - y \right) + \left( 1 - x_k \right) y \right)
\]

\[
= \sum_{k} a_k x_k \left( 1 - y \right) + \sum_{k} a_k \left( 1 - x_k \right) y
\]

\[
= \left( \sum_{k} a_k x_k \right) \left( 1 - y \right) + \left( 1 - \sum_{k} a_k x_k \right) y
\]

\[
= \left( \sum_{k} a_k x_k \right) \ast y \quad \tag{A\cdot20}
\]