Comments on the Equitz-Cover Proof of Necessity in SR

Equitz and Cover prove the necessity of the Markov condition $X - X_2 - X_1$ for successive refinement (SR) of the IID source $\{X_k\}$ from D_1 to $D_2 \leq D_1$ as follows.

First, they argue that the SR problem is a special case of the 2-descriptions problem of multiple descriptions theory in which no restriction is placed on one of the marginal distortions, namely D_0 in their notation. Then they observe that the condition that SR places on the refined decoding, namely that the sum rate $R_0 + R_1$ must equal $R(D_2)$, says that there is no excess rate for the refined decoding task. (Here, $R(\cdot)$ is the rate-distortion function of $\{X_k\}$ with respect to a specified single-letter distortion measure $d(x, \hat{x})$.) This allows them to invoke Ahlswede's result that the El Gamal - Cover (EGC) region for the 2-descriptions problem is tight in any no-excess-rate case to help them derive their necessity condition for the SR problem. Specifically, it implies that every quadruple (R_0, R_1, D_1, D_2) that is admissible in the problem of SR from D_1 to D_2 is generated by some conditional pmf $p(\hat{x}_0, \hat{x}_1, \hat{x}_2|x)$ for RV's $\hat{X}_0, \hat{X}_1, \hat{X}_2$ jointly distributed with a canonical source RV X for which $Ed(X, \hat{X}_m) \leq D_m, m = 0, 1, 2,$ $R_i \geq I(X; \hat{X}_i), i = 0, 1,$ and $R_0 + R_1 \geq I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1)$. It follows that SR admissibility requires that

$$R(D_2) = R_0 + R_1 \geq I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1)$$

$$= I(X; \hat{X}_2) + I(X; \hat{X}_1|\hat{X}_2) + I(X; \hat{X}_0|\hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1)$$

$$\geq I(X; \hat{X}_2) \geq R(D_2),$$

where the first equality reiterates that the refined decoding is (R,D)-optimum, the first inequality comes from the EGC condition for the rate sum, the second equality is the chain rule for mutual information, the second inequality is a consequence of the nonnegativity of (conditional) mutual information, and the final inequality comes from the definition of the rate-distortion function coupled with the EGC condition $Ed(X, \hat{X}_2) \leq D_2$. Since the two extremes of this inequality string are identical, the inequalities in it all must be equalities. The middle one of said inequalities therefore implies that each of the last three mutual informations in the step in which the chain rule expansion was used must be 0. The first of these, namely $I(X; \hat{X}_1|\hat{X}_2) = 0$, is the Equitz-Cover-Koshelev (ECK) Markov condition, $X - \hat{X}_2 - \hat{X}_1$, since it says that, given $\hat{X}_2$, X and $\hat{X}_1$ are conditionally independent.

Note that we have established the necessity of satisfaction of the ECK condition in the SR problem without having used the fact that SR from D_1 to D_2 also requires that $I(X; \hat{X}_1) = R(D_1)$, i.e., that the coarse reconstruction also is (R,D)-optimum. We shall return to this fact shortly. First, let's see why the other two mutual information expressions in the chain rule step that must vanish should indeed be zero for any solution of the SR problem.

Satisfaction of the condition $I(\hat{X}_0; \hat{X}_1) = 0$ is essential because otherwise there is some common information in $\hat{X}_0$ and $\hat{X}_1$ that is being sent over both transmision paths. To see why this is inconsistent with SR, we note from Shannon's first theorem that $R_i \geq H(\hat{X}_i), i = 0, 1$ because $\hat{X}_i$ is generated on the basis of information supplied at rate R_i. If $I(\hat{X}_0; \hat{X}_1) > 0$ were to hold, then we would have

$$R_0 + R_1 \geq H(\hat{X}_0) + H(\hat{X}_1) > H(\hat{X}_0) + H(\hat{X}_1|\hat{X}_0) = H(\hat{X}_0, \hat{X}_1).$$

But $H(\hat{X}_0, \hat{X}_1)$ must be at least as great as H($\hat{X}_2$) because $\hat{X}_2$ is a function of $(\hat{X}_0, \hat{X}_1)$. Thus, violation of $I(\hat{X}_0; \hat{X}_1) = 0$ would imply that $R_0 + R_1 > H(\hat{X}_2)$. But this would mean that there is rate loss in the refined decoding because the $\hat{X}_2$ it uses to approximate X could have been conveyed using a sum rate arbitrarily close to $H(\hat{X}_2)$ that is strictly less than the rate $R_0 + R_1$ that was used.

Similarly, satisfaction of the condition $I(X; \hat{X}_0|\hat{X}_1, \hat{X}_2) = 0$ also is essential to achievement of SR. This is because violation of it would imply that $\hat{X}_0$ has information about X that was not incorparted into $\hat{X}_2$ when it was calculated from $(\hat{X}_0, \hat{X}_1)$. Hence, a reduced version of $\hat{X}_0$, call it $\hat{X}_0^*$, could have been sent instead of $\hat{X}_0$ using a rate $R_0^* < R_0$ and being such that the same $\hat{X}_2$ could be calculated from $(\hat{X}_0^*, \hat{X}_1)$ that was calculated from $(\hat{X}_0, \hat{X}_1)$. This would mean that $\hat{X}_2$ could have been conveyed to the refined decoder using a sum rate of only $R_0^* + R_1$ which is strictly less than $R_0 + R_1$. That, in turn, would contradict the fact that the quadruple (R_0, R_1, D_1, D_2) satisfies the no-excess-rate condition, thereby preventing it from being an SR 4-tuple.

Where then, if at all, does the (R,D)-optimality of the coarse description come into the necessity proof? The answer is that it does not. Since that (R,D)-optimality, namely $I(X; \hat{X}_1) = R(D_1)$, is a condition for SR, we may conclude that SR from D_1 to D_2 is not acheivable unless there exists a random vector $(\hat{X}_0, \hat{X}_1, \hat{X}_2)$ jointly distributed with X that satisfies all the conditions associated with membership in the EGC region for the 2-descriptions problem when D_0 is arbitrarily large (any D_0 greater than or equal to $D_{\max} = \min_{\hat{x}} Ed(X, \hat{x})$ will do) and satisfies both of the EGC conditions, $ED(X, \hat{X}_1) \leq D_1$ and $ED(X, \hat{X}_2) \leq D_2$) with equality. In general, there is no such $(\hat{X}_0, \hat{X}_1, \hat{X}_2)$-vector because many problems (perhaps one should say most problems) are not successively refinable. In particular, Equitz and Cover show that Gerrish's problem (cf. p 61 of T. Berger, *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1971) is not SR for certain (D_1, D_2)-pairs.