<http://www.diva-portal.org>

Preprint

This is the submitted version of a paper presented at *2018 Data Compression Conference, DCC 2018, 27 March 2018 through 30 March 2018*.

Citation for the original published paper:

Vu, M T., Oechtering, T J., Skoglund, M. (2018)

Gaussian hierarchical identification with pre-processing

In: *Data Compression Conference Proceedings* (pp. 277-286). Institute of Electrical and Electronics Engineers Inc.

<https://doi.org/10.1109/DCC.2018.00036>

N.B. When citing this work, cite the original published paper.

The work was supported by the Swedish Research Council (VR) project under Grant 2016-03853

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-238076>

Gaussian Hierarchical Identification with Pre-processing

Minh Thanh Vu, Tobias J. Oechtering, Mikael Skoglund
Department of Information Science and Engineering
KTH-Royal Institute of Technology

Abstract

In this work we consider a two-stage identification problem with pre-processing where the users' data and observation are Gaussian distributed. In the first stage the processing unit returns a list of compatible users using the information from the first storage layer and the pre-processed observation. Then, the refined search is performed in the second stage where the processing unit returns the exact user's identity and a corresponding reconstruction sequence. We provide a complete rate-distortion trade-off for the Gaussian setting.

1 Introduction & Main result

The identification problem was initiated by Willems in [1] where he characterized the identification capacity. The works [2, 3] extended the results by considering compression, distortion and identification trade-offs. Our current work is motivated by the following example. Consider a forensic application, where investigators would like to identify suspects quickly and view their criminal history. The search complexity can be reduced by dividing the whole processing into two stages. In the first one, the processing center returns a list of possible suspects. In the second stage the search is refined and the record is returned. The above application is modelled as in Fig. 1. In the enrollment phase data from each of M users are stored in two indices $j_i \in [1 : |\mathcal{M}_1|]$, $k_i \in [1 : |\mathcal{M}_2|]$ in two storage nodes, respectively. The observation y^n , which is provided to the observer and correlated with the true user, is first passed through a pre-processing filter $p_{Z|Y}$. In the first processing stage a list of compatible users $\mathcal{L}$, with a constrained size, is returned based on the first layer information $(j_i)_{i=1}^M$, and the pre-processed observation z^n , i.e.,

$$\mathcal{L} = g_1(z^n, (j_i)_{i=1}^M), \quad |\mathcal{L}| \leq 2^{n\Delta}.$$

In the second step the system returns a user's identity as well as a reconstruction sequence based on the observation y^n and information in the list from both layers $(j_{\mathcal{L}}, k_{\mathcal{L}})$, i.e.,

$$\hat{w} = g_2(y^n, j_{\mathcal{L}}, k_{\mathcal{L}}), \quad \hat{x}^n = g_3(y^n, j_{\hat{w}}, k_{\hat{w}}, \hat{w}).$$

We note that the two storage nodes belong to the same cluster, i.e., we do not consider clustering in our setting as in [4] for example. For a more detailed introduction of the setup, the reader is referred to [5]. In the current work we consider the setup in which for each $w \in \mathcal{W}$ the corresponding user's data $X^n(w)$ are iid Gaussian

Figure 1: An overview of the two stage identification system. We assume that there always exists a user W which has been enrolled previously and to which the observation Y^n is the output of a memoryless channel $p_{Y|X}$ with the input $X^n(W)$. Furthermore, W is uniformly distributed over $[1 : M]$ and independent of users' data. The first and second layer information are represented by the collections $(J_i)_{i=1}^M$ and $(K_i)_{i=1}^M$, respectively.

$X_i(w) \sim \mathcal{N}(0, \sigma_X^2)$. The observation Y^n , and the pre-processing output Z^n are jointly iid Gaussian distributed according to the following relation

$$Y_i = X_i(W) + N_{1i}, \quad Z_i = Y_i + N_{2i},$$

where $N_{1i} \sim \mathcal{N}(0, \sigma_{N_1}^2)$ and $N_{2i} \sim \mathcal{N}(0, \sigma_{N_2}^2)$ are independent Gaussian random variables, also of the users' data.

Definition 1. A rate-distortion tuple (R, R_1, R_2, R_L, D) is achievable if for every $\epsilon > 0$, there exists a data processing scheme of length n such that

$$\begin{aligned} \frac{1}{n} \log M &> R - \epsilon, \quad \frac{1}{n} \log |\mathcal{M}_1| < R_1 + \epsilon \\ \frac{1}{n} \log |\mathcal{M}_2| &< R_2 + \epsilon, \quad \Delta < R_L + \epsilon, \quad \Pr(W \notin \mathcal{L}) < \epsilon, \\ \Pr(W \neq \hat{W}) &< \epsilon, \quad \mathbb{E}[d(X^n, \hat{X}^n)] < D + \epsilon, \end{aligned}$$

for all sufficiently large n , where the distortion measure d is the squared error distance. The set of all achievable tuples is denoted by $\mathcal{R}_{GS}$.

Our result can be summarized as:

Theorem 1. Assume that $R_L \leq R$ and $0 \leq D \leq \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$. Then the corresponding rate-distortion region $\mathcal{R}_{GS}$ is given by

$$R_L \leq R < R_\gamma = \min \left\{ \frac{1}{2} \log_2 \left(1 + \frac{\sigma_X^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2} \right) + R_L, \frac{1}{2} \log_2 \left(\frac{\sigma_Y^2}{\sigma_{N_1}^2} \right) \right\},$$

$$\begin{aligned}
R_1 &\geq \frac{1}{2} \log_2 \left(\frac{\sigma_X^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)} \right), \\
R_1 + R_2 &\geq \frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \Gamma, \quad R_1 + R_2 - R \geq \Gamma,
\end{aligned} \tag{1}$$

where

$$\Gamma = \frac{1}{2} \max \left\{ \underbrace{\log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 D}}_{:=h_0(R)}, \underbrace{\log_2 \frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2}}_{:=h_1(R)}, \underbrace{\log_2 \frac{\sigma_X^2}{\sigma_Y^2} \frac{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}}_{:=h_2(R)} \right\}.$$

As a corollary, in the case that one of the two assumptions is violated then the corresponding terms related to the mentioned parameters in (1) is omitted. For example the rate-distortion trade-off when $R_L > L$ and $0 \leq D \leq \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$ is given by

$$\begin{aligned}
R_1 + R_2 &\geq R + \frac{1}{2} \max \left\{ \log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 D}, \log_2 \frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2} \right\}, \\
0 \leq R &< \frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_{N_1}^2}.
\end{aligned} \tag{2}$$

In the next section we provide a sketch including several detailed descriptions of key steps of the proof. An important step is the study of relations among functions $h_0(R)$, $h_1(R)$, and $h_2(R)$ that allows us to see in which regime we are operating. Those functions are first obtained in an outer bound. In the next step it is shown how the outer bound is achieved. Thereby the functions play the role to decide how to parameterize the variances of the auxiliary random variables. Theorem 1 is a non-trivial extension of [5, Theorem 1].

2 Sketch of Proof

We first establish an outer bound on the achievable rate-distortion region. The outerbound provides the characterization of the sub-regions that can be achieved with differently parameterized coding schemes presented in Section 2.4. The characterization of the transition points is provided in Section 2.3.

2.1 Study of extreme cases

We first consider extreme cases which provide some points and hints about the whole rate-distortion region.

- Our setup can be regarded as a blowing up of the Heegard-Berger [6] scheme where there is no constraint on the distortion in the first layer, i.e., the distortion constraint in the first layer is “viewed” as ∞ . Hence when additionally $R_L = R = 0$, which also reduces the setting to the Wyner-Ziv problem, the rate region should collapse into

$$R_1 + R_2 \geq \frac{1}{2} \log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 D}. \tag{3}$$

- For a given R_L the identification capacity is the minimum of the first stage identification capacity $\frac{1}{2} \log_2 \left(1 + \frac{\sigma_X^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2} \right) + R_L$ and the identification capacity $\frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_{N_1}^2}$ when the processing unit has the full access to both storage nodes.

Additionally, given a fixed target distortion D and a fixed list size R_L which is large enough such that $R_\gamma = 1/2 \log_2(\sigma_Y^2/\sigma_{N_1}^2)$, if we increase the identification rate R to the threshold R_γ then the achieved distortion must be lower than the target distortion D . One can explain this observation as follows. In order for the identification rate to come close to the identification capacity, the compressed information must be close to the corresponding user's data, i.e., the distortion level for stored sequences will be extremely small. This provides a hint that there will be a transition point from a region where the distortion constraint is active to a region where the distortion constraint is inactive when R increases. When the list size R_L is small or moderate, there exist additional transition points where the identification rate is limited at the first stage.

2.2 An outerbound

Suppose that the rate-distortion tuple (R, R_1, R_2, R_L, D) is achievable, i.e., for a given $\epsilon > 0$ there exists a data processing scheme for *all* sufficiently large n . The distortion constraint implies that for such n

$$\begin{aligned} D + \epsilon &> \mathbb{E}[d(X^n(W), g_3(\hat{W}, J_{\hat{W}}, K_{\hat{W}}, Y^n))] \\ &\geq \frac{1}{n} \sum_{i=1}^n \inf_{g_i} \mathbb{E}[d(X_i(W), g_i(W, (J_l)_{l=1}^M, (K_l)_{l=1}^M, \hat{W}, J_{\hat{W}}, K_{\hat{W}}, Y^n))]. \end{aligned} \quad (4)$$

Our coding scheme induces the following Markov chain

$$X^n(W) - (Y^n, W, J_W, K_W) - (Z^n, (J_l)_{l \neq W}, (K_l)_{l \neq W}, \hat{W}, J_{\hat{W}}, K_{\hat{W}}), \quad (5)$$

where we use $(J_l)_{l \neq W}$ as a shorthand notation of $(J_l)_{l=1, l \neq W}^M$ and similarly for $(K_l)_{l \neq W}$. Since the distortion measure is the squared error, this implies that

$$D + \epsilon > \sum_{i=1}^n \frac{1}{n} \mathbb{E}[d(X_i(W), \mathbb{E}[X_i(W)|W, J_W, K_W, Y^n])]. \quad (6)$$

The constraint (6) can be interpreted as following. Assume that a genie provides us the exact information (W, J_W, K_W) and we employ the optimal estimator in the squared error sense to reconstruct the correct user's data sequence based on the provided information and the available information $(\hat{W}, J_{\hat{W}}, K_{\hat{W}}, Y^n)$. Then the optimal estimator depends only on (W, J_W, K_W, Y^n) and the resulting distortion is still below the given target.

Since Gaussian distribution maximizes the differential entropy of a given variance, eq. (6) implies that

$$h(X^n(W)|J_W, K_W, W, Y^n) \leq \frac{n}{2} \log_2(2\pi e(D + \epsilon)). \quad (7)$$

It can be seen that the assumption $D \leq \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$ is to make the constraint (7) possibly active. Otherwise the distortion constraint can be dropped. Next, using a variant of Fano's inequality we arrive at the following expression, cf. [5, Eq. (33)] for the details,

$$\begin{aligned} n(R - \epsilon) &\leq \log M \leq I(W, J_W; Z^n) + n(R_L + \epsilon_n) \\ &= h(Z^n) - h(Z^n|W, J_W) + n(R_L + \epsilon_n) \\ &= \frac{n}{2} \log_2(2\pi e \sigma_Z^2) - h(Z^n|W, J_W) + n(R_L + \epsilon_n), \end{aligned} \quad (8)$$

where $\epsilon_n = 2\epsilon + \frac{1}{n}\epsilon \log_2 M$. This leads to

$$h(Z^n|W, J_W) \leq \frac{n}{2} \log_2(2\pi e \sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon}).$$

Due to the entropy power inequality we obtain

$$\begin{aligned} 2\pi e(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - \sigma_{N_2}^2) &\geq 2^{\frac{2}{n}h(Y^n|W, J_W)} \\ 2\pi e(\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)) &\geq 2^{\frac{2}{n}h(X^n|W, J_W)}. \end{aligned} \quad (9)$$

Since $h(X^n|W, J_W) > -\infty$ we therefore need the following condition

$$(R - \epsilon)(1 - \epsilon) - R_L - 2\epsilon < \frac{1}{2} \log_2 \left(\frac{\sigma_Z^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2} \right), \quad (10)$$

which leads to $R \leq \frac{1}{2} \log_2 \sigma_Z^2 / (\sigma_{N_1}^2 + \sigma_{N_2}^2) + R_L$, as we take $\epsilon \rightarrow 0$. Again the assumption $R_L \leq R$ is to make the above constraint possibly active. Otherwise the list size constraint can be dropped. Additionally, corresponding to [5, Eq. (35)] we obtain

$$\begin{aligned} n(R - \epsilon) &\leq I(W, J_W, K_W; Y^n) + 1 + \epsilon \log_2 M \\ &= \frac{n}{2} \log_2(2\pi e \sigma_Y^2) - h(Y^n|W, J_W, K_W) + 1 + \epsilon \log_2 M, \end{aligned}$$

where the Markov chain $X - Y - Z$ is necessary for the first inequality, which leads to

$$h(Y^n|W, J_W, K_W) \leq \frac{n}{2} \log_2(2\pi e \sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon}).$$

From the entropy power inequality we know that

$$2\pi e \sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} \geq 2^{\frac{2}{n}h(Y^n|W, J_W, K_W)} > 2\pi e \sigma_{N_1}^2. \quad (11)$$

Since $h(X^n|W, J_W, K_W) > -\infty$, there exists an α_1 with $0 \leq \alpha_1 < 1$, which depends on other parameters, such that

$$h(Y^n|W, J_W, K_W) = \frac{n}{2} \log_2(2\pi e((1 - \alpha_1)\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} + \alpha_1\sigma_{N_1}^2)), \quad (12)$$

and

$$h(X^n(W)|W, J_W, K_W) \leq \frac{n}{2} \log_2(2\pi e(1 - \alpha_1)(\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)). \quad (13)$$

From (11) we also obtain a constraint on the rate R , namely $R \leq \frac{1}{2} \log_2 \sigma_Y^2 / \sigma_{N_1}^2$. Thus $0 \leq R_L \leq R \leq R_\gamma$. The inequalities in (9) indicate that

$$\begin{aligned} n(R_1 + \epsilon) &\geq I(X^n(W); J_W, W) \\ &\geq \frac{n}{2} \log_2 \left(\frac{\sigma_X^2}{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)} \right). \end{aligned}$$

Taking $\epsilon \rightarrow 0$ we obtain the first condition on R_1 in (1). Similarly, corresponding to [5, Eq. (37)] we obtain

$$\begin{aligned} n(R_1 + R_2 + \epsilon) &= \underbrace{h(Y^n) - h(Y^n|W, J_W)}_{\Delta_1} + \underbrace{h(X^n(W)|Y^n) - h(X^n(W)|J_W, K_W, W, Y^n)}_{\Delta_2} \end{aligned} \quad (14)$$

The first term in (14) is bounded based on (9) as

$$\Delta_1 \geq \frac{n}{2} \left(\log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - \sigma_{N_2}^2} \right). \quad (15)$$

The second term in (14) is bounded in three different ways. From (7) we obtain

$$\Delta_2 \geq \frac{n}{2} \log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 (D + \epsilon)}. \quad (16)$$

This implies in combination with (15) that

$$R_1 + R_2 \geq \frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2} \log_2 \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2 D}. \quad (17)$$

Secondly, the expressions in (12) and (13) lead to

$$\begin{aligned} \Delta_2 &\geq \frac{n}{2} \log_2 \left(\frac{\sigma_X^2 ((1 - \alpha_1) \sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} + \alpha_1 \sigma_{N_1}^2)}{\sigma_Y^2 (1 - \alpha_1) (\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)} \right) \\ &\stackrel{(a)}{\geq} \frac{n}{2} \log_2 \left(\frac{\sigma_X^2}{\sigma_Y^2 (\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)} \inf_{0 \leq \alpha_1 < 1} \left(\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} + \frac{\alpha_1}{1 - \alpha_1} \sigma_{N_1}^2 \right) \right). \end{aligned}$$

We note that due to the inequality (11) the term $\frac{\sigma_X^2}{\sigma_Y^2 (\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2)}$ is positive hence (a) is valid. Note also that since α_1 might depend on ϵ and n , we should avoid taking the limit directly. Since $\frac{\alpha_1}{1 - \alpha_1}$ is an increasing and positive function of α_1 on $[0, 1)$, the infimum is attained at $\alpha_1 = 0$. Hence

$$\Delta_2 \geq \frac{n}{2} \log_2 \left(\frac{\sigma_X^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon}}{\sigma_Y^2 2^{-2(R-\epsilon)(1-\epsilon)+2\epsilon} - \sigma_{N_1}^2} \right). \quad (18)$$

This implies by taking $\epsilon \rightarrow 0$ that

$$R_1 + R_2 \geq \frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2} \log_2 \left(\frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2} \right). \quad (19)$$

We optimize over the additional parameter α_1 to simplify the corresponding outer-bound. It will be shown latter that this optimized step does not break the optimality of the bound, i.e., the outerbound can still be achievable.

Lastly, using a similar argument as the one which leads to (18) we obtain, (cf. (10) for the regularity condition)

$$\Delta_2 \geq \frac{n}{2} \log_2 \left(\frac{\sigma_X^2}{\sigma_Y^2} \frac{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - \sigma_{N_2}^2}{\sigma_Z^2 2^{-2((R-\epsilon)(1-\epsilon)-R_L)+4\epsilon} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)} \right). \quad (20)$$

Thus

$$R_1 + R_2 \geq \frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2} \log_2 \frac{\sigma_X^2}{\sigma_Y^2} \frac{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2}{\sigma_Z^2 2^{-2(R-R_L)} - (\sigma_{N_1}^2 + \sigma_{N_2}^2)}, \quad (21)$$

by taking $\epsilon \rightarrow 0$. Combining these three bounds we obtain

$$R_1 + R_2 \geq \frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \Gamma. \quad (22)$$

Additionally, we have the following constraint which corresponds to [5, Eq. (39)]

$$\begin{aligned} & n(R_1 + R_2 + \epsilon) - \log_2 M \\ & \geq I(X^n(W); W, J_W, K_W | Y^n) - (1 + \epsilon \log_2 M) \\ & = \Delta_2 - (1 + \epsilon \log_2 M). \end{aligned} \quad (23)$$

which implies that

$$R_1 + R_2 - R \geq \Gamma. \quad (24)$$

As $R \rightarrow R_\gamma$ either $h_1(R)$ or $h_2(R)$ goes to infinity. However since both R_1 and R_2 are finite we must have

$$0 \leq R_L \leq R < R_\gamma. \quad (25)$$

In summary we obtain an outerbound which is given by (1).

2.3 Analysis of the outerbound

We show that for fixed D and R_L the three functions $h_0(R)$, $h_1(R)$ and $h_2(R)$, which are the three defining functions of Γ , provide the key for the transition behavior from one extreme case to another. Thus, the dominant term specifies how to pick the auxiliary random variables. Since these functions are monotone in R the intersections characterize the phase transitions. In this subsection we therefore find all possible

phase transition points.

Consider the following equation $h_1(R) = h_2(R)$ which implies that

$$R_{cr12} = \frac{1}{2} \log_2 \frac{\sigma_Z^2 2^{2R_L} - \sigma_Y^2}{\sigma_{N_2}^2} \geq R_L, \quad (26)$$

since $R_L \geq 0$. Note, however, that R_{cr12} can lie outside the interval $[R_L, R_\gamma)$. We observe that $h_1(0) = 0 \geq h_2(0)$ as

$$\sigma_Y^2 \left(1 - \frac{\sigma_{N_1}^2}{\sigma_Z^2 2^{2R_L} - \sigma_{N_2}^2}\right) \geq \sigma_X^2 \quad (27)$$

holds since $R_L \geq 0$. Moreover it can be seen that both $h_1(R)$ and $h_2(R)$ are increasing functions of R in their corresponding domains. Thus for $0 \leq R \leq \min\{R_\gamma, R_{cr12}\}$, $h_1(R) \geq h_2(R)$.

Next, note that $h_1(0) = 0$ and $h_1(R) \rightarrow \infty$ as $R \rightarrow \frac{1}{2} \log_2 \sigma_Y^2 / \sigma_{N_1}^2$. Therefore, there is a unique point $R_{cr01} \in [0, \frac{1}{2} \log_2 \sigma_Y^2 / \sigma_{N_1}^2)$ such that $h_1(R) = h_0(R)$, i.e.,

$$\frac{\sigma_{N_1}^2}{\sigma_Y^2 D} = \frac{2^{-2R_{cr01}}}{\sigma_Y^2 2^{-2R_{cr01}} - \sigma_{N_1}^2}, \quad (28)$$

above which the $h_1(R)$ dominates $h_0(R)$. Solving for R_{cr01} we obtain

$$R_{cr01} = \frac{1}{2} \log_2 \frac{\sigma_Y^2 (1 - \frac{D}{\sigma_{N_1}^2})}{\sigma_{N_1}^2}. \quad (29)$$

Analogously, there exists a unique point $R_{cr02} \in [R_L, R_L + \frac{1}{2} \log_2 \sigma_Z^2 / (\sigma_{N_1}^2 + \sigma_{N_2}^2))$ which is the solution of the equation $h_2(R) = h_0(R)$, i.e.,

$$D = \sigma_{N_1}^2 \left(1 - \frac{\sigma_{N_1}^2}{\sigma_Z^2 2^{-2(R_{cr02} - R_L)} - \sigma_{N_2}^2}\right). \quad (30)$$

When $D \rightarrow 0$, we observe that as $R \rightarrow R_\gamma$ either $h_1(R)$ or $h_2(R)$ goes to infinity. Hence at least one of the points R_{cr01} or R_{cr02} lies in the interval $[R_L, R_\gamma)$. If $D \rightarrow \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$, then R_{cr01} might lie outside the domain since $R_L \leq R < R_\gamma$. In this case R_{cr02} is always inside. An illustration of possible cases is given in Fig. 2

2.4 Achievability

Fig. 2 shows which function to focus for a certain rate, i.e., in each case we find auxiliary random variables to match the corresponding dominant function.

Fix a value of D and R_L where $0 \leq D \leq \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_Y^2}$ and $0 \leq R_L < R_\gamma$. Due to space limit we present here the case where $R_{cr12} < R_\gamma$ and $R_{cr01} \leq R \leq R_{cr12}$. The other cases can be done following the same principle.

Since $R_{cr12} \geq R_L$ holds, cf. (26), the above interval should be truncated if necessary so that $R \geq R_L$ holds. For $R_{cr01} \leq R \leq R_{cr12}$, $h_1(R)$ is the dominant component

Figure 2: $R_{cr12} < R_\gamma$ and $R_L \leq R_{cr01} \leq R_{cr12}$. We can see that when $R_L \leq R \leq R_{cr01}$, $h_0(R)$ dominates over $h_1(R)$ and $h_2(R)$. $h_1(R)$ is the dominant component when $R_{cr01} < R \leq R_{cr12}$. When $R_{cr12} < R < R_\gamma$ then $h_2(R)$ dominates the other two functions.

in the outerbound since $h_1(R) \geq h_0(R)$ when $R \geq R_{cr01}$ and $h_1(R) \geq h_2(R)$ when $R \leq R_{cr12}$. Let $X = V + N_0$ where V and N_0 are independent Gaussian random variables, where $\sigma_V^2 = \sigma_Y^2(1 - 2^{-2R}) \leq \sigma_X^2$ since $R < R_\gamma$. V should be understood as the output of the *test channel* $p_{V|X}$, cf. [7, p. 311]. Then, let $V = U + N'_0$ where U and N'_0 are independent Gaussian random variables such that $\sigma_U^2 = \sigma_Z^2(1 - 2^{-(R-R_L)}) \geq 0$ if $R \geq R_L$. Similarly, U is the output of the test channel $p_{U|V}$. We note that

$$2^{-2R}(\sigma_Z^2 2^{2R_L} - \sigma_Y^2) \geq \sigma_{N_2}^2 \text{ or } \sigma_Y^2(1 - 2^{-2R}) \geq \sigma_Z^2(1 - 2^{-2(R-R_L)}), \quad (31)$$

since $R \leq R_{cr12}$. This means that $\sigma_U^2 \leq \sigma_V^2$. By our choice of U and V the relation $U - V - X - Y - Z$ holds. We next examine whether the chosen random variables satisfy the constraints corresponding to the fixed parameters. The condition $I(Z; U) = R - R_L$ is satisfied by the chosen U . Furthermore, $I(Y; V) = R$ due to the choice of V . This means that the choice of U and V does not violate the constraint

$$R \leq \min\{I(Z; U) + R_L, I(Y; V)\}. \quad (32)$$

Next, we calculate

$$h(X|V, Y) = \frac{1}{2} \log_2 \left(2\pi e \frac{\sigma_{N_1}^2 \sigma_Y^2 2^{-2R} - \sigma_{N_1}^2}{\sigma_Y^2} \right) \stackrel{(*)}{\leq} \frac{1}{2} \log_2(2\pi e D)$$

where $(*)$ is valid due to (28) as $R \geq R_{cr01}$, i.e., the distortion level D is attainable by the MMSE decoder. Since

$$I(X; U) + I(X; V|U, Y) = I(Y; U) + I(X; V|Y) \quad (33)$$

we obtain that

$$R_1 + R_2 \geq \frac{1}{2} \log_2 \frac{\sigma_Y^2}{\sigma_Z^2 2^{-2(R-R_L)} - \sigma_{N_2}^2} + \frac{1}{2} \log_2 \frac{\sigma_X^2 2^{-2R}}{\sigma_Y^2 2^{-2R} - \sigma_{N_1}^2}. \quad (34)$$

The other constraints be seen to be satisfied similarly.

References

- [1] F. M. J. Willems, T. Kalker, and J.-P. Linnartz, “On the capacity of a biometrical identification system,” in *IEEE International Symposium on Information Theory*, 2003.
- [2] E. Tuncel, “Capacity/storage tradeoff in high-dimensional identification systems,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2097–2106, 2009.
- [3] E. Tuncel and D. Gündüz, “Identification and lossy reconstruction in noisy databases,” *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 822–831, 2014.
- [4] F. Farhadzadeh and F. M. J. Willems, “Identification rate, search and memory complexity tradeoff: Fundamental limits,” *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6173–6188, 2016.
- [5] M. T. Vu, T. Oechtering, and M. Skoglund, “Hierarchical identification with pre-processing,” in *IEEE International Symposium on Information Theory*. IEEE, 2017.
- [6] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, 1985.
- [7] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.