On Multiple Descriptions and Team Guessing

Rudolf Ahlswede

Abstract

Witsenhausen's hyperbola bound for the multiple description problem without excess rate in case of a binary source is not right for exact joint reproductions. However, this bound is tight for almost–exact joint reproductions (Theorem 1, conjectured by Witsenhausen). The proof is based on an *approximative* form of the team guessing lemma for *sequences* of random variables. (This result may be of interest also for team guessing.) The hyperbola bound is also tight for exact joint reproductions and arbitrarily small, but possible, excess rate (Theorem 2). The proof of this result uses our covering lemma.

1 The Problem of Multiple Descriptions

During the last years a strong interest has developed in a certain source–coding problem called the "problem of multiple descriptions". Since the origin of this problem and the motivations for its study have already been extensively described (see [1]–[9]), we begin immediately with the formal setup.

Let $(X_t)_{t=1}^{\infty}$ be a sequence of independent and identically distributed (i.i.d.) random variables (RV's) with values in a finite set $\mathcal{X}$, that is, a discrete memoryless source (DMS). We are given three finite reconstruction spaces, $\hat{\mathcal{X}}_0$, $\hat{\mathcal{X}}_1$, and $\hat{\mathcal{X}}_2$ together with associated per letter distortion measures

$$d_i : \mathcal{X} \times \hat{\mathcal{X}}_1 \to \mathbb{R}, \ \ i = 0, 1, 2. \tag{1.1}$$

For a function F defined on a product space $\mathcal{Y}^n$ we use the notation

$$\text{rate}(F) = \frac{1}{n} \log \|F\|, \ \|F\| = \text{ the cardinality of the range of } F. \tag{1.2}$$

The quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable, if for all large n description functions $f_i : \mathcal{X}^n \to \mathcal{F}_i$ $(i = 1, 2)$ and reconstruction functions $g_i : \mathcal{F}_i \to \hat{\mathcal{X}}_i^n$ $(i = 1, 2)$, $g_0 : \mathcal{F}_1 \times \mathcal{F}_2 \to \hat{\mathcal{X}}_0^n$ exist such that

a) $\quad \text{rate}(f_i) \leq R_i, \ i = 1, 2$

and for

$$\hat{X}_i^n = (\hat{X}_{i1}, \ldots, \hat{X}_{in}) = g_i\big(f_i(X^n)\big), \ i = 1, 2$$
$$\hat{X}_0^n = (\hat{X}_{01}, \ldots, \hat{X}_{0n}) = g_0\big(f_i(X^n), f_2(X^n)\big),$$

b) $\quad E \sum_{t=1}^{n} d_i(X_t, \hat{X}_{it}) \le D_i n, \ i = 0, 1, 2.$

2 Witsenhausen's Hyperbola Conjecture for a Binary Source

We now consider a seemingly special characterization problem, which alone has already received great attention in the literature. Let X be binary and uniform, that is, $H(X) = 1$, and let all of the distortion measures equal the Hamming distance. Consider the closure $\overline{\mathcal{D}}_\delta(\vec{R})$ of the cross section $\mathcal{D}_\delta(\vec{R})$ of $\mathcal{D}(\vec{R})$, which is defined by choosing $D_0 = \delta$.

We say that $\vec{R}$ has no excess rate at $D_0 = 0$ if $R_1 + R_2 = H(X) = 1$. Let $\mathcal{L}$ (for line segment) denote the set of those $\vec{R}$'s Witsenhausen has established in [1] the hyperbola bound

$$\overline{\mathcal{D}}_0(\vec{R}) \subset \mathcal{P} \text{ for } \vec{R} \in \mathcal{L} \tag{2.1}$$

where

$$\mathcal{P} = \left\{ (D_1, D_2) : \left(D_1 + \frac{1}{2}\right)\left(D_2 + \frac{1}{2}\right) \ge \frac{1}{2}; D_1, D_2 \ge 0 \right\}.$$

Instead of considering the case $D_0 = 0$, one can study the situation where D_0 is arbitrarily small but positive. Since for $\delta > \delta' \overline{\mathcal{D}}_\delta(\vec{R}) \supset \overline{\mathcal{D}}_{\delta'}(\vec{R})$, it is natural to define

$$\overline{\mathcal{D}}_+(\vec{R}) = \bigcap_{\delta > 0} \overline{\mathcal{D}}_\delta(\vec{R}). \tag{2.2}$$

Since this is the way in which quantities are usually defined in Shannon's rate–distortion theory, $\overline{\mathcal{D}}_+$ may be termed a *distortion–rate map* at $D_0 = 0$ in the "Shannon sense".

Witsenhausen conjectured that for

$$\overline{\mathcal{D}}_+(\mathcal{L}) = \bigcup_{\vec{R} \in \mathcal{L}} \overline{\mathcal{D}}_+(\vec{R})$$
$$\overline{\mathcal{D}}_+(\mathcal{L}) = \mathcal{P}. \tag{2.3}$$

Generally, one is interested in characterizing $\mathcal{Q}$, the set of achievable quintuples, or its closure $\overline{\mathcal{Q}}$. In particular, one is interested in rate–distortion regions and distortion–rate regions, which are the analogs to the following classical rate–distortion function and distortion–rate function.

$\mathcal{R}(\vec{D})$ denotes the set of rates $\vec{\mathcal{R}} = (R_1, R_2)$ achievable for distortion $\vec{D} = (D_0, D_1, D_2)$, and $\mathcal{D}(\vec{R})$ stands for the set of distortions $\vec{\mathcal{D}} = (D_0, D_1, D_2)$ achievable for rate $\vec{R} = (R_1, R_2)$. Often it is more convenient to work with their closures $\overline{R}(\vec{D})$ and $\overline{D}(\vec{R})$.

Several authors have made an effort to prove (2.3). In the special symmetric case $D_1 = D_2 = D$, one can calculate that

$$D' = \min\{D : (D, D) \in \mathcal{P}\} = 2^{-1}(\sqrt{2} - 1) \sim 0.207.$$

Wolf et al. [2] have shown that $D_{\min} = \min\{D : (D, D) \in \overline{\mathcal{D}}_+(\mathcal{L})\} \geq 6^{-1}$, and Witsenhausen and Wyner [5] have improved this result to $D_{\min} \geq 5^{-1}$. Finally, Berger and Zhang [7] proved the equality $D_{\min} = D'$.

Whereas the inequality $D_{\min} \leq D'$ readily follows from a general achievable region due to El Gamal and Cover [6], their proof for the opposite inequality is rather complicated. We completely settle Witsenhausen's conjecture with Theorem 1.

Theorem 1. *For X binary and uniform and the Hamming distance measures,*

$$\overline{\mathcal{D}}_+(\mathcal{L}) = \mathcal{P}.$$

Remarks:

1) A noticeable phenomenon about the result by Berger and Zhang is that at least in one point the hyperbola bound, which was derived for the case $D_0 = 0$, coincides with the true value in the Shannon case ($D_0 \to 0$). This motivated us in proving Witsenhausen's conjecture by continuity considerations, which let to the team–guessing Lemma 3, an improvement of the original team–guessing lemma.

2) We also studied continuity properties of $\overline{\mathcal{R}}(\vec{D})$ and $\overline{\mathcal{D}}(\vec{R})$ for general sources. The results are stated in the Appendix. A reader interested in these mathematical delicacies can find the proofs in [17].

It is important to notice that $\overline{\mathcal{D}}$ is *not* everywhere continuous; in particular, $\overline{\mathcal{D}}_\delta(\vec{R})$ is *not continuous* at $\delta = 0$. Even worse, $\overline{\mathcal{D}}_0(\vec{R})$ is not even convex. This led us to another zero–distortion problem, which we will define and whose solution we will present next.

3 A Zero–Distortion Problem

Instead of allowing arbitrary small distortion D_0, but not excess rate, one can consider the case of no distortion, but arbitrarily small excess rate. For X binary and uniform and the Hamming distortion measure we can thus consider the set

$$\lim_{\vec{\varepsilon} \to (0,0)} \overline{\mathcal{D}}_0(\mathcal{L} + \vec{\varepsilon}).$$

With the help of our covering lemma [11, part I] we show that a certain trade–off exists between the distortion and the rate to the extent that

$$\overline{\mathcal{D}}_+(\mathcal{L}) \subset \overline{\mathcal{D}}_0(\mathcal{L} + \vec{\varepsilon}) \text{ for every } \vec{\varepsilon} = (\varepsilon_1, \varepsilon_2) \text{ with } \varepsilon_i > 0, \; i = 1, 2. \tag{3.1}$$

We actually prove Theorem 2.

Theorem 2. *For X binary and uniform and the Hamming distortion measures,*

$$\lim_{\vec{\varepsilon} \to (0,0)} \overline{\mathcal{D}}_0(\mathcal{L} + \vec{\varepsilon}) = \mathcal{P}.$$

Next we state improved versions of the team–guessing lemma. They are also be of interest otherwise (see [3] and [4]).

4 On Team Guessing

The philosophy of team guessing is outlined in [3] and [4]. The result of [1] is the following lemma.

Team Guessing Lemma 1. *Let U, V, and W be 0–1–valued RV's defined on the same probability space. If*

a) *U and V are independent and $Pr(W = 0) = 1/2$,*

then

b) $\big(Pr(U \neq W), Pr(V \neq W)\big) \in \mathcal{P}$.

By an elementary continuity argument the following improvement is readily established. Define for any $\tau > 0$

$$\mathcal{P}(\tau) = \left\{ (D_1, D_2) : \left(D_1 + \frac{1}{2}\right)\left(D_2 + \frac{1}{2}\right) \geq \frac{1}{2}\tau ; D_1, D_2 \geq 0 \right\}. \tag{4.1}$$

Team Guessing Lemma 2. *A function $c : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ exists with $\lim_{r_1, r_2 \to 0} c(r_1, r_2) = 1$ such that for any 0–1–valued RV's U, V, and W with the properties*

a) *$I(U \wedge V) \leq \delta_1$ and $H(W) \geq 1 - \delta_2$*

we also have

b) $\big(Pr(U \neq W), Pr(V \neq W)\big) \in \mathcal{P}\big(c(\delta_1, \delta_2)\big)$.

Our main generalization is the next lemma.

Team Guessing Lemma 3. *A function $\omega : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ exists with $\lim_{r_1, r_2 \to 0} \omega(r_1, r_2) = 1$ such that for any positive integer n and any sequences $U^n = (U_1, \ldots, U_n)$, $V^n = (V_1, \ldots, V_n)$, and $W^n = (W_1, \ldots, W_n)$ of 0–1–valued RV's, the properties*

a) $\frac{1}{n}I(U^n \wedge V^n) \leq \varepsilon,\ \frac{1}{n}H(W^n) \geq 1-\eta$

imply

b) $\left(\frac{1}{n}\sum_{t=1}^n Pr(U_t \neq W_t), \frac{1}{n}\sum_{t=1}^n Pr(V_t \neq W_t)\right) \in \mathcal{P}\big(\omega(\varepsilon,\eta)\big)$.

The following proof makes use of what has been called in [15] a *wringing technique*. These techniques originated with [13] and were first made a powerful instrument for proving strong converses by Dueck [14]. (For a systematic analysis and improved forms see also [15].)

Wringing Lemma [14]. *Let $Y^n = (Y_1, \dots, Y_n)$ and $Z^n = (Z_1, \dots, Z_n)$ be RV's with values in $\mathcal{Y}^n$ resp. $\mathcal{Z}^n$. If $I(Y^n \wedge Z^n) \leq \sigma$, then for any $\delta > 0$, $t_1, \dots, t_k \in \{1, 2, \dots, n\}$, $k \leq \sigma/\delta$ exist such that*

$$I(Y_t \wedge Z_t | Y_{t_1}, Z_{t_1} \dots Y_{t_k} Z_{t_k}) \leq \delta,\ t = 1, 2, \dots, n.$$

This lemma says that conditional on a relatively small number of suitable component variables all corresponding component variables are much closer to independence than are Y^n and Z^n. The lemma was used by Dueck in [14] for strong converse proofs with σ, δ held constant, but it turns out to be good enough in the present "rate–distortion situation", where $\sigma = \varepsilon n$ (ε arbitrary small).

5 Proof of Team–Guessing Lemma 3

Application of the wringing lemma with $\sigma = \varepsilon n$ and $\delta = \varepsilon l$, $1 \leq l \leq n$, guarantees the existence of $t_1, \dots, t_k \in \{1, 2, \dots, n\}$ such that

$$I(U_t \wedge V_t | S) \leq \varepsilon l,\ t = 1, 2, \dots, n \tag{5.1}$$

where

$$S = U_{t_1} V_{t_1} \dots U_{t_k} V_{t_k},\ k \leq n l^{-1}. \tag{5.2}$$

Clearly, for the range $\mathcal{S}$ of S

$$|\mathcal{S}| = 2^{2k}. \tag{5.3}$$

The proof proceeds by showing that for a very large proportion of the triples of RV's $\big(U_t(s), V_t(s), W_t(s)\big)$ with joint distributions $P_{U_t V_t W_t | S=s}$ ($s \in \mathcal{S}; 1 \leq t \leq n$), Team–Guessing–Lemma 2 can be applied with sufficiently small δ_t. From here the final result is derived by using the convexity of $\mathcal{P}(\tau)$.

We now extract "good" components t and the "good" subsets of $\mathcal{S}$ for those t. As set of "good" components we define for $\gamma > 0$

$$\mathcal{N} = \mathcal{N}(n, \gamma) = \big\{t : H(W_t | S) \geq 1 - (1+\gamma)(\eta + 2l^{-1}), 1 \leq t \leq n\big\}. \tag{5.4}$$

Since

$$\sum_{t=1}^{n} H(W_t|S) \geq H(W^n|S)$$
$$\geq (1-\eta)n - 2k \geq (1-\eta-2l^{-1})n$$

and since $H(W_t|S) \leq 1$, we have

$$(1-\eta-2l^{-1})n \leq \sum_{t\in\mathcal{N}} H(W_t|S) + \sum_{t\in\{1,\dots,n\}\setminus\mathcal{N}} H(W_t|S)$$
$$\leq |\mathcal{N}| + \big(1-(1-\gamma)(\eta+2l^{-1})\big)(n-|\mathcal{N}|)$$

and, therefore,

$$|\mathcal{N}| \geq \gamma(1+\gamma)^{-1}n. \tag{5.5}$$

Define now for every $t \in \mathcal{N}$ the "good" subset

$$\begin{aligned}
\mathcal{S}_t &= \mathcal{S}'_t \cap \mathcal{S}''_t, \text{ where} \\
\mathcal{S}'_t &= \big\{s \in \mathcal{S} : I(U_t \wedge V_t|S=s) \leq \varepsilon l^2\big\} \\
\mathcal{S}''_t &= \big\{s \in \mathcal{S} : H(W_t|S=s) \geq 1-(1+\gamma)^2(\eta+2l^{-1})\big\}.
\end{aligned} \tag{5.6}$$

By (5.1) and the definition of $\mathcal{S}'_t$

$$\varepsilon l \geq I(U_t \wedge V_t|S) \geq \varepsilon l^2 \Pr(S \notin \mathcal{S}'_t)$$

and, therefore, $\Pr(S \in \mathcal{S}'_t) \geq 1 - l^{-1}$.

Similarly, by (5.4) and the definition of $\mathcal{S}''_t$

$$1\cdot\Pr(S\in\mathcal{S}''_t) + \big(1-(1+\gamma)^2(\eta+2l^{-1})\big)\cdot\Pr(S\notin\mathcal{S}''_t) \geq H(W_t|S) \geq 1-(1+\gamma)(\eta+2l^{-1})$$

and, therefore, $\Pr(S \in \mathcal{S}''_t) \geq 1-(1+\gamma)^{-1}$. The two inequalities imply

$$\Pr(S\in\mathcal{S}_t) \geq 1-l^{-1}-(1+\gamma)^{-1},\ t\in\mathcal{N}. \tag{5.7}$$

Application of Team Guessing Lemma 2 with the parameters $\delta_1 = \varepsilon l^2$ and $\delta_2 = (1+\gamma)^2(\eta+2l^{-1})$ yields

$$\left(\Pr\big(U_t(s)\neq W_t(s)\big)+\frac{1}{2}\right)\left(\Pr\big(V_t(s)\neq W_t(s)\big)+\frac{1}{2}\right) \geq \frac{1}{2}c(\delta_1,\delta_2),\ t\in\mathcal{N}, s\in\mathcal{S}_t. \tag{5.8}$$

The inequality $\Pr(U_t\neq W_t) \geq \sum_{s\in\mathcal{S}_t}\Pr(U_t\neq W_t|S=s)\Pr(S=s)$ and (5.7) imply

$$n^{-1}\sum_{t=1}^{n}\Pr(U_t\neq W_t) \geq n^{-1}\sum_{t\in\mathcal{N}}\Pr(U_t\neq W_t)$$
$$\geq n^{-1}\big(1-l^{-1}-(1+\gamma)^{-1}\big)|\mathcal{N}|\sum_{t\in\mathcal{N}}|\mathcal{N}|^{-1}$$
$$\cdot\sum_{s\in\mathcal{S}_t}\frac{\Pr(U_t\neq W_t|S=s)\Pr(S=s)}{\Pr(S\in\mathcal{S}_t)}.$$

Since $1 \geq |\mathcal{N}|n^{-1}\big(1 - l^{-1} - (1+\gamma)^{-1}\big)$, we also get

$$n^{-1}\sum_{t=1}^{n}\Pr(U_t \neq W_t) + \frac{1}{2} \geq |\mathcal{N}|n^{-1}\big(1 - l^{-1} - (1+\gamma)^{-1}\big)$$

$$\left[\sum_{t\in\mathcal{N}}|\mathcal{N}|^{-1}\cdot\sum_{s\in\mathcal{S}_t}\frac{\Pr(U_t \neq W_t|S=s)\Pr(S=s)}{\Pr(S\in\mathcal{S}_t)} + \frac{1}{2}\right], \tag{5.9}$$

and the same inequality holds with U_t replaced by V_t. The convexity of $\mathcal{P}\big(c(\delta_1\delta_2)\big)$, (5.8), and (5.9) imply the relation

$$\left(n^{-1}\sum_{t=1}^{n}\Pr(U_t \neq W_t), n^{-1}\sum_{t=1}^{n}\Pr(V_t \neq W_t)\right)$$
$$\in \big[|\mathcal{N}|n^{-1}\big(1 - l^{-1} - (1+\gamma)^{-1}\big)\big]^2 \cdot \mathcal{P}\big(c(\delta_1,\delta_2)\big). \tag{5.10}$$

Now we make an explicit choice of all of the parameters that are dependent upon ε and η, and then we verify that all of the demands can be met. Define

$$l = \varepsilon^{-2/5} \quad \gamma = \min(2^{-1/2}\varepsilon^{-1/10} - 1, \eta^{-2/5} - 1). \tag{5.11}$$

Then, clearly, $\delta_1 = \varepsilon l^2 = \varepsilon^{1/5}$ and

$$\begin{aligned}\delta_2 &= (2l^{-1} + \eta)(1+\gamma)^2 \\ &= (2\varepsilon^{2/5} + \eta)\min(2^{-1}\varepsilon^{-1/5}, \eta^{-4/5}) \\ &\leq \varepsilon^{1/5} + \eta^{1/5}.\end{aligned}$$

Finally, $\big[\gamma/(1+\gamma)\big(1 - \varepsilon^{2/5} - (1+\gamma)^{-1}\big)\big]^2 = \big(\gamma/(1+\gamma)\big)^2\big(\gamma/(1+\gamma) - \varepsilon^{2/5}\big)^2$, and since $\lim_{\varepsilon,\eta\to 0}\big(\gamma/(1+\gamma)\big) = 1$, the choice

$$\omega(\varepsilon,\eta) = \left(\frac{\gamma}{1+\gamma}\right)^2\left(\frac{\gamma}{1+\gamma} - \varepsilon^{2/5}\right)^2 c(\varepsilon^{1/5}, \varepsilon^{11/5} + \eta^{1/5}) \tag{5.12}$$

is suitable.

Remark: We use Lemma 3 only for $\eta = 0$. As a natural problem we suggest finding the exact regions of errors, to be guaranteed for all choices of RV's for every ε and η. Are these regions independent of n? The results may also be generalized to arbitrary RV's.

6 Proof of Theorem 1

Let $(D_1, D_2, 0, R_1, R_2) \in \overline{\mathcal{Q}}$ and $R_1 + R_2 = 1$; then for any $\alpha > 0$ and large enough n

$$f_i : \mathcal{X}^n \to \mathcal{F}_i \quad g_i : \mathcal{F}_i \to \mathcal{X}^n (i = 1, 2)$$

$$g_0 : \mathcal{F}_1 \times \mathcal{F}_2 \to \mathcal{X}^n$$

exist such that

$$\text{rate}(f_i) < R_i + \alpha, \tag{6.1}$$

and for

$$\hat{X}_i^n = g_i\big(f_i(X^n)\big)$$

and

$$\hat{X}_0^n = g_0\big(f_1(X^n), f_2(X^n)\big)$$

$$\frac{1}{n}\sum_{t=1}^{n} \Pr(\hat{X}_{it} \neq X_t) < D_i + \alpha,\ i = 1, 2 \tag{6.2}$$

and

$$\frac{1}{n}\sum_{t=1}^{n} \Pr(\hat{X}_{0t} \neq X_t) < \alpha. \tag{6.3}$$

We show first that (6.1) and (6.3) imply

$$\frac{1}{n} I(\hat{X}_1^n \wedge \hat{X}_2^n) \leq 2\alpha + h(\alpha). \tag{6.4}$$

For this, notice that by (6.3) and Fano's inequality

$$H(X^n | \hat{X}_0^n) \leq \sum_{t=1}^{n} H(X_t | \hat{X}_{0t}) \leq n h(\alpha),$$

and, therefore,

$$H\big(f_1(X^n), f_2(X^n)\big) \geq H(\hat{X}_0^n) \geq H(X^n) - H(X^n | \hat{X}_0^n) \geq n\big(1 - h(\alpha)\big).$$

Since also by (6.1)

$$H\big(f_1(X^n)\big) + H\big(f_2(X^n)\big) \leq n(1 + 2\alpha),$$

we conclude that

$$I\big(f_1(X^n) \wedge f_2(X^n)\big) \leq n\big(2\alpha + h(\alpha)\big)$$

and thus (6.4) by data processing.

We now apply Team Guessing Lemma 3 to the situation

$$U^n = \hat{X}_1^n \quad V^n = \hat{X}_2^n \quad W^n = X^n \quad \varepsilon = 2\alpha + h(\alpha) \quad \eta = 0,$$

and conclude that $\overline{\mathcal{D}}_+(\mathcal{L}) \subset \mathcal{P}$. The opposite implication follows by specialization of the El Gamal/Cover region [6]. Note that an argument has to be added in the calculations for this performed in [7]. It is given in the Appendix.

7 Proof of Theorem 2

Since in the proof of the converse part of Theorem 1 we allowed small excess rate, and since now (6.3) obviously holds, the same proof gives the converse of Theorem 2. Thus only (3.1) remains to be proven. This will be done by changing, with a small increase of rate, descriptions with a small average distortion D_0 to descriptions with $D_0 = 0$.

Actually, we proceed in two steps via Lemmas 1 and 2 to follow, which say how in classical rate–distortion theory coding functions can be modified in order to pass from an average distortion to a maximal distortion and finally to a zero distortion. The proof of Lemma 1 uses a special case of the following covering lemma.

Covering Lemma [11, part I].] *If for a hypergraph $(\mathcal{V}, \mathcal{E})$ $\min_{v \in \mathcal{V}} \deg(v) \geq d$, then a covering $\mathcal{C} \subset \mathcal{E}$ of $\mathcal{V}$ exists with*

$$|\mathcal{C}| \leq |\mathcal{E}| d^{-1} \log |\mathcal{V}| + 1.$$

An important special case is Covering Lemma 2 of [11, part II] which we called the link between channel and source coding. It also led to the notion of codes produced by permutations, etc.

Here we need another special case. We can always choose the alphabet $\mathcal{X}$ as $\{0, 1, \ldots, a-1\}$ and endow $\mathcal{X}$ with a group structure by adding numbers $\mod a$. We again denote this group by $\mathcal{X}$ and let $\mathcal{X}^n$ stand for the direct sum with n summands isomorphic to $\mathcal{X}$. Now, for every $A \subset \mathcal{X}^n$ we can define the hypergraph $\mathcal{H}_A = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{X}^n$ and $\mathcal{E}_A = \{A + x^n : x^n \in \mathcal{X}^n\}$.

Clearly, for every $x^n \in \mathcal{X}^n$

$$\deg(x^n) = |\{E \in \mathcal{E}_A : x^n \in E\}| = |A| \tag{7.1}$$

and, therefore, the covering lemma implies the following.

Covering Lemma 3. *For every $A \subset \mathcal{X}^n$, $u_1, \ldots, u_k \in \mathcal{X}^n$ exist with $\bigcup_{i=1}^n A + u_i = \mathcal{X}^n$, if $k > |A|^{-1} |\mathcal{X}|^n \log |\mathcal{X}|^n$.*

Henceforth, we assume $\hat{\mathcal{X}} = \mathcal{X}$. We call $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}_+$ accurate if for all $x, x' \in \mathcal{X}$,

$$d(x, x') = 0 \Leftrightarrow x = x', \tag{7.2}$$

and we call d *translation invariant* if for all $x, x', x'' \in \mathcal{X}$,

$$d(x + x'', x' + x'') = d(x, x'). \tag{7.3}$$

Examples of this include the Hamming distance, in particular in our case $\mathcal{X} = \{0, 1\}$, and the Lee distance.

A. From Average to Maximal Distortion

Lemma 1. *Let X be uniform, that is, $Pr(X=x)=|\mathcal{X}|^{-1}$ for $x\in\mathcal{X}$, and let d be translation invariant. Further, let γ be a positive number. Suppose now that for $f:\mathcal{X}^n\to\mathcal{F}$ and $g:\mathcal{F}\to\hat{\mathcal{X}}^n$*

$$|\mathcal{X}|^{-n}\sum_{x^n\in\mathcal{X}^n} d\big(x^n, g\big(f(x^n)\big)\big)\le Dn, \tag{7.4}$$

functions $f^*:\mathcal{X}^n\to\mathcal{F}^*\supset\mathcal{F}$ and $g^*:\mathcal{F}^*\to\hat{\mathcal{X}}^n$ exist such that

1) $f^*\equiv f$
   on $A_\gamma=\big\{x^n: d\big(x^n, g\big(f(x^n)\big)\big)\le(1+\gamma)Dn\big\}$
2) $g^*\equiv g$ *on* $f(A_\gamma)$
3) $d\big(x^n, g^*\big(f^*(x^n)\big)\big)\le(1+\gamma)Dn$, *for all* $x^n\in\mathcal{X}^n$
4) *rate*$(f^*)\le$ *rate*$(f)+\frac{1}{n}\Big[\log n+\frac{1}{\gamma}+\log^{(2)}|\mathcal{X}|\Big]$.

Proof. By Chebyshev's inequality $|\mathcal{X}|^{-n}|\mathcal{X}^n-A_\gamma|(1+\gamma)D\le D$ and, therefore,

$$|A_\gamma|\ge\gamma(1+\gamma)^{-1}|\mathcal{X}|^n. \tag{7.5}$$

Now apply Covering Lemma 3 to A_γ. Thus for

$$k=\big[\big(\gamma^{-1}(1+\gamma)\log|\mathcal{X}|\big)n\big], \tag{7.6}$$

$u_1,\dots,u_k\in\mathcal{X}^n$ exist with

$$\bigcup_{i=1}^{k} A_\gamma+u_i=\mathcal{X}^n. \tag{7.7}$$

Obviously, this can be achieved with $u_1=(0,\dots,0)$; otherwise, just subtract u_1 from all u_i. From the covering we pass to a partition $\{A_i:1\le i\le k\}$, where $A_1=A_\gamma$ and for $i>1$

$$A_i=(A_\gamma+u_i)\setminus\bigcup_{i'<i}A_{i'}. \tag{7.8}$$

Now define $f^*:\mathcal{X}^n\to\mathcal{F}^*=\mathcal{F}\cup\mathcal{F}\times\{2,\dots,k\}$ by

$$f^*(x^n)=\begin{cases} f(x^n), & x^n\in A_1,\\ \big(f(x^n-u_i),i\big), & x^n\in A_i,\ i\ge 2\end{cases} \tag{7.9}$$

and $g^* : \mathcal{F}^* \to \hat{\mathcal{X}}^n$ by

$$g^*(l) = \begin{cases} g(l), & l \in \mathcal{F} \\ g(m) + u_i, & l = (m,i) \in \mathcal{F} \times \{i\},\ i \geq 2. \end{cases} \tag{7.10}$$

Clearly, (1) and (2) are met. To verify (3), just observe that by the translation invariance of d for $x^n \in A_i$

$$\begin{aligned} d\big(x^n, g^*\big(f^*(x^n)\big)\big) &= d\big(x^n, g\big(f(x^n - u_i)\big) + u_i\big) \\ &= d\big(x^n - u_i, g\big(f(x^n - u_i)\big)\big), \end{aligned}$$

and since $x^n - u_i \in A_\gamma$

$$d\big(x^n, g^*\big(f^*(x^n)\big)\big) \leq (1+\gamma)Dn.$$

Finally, $\mathrm{rate}(f^*) \leq \mathrm{rate}(f) + (1/n)\log k \leq \mathrm{rate}(f) + (1/n)\log n + (1/\gamma) + \log^{(2)}|\mathcal{X}|$, and thus (4) follows.

B. From Small Maximal to Zero Distortion

Define $\mathbf{d} = \min\big\{d(x,\hat{x}) : (x,\hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}} \text{ with } d(x,\hat{x}) > 0\big\}$.

Lemma 2. *Let d be accurate and $0 < D \leq \mathbf{d}2^{-1}$. For $f : \mathcal{X}^n \to \mathcal{F}$, $g : \mathcal{F} \to \hat{\mathcal{X}}^n$ with $d\big(x^n, g\big(f(x^n)\big)\big) \leq Dn$ for all $x^n \in \mathcal{X}^n$, $f^0 : \mathcal{X}^n \to \mathcal{F}^0$, $G : \mathcal{F} \times \mathcal{F}^0 \to \hat{\mathcal{X}}^n$ exist such that for $F = (f, f^0)$,*

1) $d\big(x^n, G\big(F(x^n)\big)\big) = 0$ *for all* $x^n \in \mathcal{X}^n$
2) *rate*$(f^0) \leq h(D\mathbf{d}^{-1}) + D\mathbf{d}^{-1}\log(a-1)$, $a = |\mathcal{X}|$.

Proof. Clearly, for any $\hat{x}^n \in \hat{\mathcal{X}}^n$ and any $x^n \in S(\hat{x}^n) = \big\{x^n : g\big(f(x^n)\big) = \hat{x}^n\big\}$,

$$Dn \geq d(x^n, \hat{x}^n) = \sum_{t : x_t \neq \hat{x}_t} d(x_t, \hat{x}_t) \geq |\{t : x_t \neq \hat{x}_t\}|\mathbf{d}.$$

Therefore,

$$S(\hat{x}^n) \subset \big\{x^n : |\{t : x_t \neq \hat{x}_t\}| \leq D\mathbf{d}^{-1}n\big\} \tag{7.11}$$

and

$$\begin{aligned} |S(\hat{x}^n)| &\leq \big|\big\{x^n : |\{t : x_t \neq \hat{x}_t\}| \leq D\mathbf{d}^{-1}n\big\}\big| \\ &\leq \sum_{s=0}^{D\mathbf{d}^{-1}} \binom{n}{s} (a-1) D\mathbf{d}^{-1}n \\ &< \exp\big[\big(h(D\mathbf{d}^{-1}) + D\mathbf{d}^{-1}\log(a-1)\big)n\big]. \end{aligned}$$

Since the sets $S(\hat{x}^n)$, $\hat{x}^n \in \hat{\mathcal{X}}^n$, are disjoint, a function $f^0 : \mathcal{X}^n \to \mathcal{F}^0$ exists whose restrictions to these sets are injective, satisfying

$$\text{rate}(f^0) \leq h(D\mathbf{d}^{-1}) + D\mathbf{d}^{-1}\log(a-1). \tag{7.12}$$

Now, obviously, $F = (f, f^0) : \mathcal{X}^n \to \mathcal{F} \times \mathcal{F}^0$ is injective and a $G : \mathcal{F} \times \mathcal{F}^0 \to \mathcal{X}^n$ esists such that GF is the identity map on $\mathcal{X}^n$.

C. Proof of the Direct Part of Theorem 2

Because by Theorem 1 a pair (D_1, D_2) in $\mathcal{P}$ is achievable with an arbitrarily small average distortion D_0 by suitable descriptions f_1, f_2 and reproductions g_2, g_2, g_0, because the Hamming distance is accurate and translation invariant, and because our X is uniform, we can apply Lemmas 1 and 2 to $(f, g) = \big((f_1, f_2), g_0\big)$. Since D_0 can be made arbitrarily small, the additional rates to be transmitted to any one (or both) decoders can be kept arbitrarily small.

Acknowledgement

The author wishes to thank T. Berger and Z. Zhang for a preprint of their paper [7] and for informing him on the occasion of the Kiefer/Wolfowitz Memorial Conference about the existing literature on multiple descriptions.

Appendix I

Continuity Properties

For a distortion measure d we set

$$\begin{aligned} \overline{d} &= \max\big\{d(x, \hat{x}) : (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}\big\}, \\ \underline{d} &= \min\big\{d(x, \hat{x}) : (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}\big\}. \end{aligned} \tag{1}$$

Obviously, it suffices to consider only those $\vec{D} = (D_0, D_1, D_2)$ which are in

$$\Delta = \prod_{i=0}^{2} [\underline{d}_i, \overline{d}_i] \subset \mathbb{R}^3, \tag{2}$$

and only those $\vec{R} = (R_1, R_2)$ which are in

$$\Sigma = [0, \log |\mathcal{X}|]^2. \tag{3}$$

Generally, one is interested in characterizing $\mathcal{Q}$, the set of achievable quintuple, or its closure $\overline{\mathcal{Q}}$. For this, one first studies certain cross sections of $\mathcal{Q}$ and $\overline{\mathcal{Q}}$. In particular, one is interested in rate–distortion regions and distortion–rate regions, which are the analoga to the classical rate–distortion function and distortion–rate function.

$\mathcal{R}(\vec{D})$ shall denote the set of rates $\vec{R} = (R_1, R_2) \in \Sigma$ achievable for distortion $\vec{D} = (D_0, D_1, D_2) \in \Delta$, and $\mathcal{D}(\vec{R})$ stands for the set of distortion $\vec{D} = (D_0, D_1, D_2) \in \Delta$ achievable for rate $\vec{R} = (R_1, R_2) \in \Sigma$. Often it is more convenient to work with their closures $\overline{\mathcal{R}}(\vec{D})$ and $\overline{\mathcal{D}}(\vec{R})$ in the Euclidean topologies in Σ resp. Δ. Since these regions are also bounded, they are compact.

We investigate continuity properties of the maps

$$\overline{\mathcal{D}} : \Delta \to \text{comp}(\Sigma), \text{ the set of compact subsets of } \Sigma \tag{4}$$

and

$$\overline{\mathcal{D}} : \Sigma \to \text{comp}(\Delta). \tag{5}$$

Here the appropriate topologies for our purposes are the Euclidean topologies in the domains and the Hausdorff topologies in the ranges. Recall that the Hausdorff distance ρ between compact sets in metric spaces is given by

$$\rho(A, B) \triangleq \max\left(\max_{a\in A}\min_{b\in B} \text{dist}(a, b), \max_{b\in B}\min_{a\in A} \text{dist}(a, b)\right), \tag{6}$$

where dist denotes the Euclidean distance. We also consider the projections of Δ

$$\Delta_I \triangleq \prod_{i\in\{0,1,2\}\setminus I} [\underline{d}_i, \overline{d}_i] \prod_{i\in I} [d_i] \text{ for } I \subset \{0, 1, 2\} \tag{7}$$

endowed again with their Euclidean topologies and the restrictions $\overline{\mathcal{R}}_I$ of $\overline{\mathcal{R}}$ to Δ_I.

A. Continuity Properties of $\overline{\mathcal{R}}$

Some simple results for general sources and distortion measures are readily established.

Proposition 1. $\mathcal{R}_I$ is continuous in $\text{int}(\Delta_I)$, the interior of Δ_I, for all $I \subset \{0, 1, 2\}$. In particular, $\overline{\mathcal{R}}$ is continuous in $\text{int}(\Delta)$. Thus we are left with the study of continuity properties on the *boundary* of Δ resp. Δ_I.

Henceforth, $bd(A)$ stands for the boundary of a set A. Already in classical rate–distortion theory the rate–distortion function $\overline{R}(D)$ is generally discontinuous at $D = 0$. If we choose, for instance, the Hamming distance for d, then for a source with generic vairable X and $P_X(x) > 0$ for all $x \in \mathcal{X}$ we have

$$\lim_{D\to 0} \overline{R}(D) = H(X) \quad \overline{R}(0) = \log|\mathcal{X}|, \tag{8}$$

and, therefore, Proposition 2 follows.

Proposition 2. For the Hamming distortion measure, $\overline{R}$ is continuous at $D = 0$ iff X is uniform.

This obvious fact extends to multiple descriptions.

Continuity Theorem. *Let $d_i (i = 0, 1, 2)$ be accurate and translation invariant, and let X be uniform. Then $\overline{\mathcal{R}}$ is continuous everywhere in Δ, in particular, also on $bd(\Delta)$.*

B. Continuity Properties of $\overline{\mathcal{D}}$

Proposition 3. a) $\overline{\mathcal{D}}$ is continuous in $\mathrm{int}(\Sigma)$ and b) $\overline{\mathcal{D}}(0, \cdot)$ and $\overline{\mathcal{D}}(\cdot, 0)$ are continuous in $(0, \log |\mathcal{X}|)$.

However, $\overline{\mathcal{D}}$ is *not* continuous on $bd(\Sigma)$ *even under the assumptions of Theorem 1*, because $\overline{\mathcal{D}}_\delta(\vec{R})$ already behaves rather pathologically at $\delta = 0$ for $\vec{R} \in bd(\Sigma)$.

Example: Let X be binary and uniform. Since for $\vec{R} \in \mathcal{L}$ 2^{nR_1} and 2^{nR_2} are integral (and thus realizable by a code) for all large n only if $R_1 = 0$ or 1, we have

$$\overline{\mathcal{D}}_0(\mathcal{L}) = \bigcup_{\vec{R} \in \left\{(0,1),(1,0)\right\}} \overline{\mathcal{D}}_0(\vec{R}) = \left\{ \left(0, \frac{1}{2}\right), \left(\frac{1}{2}, 0\right) \right\}. \tag{9}$$

Thus this set is not even convex because time sharing fails. On the other hand, by Theorem 2

$$\lim_{\vec{\varepsilon} \to (0,0)} \overline{\mathcal{D}}_0(\mathcal{L} + \vec{\varepsilon}) = \mathcal{P}.$$

Therefore, the map $\overline{\mathcal{D}}_0$ (and a *fortiori* also $\overline{\mathcal{D}}$) is *discontinuous* in almost all points of $\mathcal{L}$, and $\overline{\mathcal{D}}_0(\mathcal{L})$ is not a reasonable notion of a "zero–distortion region".

C. On Zero–Distortion Problems

It seems that the region characterized in Theorem 2 is a canonical notion for a zero–distortion region. What other notions of zero–distortion regions without excess rate exist? Out of curiosity, we present two which are obtained by modifying the definition of achievability.

Notion 1 — $\mathcal{E}_0(\mathcal{L})$: Notice that for $(R_1, R_2) = (1/2, 1/2)$ $2^{R_i n}$ is integral and thus realizable by a code only for n even and not for all large n. This suggests a weaker concept of achievability, as follows.

The quintuple $(0, D_1, D_2, R_1, R_2)$ is *occasionally achievable*, if a sequence of codes $(f_1^n, f_2^n, \dots)_{j=1}^\infty$ exists with rate $(f_t^{n_j}) \leq R_i$ $(i = 1, 2)$ for $j = 1, 2, \dots,$ and distortions not exceeding $(0, D_1, D_2)$. This leads to the definitions $\mathcal{E}_0(\vec{R}) = \big\{(D_1, D_2) : (0, D_1, D_2, R_1, R_2)$ occasionally achievable$\big\}$ and $\mathcal{E}_0(\mathcal{L}) = \bigcup_{\vec{R} \in \mathcal{L}} \mathcal{E}_0(\vec{R})$.

Here again, it can happen that $\mathcal{E}_0(\vec{R}) = \varnothing$ for certain $\vec{R} \in \mathcal{L}$. For example, $\vec{R} = (\pi^{-1}, 1 - \pi^{-1})$ has this property because $2^{n\pi^{-1}}$ is irrational for all n.

Notion 2 — $\mathcal{F}_0(\mathcal{L})$: We call $(0, D_1, D_2)$ $\mathcal{L}$–achievable, if for all large n codes exist with distortions not exceeding $(0, D_1, D_2)$, and rate $\vec{R}(n) = \big(R_1(n), R_2(n)\big) \in \mathcal{L}$. This leads to the notion that $\mathcal{F}_0(\mathcal{L}) = \big\{(D_1, D_2) : (0, D_1, D_2 1) \text{ is } \mathcal{L}\text{–achievable}\big\}$. (Of course, one cound also define occasional $\mathcal{L}$–achievability and get a still different notion *a priori*).

$\mathcal{F}_0(\mathcal{L})$ formalizes Witsenhausen's concept of a zero–distortion region. It is mentioned in [1] without proof that $\mathcal{F}_0(\mathcal{L})$ is smaller that $\mathcal{P}$. We expect that $\overline{\mathcal{E}}_0(\mathcal{L}) = \overline{\mathcal{F}}_0(\mathcal{L})$. Can $\overline{\mathcal{E}}_0(\mathcal{L})$ or $\overline{\mathcal{F}}_0(\mathcal{L})$ be characterized?

Appendix II

A Missing Step in [7]

We complete here the calculations of [7] for the relation $\overline{\mathcal{D}}_+(\mathcal{L}) \supset \mathcal{P}$. For zero–one–valued RV's U, V, and X with $I(U \wedge V) = 0$ and $H(X) = 1$, it has to be shown that for $D_1 = \Pr(U \neq X)$, $D_2 = \Pr(V \neq X)$, the inequality

$$h(D_1) + h(D_2) \geq 1 \quad 0 \leq D_1 \quad D_2 \leq 1/2 \tag{10}$$

holds. This is a consequence of the following simple proposition.

Proposition. For any discrete valued RV's U, V, and W

$$H(W) - I(U \wedge V) \leq H(W|U) + H(W|V).$$

Proof.

$$\begin{aligned} H(UV) &\leq H(UVW) = H(V|WU) + H(W|U) + H(U) \\ &\leq H(V|W) + H(W|U) + H(U) \end{aligned}$$

and therefore

$$H(W) + H(UV) \leq H(W,V) + H(W,U)$$

or

$$H(W) - I(U \wedge V) \leq H(W|V) + H(W|U).$$

Consequently, for $W = X$ and independent U and V,

$$H(X) \leq H(X|U) + H(X|V). \tag{11}$$

If, in addition, $H(X) = 1$ and U and V are zero–one–valued, then also $h(D_1) \geq H(X|U)$ and $h(D_2) \geq H(X|V)$ (Fano), and (10) follows.

References

[1] H.S. Witsenhausen, "On source networks with minimal breakdown degradation", Bell Syst. Tech. J., vol. 59, no. 6, 1083–1087, July–Aug. 1980.

[2] J.K. Wolf, A.D. Wyner, and J. Ziv, "Source coding for multiple descriptions", Bell Syst. Tech. J., vol. 59, no. 8, 1417–1426, Oct. 1980.

[3] H.S. Witsenhausen, “On team guessing with independent information”, Math. Oper. Res., vol. 6, no. 2, 293–304, 1981.

[4] H.S. Witsenhausen, “Team guessing with lacunary information”, Math. Oper. Res., vol. 8, no. 1, 110–121, Feb. 1983.

[5] H.S. Witzenhausen and A.D. Wyner, “Source coding for multiple descriptions II: A binary source”, Bell. Syst. Tech. J., vol. 60, no. 10, 2281–2292, 1981.

[6] A.A. El Gamal and T.M. Cover, “Achievable rates for multiple descriptions”, IEEE Trans. Inform. Theory, vol. IT–28, no. 6, 851–857, Nov. 1982.

[7] T. Berger and Z. Zhang, “Minimum breakdown degradation in binary source coding”, preprint.

[8] L. Ozarow, “On a source–coding problem with two channels and three receivers”, Bell. Syst. Tech. J., vol. 59, no. 10, 1909–1921, 1980.

[9] T. Berger and Z. Zhang, “A theorem about diversity coding”, in preparation.

[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1982.

[11] R. Ahlswede, “Coloring hypergraphs: a new approach to multi–user source coding”, part I, J. Comb. Inform. Syst. Sci., vol. 4, no. 1, 76–115, 1979; part II ibid., vol. 5, no. 3, 220–268, 1980.

[12] R. Ahlswede and G. Dueck, “Good codes can be produced by a feq permutations”, IEEE Trans. Inform. Theory, vol. IT–28, no. 3, 430–443, 1982.

[13] R. Ahlswede, “On two–way communication channels and a problem by Zarankiewic”, in Trans. Sixth Prague Conf. Information Theory, Sept. 1971, 23–37, 1973.

[14] G. Dueck, “The strong converse to the coding theorem for the multiple access channel”, J. Comb. Inform. Syst. Sci., vol. 6, no. 3, 187–196, 1981.

[15] R. Ahlswede, “An elementary proof of the strong converse theorem for the multiple–access channel”, J. Comb. Inform. Syst. Sci., vol. 7, no. 3, 216–230, 1982.

[16] M.S. Pinsker, “Information and information stability of random variables and processes”, (in Russian), vol. 7, 1960. English translation, San Francisco, CA: Holden–Day, 1964.

[17] R. Ahlswede, “The distortion region of a binary source for multiple descriptions without excess rate”, IEEE Trans. Inform. Theory.

[18] R. Ahlswede, “The rate–distortion region for multiple descriptions without excess rate”, IEEE Trans. Inform. Theory, vol. IT–31, 721–726, Nov. 1985.