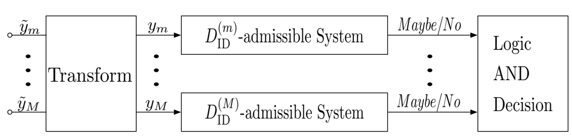
Component-based Quadratic Similarity Identification for Multivariate Gaussian Sources

Hanwei Wu and Markus Flierl
School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology, Stockholm
{hanwei, mflierl}@kth.se

In this work, we investigate the component-based quadratic similarity identification [1] for multivariate Gaussian sources. Consider a concatenation of N independent blocks of correlated zero-mean Gaussian random variables with blocklength M for D_{ID}-similarity identification, where $n = MN$ is the length of the source sequence, and D_{ID} the similarity threshold. The blocks are decorrelated by the Karhunen-Loève (KL) transform. As we consider a Gaussian source, the KL transform outputs M independent Gaussian components with their corresponding variances. Then, each individual component is processed by a $D_{\mathrm{ID}}^{(m)}$-*admissible* system, where $D_{\mathrm{ID}}^{(m)}$ is the similarity threshold for the m-th component. A database vector is labeled as *maybe*, if and only if all its transform components are determined as *maybe*.

The bit allocation for the components is achieved by a constrained minimization of the overall Pr{maybe}. We use $\Pr\{\text{maybe}\} \approx e^{-\frac{n}{M}\mathbb{E}_{\mathrm{ID}}}$ for large $\frac{n}{M}$, where $\mathbb{E}_{\mathrm{ID}}$ is the *identification exponent* [1]. This leads to a maximization of the average identification exponent of M components subject to the total *identification rate* constraint R_{ID} and the similarity constraint D_{ID}. The latter guarantees that the M-component system maintains D_{ID}-*admissibility.*



$$\max_{(R_1,D_1)\ldots,(R_M,D_M)} \quad \sum_{i=1}^{M} \mathbb{E}_{\mathrm{ID}}^{(m)}(R_m)$$

$$\text{s.t.} \quad \frac{1}{M}\sum_{m=1}^{M} R_m \le R_{\mathrm{ID}}$$

$$\text{s.t.} \quad \frac{1}{M}\sum_{m=1}^{M} D_{\mathrm{ID}}^{(m)} \ge D_{\mathrm{ID}}$$

We consider the special case where one component forms a MD_{ID}-admissible system. For the bit allocation problem, it can be shown that the sum of identification exponents is a quasiconcave function of the assigned component rates. Using quasiconcave programming [2], we show that the minimum D-*achievable* rate for this setting is achieved by the identification rate of the component with the largest variance.

[1] A. Ingber, T. Courtade, and T. Weissman, "Compression for quadratic similarity queries," *IEEE Trans. on Information Theory*, vol. 61, no. 5, pp. 2729 –2747, May 2015.

[2] K. J. Arrow and A. Enthoven, "Quasi-concave programming," *Econometrica*, vol. 29, no. 4, pp. 779–800, 1961.