Interactive Fusion in Distributed Detection: Architecture and Performance Analysis

Earnest Akofor and Biao Chen
Department of Electrical Engineering and Computer Science
Syracuse University, Syracuse NY 13244

***Abstract*—Within the Neyman-Pearson framework we investigate the effect of feedback in two-sensor tandem fusion networks with conditionally independent observations. While there is noticeable improvement in performance of the fixed sample size Neyman-Pearson (NP) test, it is shown that feedback has no effect on the asymptotic performance characterized by the Kullback-Leibler (KL) distance. The result can be extended to an interactive fusion system where the fusion center and the sensor may undergo multiple steps of interactions.**

***Index Terms*—Distributed detection, Interactive fusion, Neyman-Pearson test, Kullback-Leibler distance.**

I. Introduction

A simple tandem sensor network typically consists of two sensors, one of them serving as a fusion center and makes a final decision using its own observation as well as input from the other sensor. Practical constraints often dictate that the input from the other sensor is maximally compressed. The extreme case, as adopted in the present work, is that the observation at the other sensor is mapped to a single bit, often referred to as the local decision. Distributed detection with such a tandem network has been relatively well understood under the conditional independence assumption, i.e., the observations at distributed nodes are independent conditioned on a given hypothesis. Specifically, it was known that the optimal local sensor decision rule is in the form of a likelihood ratio test [1]. Fusion architecture, and in particular, the impact of communication direction in a two sensor system was studied in [2].

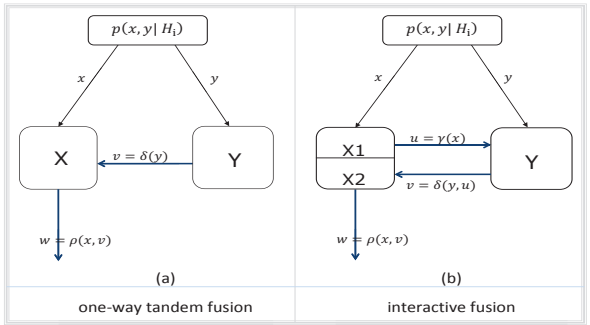

Fig. 1. One-way YX and two-way XYX processes

This paper revisits this simple tandem distributed detection network by replacing the static message passing (from the sensor node to the fusion node) with an interactive one: it is assumed that the fusion center may send an initial bit to the local sensor based on the observation at the fusion center. The local sensor then makes a local decision based on its own observation as well as the input from the fusion center before passing it back to the fusion ceneter. In the most general setting, as to be discussed in the end of Section IV, multiple rounds of interactions may occur. For the most part, we limit ourselves to a single round of interaction and we refer to this communication protocol as the so-called interactive fusion. The contrast between the traditional tandem network and the interactive fusion network is illustrated in Fig. 1.

Let x be the observation of sensor X and y be the observation of sensor Y. Then for a one-way tandem network the decision process (v, w) is based on x and y through dependencies of the form $v = \gamma(x), \quad w = \delta(y, v)$. Similarly, for the interactive model the decision processes based on observations x and y yield outputs (u, v, w), where $u = \gamma(x),\ v = \delta(y, u),\ w = \rho(x, v)$. For simplicity, we refer to the fusion architecture in Fig. 1(a) as the YX process whereas to that of Fig. 1(b) as the XYX process.

This interactive fusion network has been studied under the Bayesian framework and was shown to improve the error probability performance for fixed-sample size test [3]. This paper focuses on the Neyman-Pearson framework and we would like to address whether the additional feedback would improve

1) the performance of the fixed sample size NP test;
2) the asymptotic performance, quantified using the Kullback-Leibler distance which is known to be the error exponent of the NP test, a.k.a., the Chernoff-Stein Lemma [4].

We show that while the answer to the first question is affirmative, feedback does not improve the asymptotic performance, i.e., the answer to the second question is negative.

We note that [5], [6], and [7] examine the effect of feedback from a global fusion center to local sensors. Thus this feedback setting is noninteractive. Further, [6] and [7] considered asymptotics with respect to the number of sensors, contrary to our setup where asymptotics is with respect to the number of independent samples taken over time.

Our presentation is organized as follows. In Section II we describe the procedure for obtaining decision rules given an objective function. The obtained result in Proposition II.1 will be applied in subsequent sections. We demonstrate in Section III that iterative fusion does improve performance of a fixed sample size NP test. For large sample size, however, we show in Section IV that feedback does not improve detection performance characterized using the error exponent. Section V contains concluding remarks.

This material is based on research sponsored by AFOSR under Award FA9550-10-1-0458, by ARO under Award W911NF-12-1-0383, and by NSF under Award 1218289.

II. The underlying decision theory

Consider M simple hypotheses with observation $x \in \mathcal{X}$,

$$H_i: \; x \sim p_i(x), \qquad i = 0, 1, ..., M-1, \tag{1}$$

where $p_i(x) \equiv p(x|H_i)$ is the distribution of x under the ith hypothesis H_i. A *decision rule* is a mapping defined as

$$\gamma: \; x \mapsto i \in \{0, 1, ..., M-1\}, \tag{2}$$

where γ is a deterministic function and $\gamma(x) = i$ denotes acceptance of the ith hypothesis H_i. We refer to the assignment $\gamma(x) = i$ as a *decision* based on the observation x. Here, without loss of generality, we assume that the decision output has the same alphabet as the underlying hypothesis.

The desired decision rule γ so defined is deterministic in the sense that $p(\gamma(x) = i|x) = \delta_{i,\gamma(x)}$, which means that once x is given $\gamma(x)$ is precisely known. As the optimum decision rule is not necessarily deterministic, we consider the larger set containing all deterministic and nondeterministic decision rules. Let us write the generic decision rule as

$$\Gamma: \; x \mapsto i \in \{0, 1, ..., M-1\}, \tag{3}$$

and let $u = \Gamma(x)$. Then $\Gamma = \gamma$ denotes a deterministic choice of decision rule. Recall, [8], that the set of Γ is the convex hull of the set of γ. Therefore

$$p(\Gamma(x) = i) = \sum_g p(g) \; p\big(\gamma_g(x) = i\big) \tag{4}$$

where g is a random variable with pdf $p(g)$ and is independent of x. The decision process simply picks the appropriate $p(g)$, and hence the desired $p(\Gamma(x) = i)$.

As u is a random variable, making an optimal guess $u = i$ is equivalent to choosing $p(u = i|x)$ such that some *objective function*, which we denote by S, is optimized. Here S is a function of $p(u = i|x)$ for all $i = 0, 1, ..., M-1$ and for all data points $x \in \mathcal{X}$.

In general, $0 \le p(u = i|x) \le 1, \quad i = 0, 1, ..., M-1$. A deterministic decision rule implies that $p(u = i|x)$ takes on only the boundary values 0 and 1. For such cases, the decision rule is equivalently expressed as a partition of the sample space into disjoint decision regions, i.e.,

$$\begin{aligned} p_{\text{opt}}(u = i|x) \; &:= \; p(\gamma(x) = i|x) = \delta_{i,\gamma(x)} \\ &= \left\{ \begin{array}{ll} 1, & \gamma(x) = i \\ 0, & \gamma(x) \neq i \end{array} \right\} = I_{R_{u=i}}(x), \end{aligned} \tag{5}$$

where $I_{R_{u=i}}(x) \equiv I_{\{x:\; \gamma(x)=i\}}(x)$ is the indicator function of the region $R_{u=i}$, i.e., the *decision region* for the ith hypothesis.

In the following, we establish the general structure of the optimal decision rule for an important class of decision problems.

Proposition II.1 (Decision regions due to convex objectives)**.** *Let x be a continuous random variable, and suppose the objective function S to be maximized is either an affine or a differentiable convex function of $p(u|x)$. For each i suppose further that the set of data points*

$$C_{u=i} = \{x : \partial S/\partial p_{opt}(u = i|x) = 0\} \tag{6}$$

has zero probability measure, where $\partial S/\partial p_{opt}(u = i|x) = \partial S/\partial p(u = i|x)|_{p(u=i|x)=p_{opt}(u=i|x)}$. Then the resulting optimal decision regions are given by the following equations. For each $i = 0, 1, ..., M-1$,

$$R_{u=i} = \big\{x : \partial S/\partial p_{opt}(u = i|x) > 0\big\}. \tag{7}$$

Proof: We want to maximize S with respect to $p(u = i|x) \in [0, 1]$. If S is convex in $p(u = i|x)$ over the interval $[0, 1]$ then its maxima occur at the boundary where $p(u = i|x) = 0$ or 1. Therefore for each data point $x \in \mathcal{X}$,

$$p_{\text{opt}}(u = i|x) = 1 \iff \partial S/\partial p_{\text{opt}}(u = i|x) > 0 \tag{8}$$

and

$$p_{\text{opt}}(u = i|x) = 0 \iff \partial S/\partial p_{\text{opt}}(u = i|x) < 0. \tag{9}$$

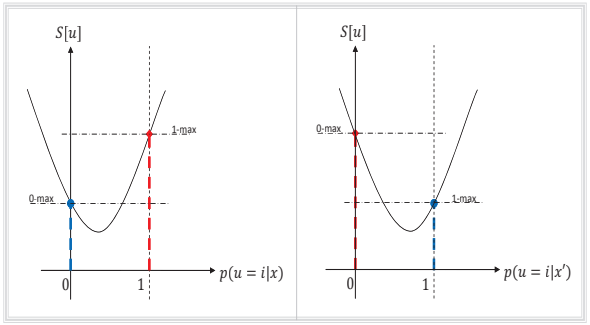

Fig. 2. Visualization of S (under H_i, x is accepted and x' is rejected)

From (6), the two regions defined by x satisfying $p_{\text{opt}}(u = i|x) = 0$ and x satisfying $p_{\text{opt}}(u = i|x) = 1$ are measure-wise complementary. Therefore, equation (8) covers both cases, and is equivalent to

$$p_{\text{opt}}(u = i|x) = I_{R_{u=i}}(x), \tag{10}$$

where $R_{u=i} = \{x : \partial S/\partial p_{\text{opt}}(u = i|x) > 0\}$. ■

In Sections III and IV we will use similar expressions to determine optimal decision regions with the probability of detection and KL distance as objective functions.

Notice that (10) is an implicit equation in $p_{\text{opt}}(u = i|x)$ since the region $R_{u=i}$ also depends on $p_{\text{opt}}(u = i|x)$. Therefore we must proceed to solve the system of equations $\{p_{\text{opt}}(u = i|x) = I_{R_{u=i}}(x) : \; i = 0, 1, ..., M\}$ to explicitly determine the decision regions. In the case of distributed networks of sensors where more than one set of local decision rules are involved, the resulting system of equations is analytically intractable and one often has to resort to numerical computation.

Recall that the set of data points satisfying (6) must be null with respect to the probability measure. Otherwise, we have a randomized decision rule $p_{\text{opt}}(u = i|x) = I_{R_{u=i}}(x) + \sum_j \rho_{ij} \; I_{C_{u=j}}(x)$, where $\rho_{ij} \in [0, 1]$ are arbitrary (i.e., free) coefficients but which must be consistent with all constraints on the optimization problem. With this randomization, the proposition holds for discrete random variables as well.

For ease of presentation, in the rest of the paper we only consider the case of $M = 2$, i.e., binary hypotheses with binary sensor output.

III. The fixed sample size Neyman-Pearson test

We will now consider a fixed sample size NP test. From Fig 1 (a) the decisions (v, w) for the YX process are based on the observations (x, y), which satisfy the conditional independence relation $p_i(x, y) = p_i(x)p_i(y)$. Similarly Fig 1 (b) describes the XYX process with decisions (u, v, w).

The objective function for the NP test is the probability of detection for which the probability of false alarm must not exceed a certain fixed value α. The optimization problem is

$$\text{maximize} \quad P_d[w] = p_1(w=1)|_{p_0(w=1)\leq\alpha} \tag{11}$$

The Lagrangian for such a problem is given by

$$L([w],\lambda) = p_1(w=1) + \lambda\ (\alpha - p_0(w=1)), \quad 0 \leq \lambda < \infty, \tag{12}$$

where $p_i(w) = \sum_{x,y,v} p(w|x,v)p(v|y)\ p_i(x)p_i(y)$ for the YX process, and $p_i(w) = \sum_{x,v,y,u} p(w|x,v)p(v|y,u)p(u|x)\ p_i(x)p_i(y)$ for the XYX process. We must minimize $L([w],\lambda)$ over λ in the dual problem, or for unbiased tests, we can equivalently solve $p_0(w=1) = \alpha$ for $\lambda(\alpha)$ since in that case both $p_1(w=1)$ and $p_0(w=1)$ are increasing in $p(w=1|x,v),\ p(v=1|y,u)$, and $p(u=1|x)$ at the optimal point.

Notice that the Lagrangian (12) has essentially the same form as a Bayes risk function such as the probability of error used in [3]. Indeed, a paralleled analysis of the decision regions due to our Lagrangian (12) shows that for a constant signal s in WGN

$$x = s + z_1, \qquad y = s + z_2, \tag{13}$$

where $z_1 \sim N(0,\sigma_x^2),\ \ z_2 \sim N(0,\sigma_y^2)$, under hypotheses

$$H_0:\ s=0, \qquad H_1:\ s=1, \tag{14}$$

the XYX process performs better than the YX process.

In this example, we have assumed for simplicity that each sensor's observation consists of only one real sample; i.e., we have the joint sample $(x,y),\ \ x,y \in \mathbb{R} = \mathcal{X} = \mathcal{Y}$. Fig. 3. shows the dependence of the probability of detection on σ_x when σ_y is fixed. The corresponding false alarm probability $P_f = \alpha = 0.2$.

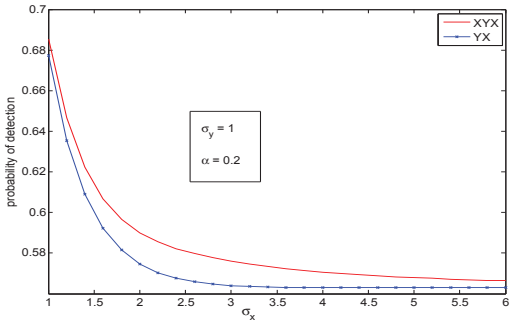

Fig. 3. Performance of XYX and YX processes

IV. The Asymptotic Neyman-Pearson Test

A. Derivation of the KL distance

Let us denote any sequence $s_1,...,s_n$ by s^n, and again consider the Lagrangian (12) of the fixed size NP test. We use the two-way XYX process here for illustration but the basic ideas apply to any number of iterative feedback steps.

Let there be n observation samples or n "processing blocks" $(x_1,y_1),...,(x_n,y_n)$. We adopt a sample-by-sample scenario in which the two sensors first go through a sequence of n iid XY decision processes $(u_1,v_1),...,(u_n,v_n)$, where $u_k = \gamma_k(x_k),\ \ v_k = \delta_k(y_k,u_k)$. In this particular scenario we in theory require that sensor X stores the sequence $v^n = v_1,...,v_n$ and then combines it with its own full n-sample observation x^n to make a final decision $w = \rho(x^n,v^n)$. Therefore,

$$p_1(w) = \sum_{x^n,v^n} p(w|x^n,v^n)p_1(x^n,v^n).$$

By Proposition II.1 the decision region for w is given by

$$R_{w=1} = \underset{\substack{R\subset\mathcal{X}^n\times\{0,1\}^n,\\ p_0(R)\leq\alpha}}{\arg\max}\ p_1(R) = \left\{(x^n,v^n) : \frac{p_1(x^n,v^n)}{p_0(x^n,v^n)} > \lambda\right\}, \tag{15}$$

where λ is the Lagrange multiplier as in (12). We could proceed by Proposition II.1 to find decision regions for v^n.

However, we are interested mainly in the asymptotic (i.e., large n) nature of the iid decision processes $\{(u_k,v_k)\}$. Recall that by the weak law of large numbers,

$$\frac{1}{n}\log\frac{p_0(x^n,v^n)}{p_1(x^n,v^n)} = \frac{1}{n}\sum_{k=1}^{n}\log\frac{p_0(x_k,v_k)}{p_1(x_k,v_k)}$$
$$\overset{n\to\infty}{\longrightarrow}\ E_{p_0(x,v)}\log\frac{p_0(x,v)}{p_1(x,v)} = D(p_0(x,v)\|p_1(x,v)),$$

where $v=\delta(y,u),\ \ u=\gamma(x)$. By the Chernoff-Stein Lemma, [4], the test with decision region

$$R_n^\alpha(p_0|p_1) = \left\{(x^n,v^n) : D(p_0\|p_1) - \alpha \leq \frac{1}{n}\log\frac{p_0(x^n,v^n)}{p_1(x^n,v^n)} \leq D(p_0\|p_1) + \alpha\right\}$$

is asymptotically optimal with error exponent

$$-\lim_{n\to\infty}\frac{1}{n}\log p_1\big((x^n,v^n)\notin R_n^\alpha(p_0|p_1)\big) = D(p_0(x,v)\|p_1(x,v)), \tag{16}$$

which is the KL distance that we will now use as our objective function for the asymptotic performance of the NP test.

We show in the following that with the KL distance as objective, interactive fusion provides no improvement over one-way tandem fusion.

B. One-step YX process

Consider a distributed network in which two sensors X and Y make independent observations x and y. Y then makes a decision $v = \gamma(y)$ and passes v to X (See Fig 1. (a)). The optimal decision v is chosen so as to maximize the KL distance

$$K[x,v] = D\big(p_0(x,v)\|p_1(x,v)\big) \tag{17}$$

at sensor X.

Since $p_i(x,v) = p_i(x)p_i(v)$, we have

$$K[x,v] = D\big(p_0(x)\|p_1(x)\big) + \sum_v p_0(v)\ \log\big(p_0(v)/p_1(v)\big) \tag{18}$$

where $p_i(v) = \sum_y p(v|y)p_i(y)$.

Theorem IV.1. *The optimal decision region at Y is given by*

$$R_{v=1} = \left\{y : \frac{\partial K[x,v]}{\partial p_{\rho y'}(v=1|y)} > 0\right\} = \left\{y : \frac{p_1(y)}{p_0(y)} > \lambda\right\}, \tag{19}$$
$$\lambda = \left(\log\frac{\beta(1-\alpha)}{\alpha(1-\beta)}\right)\Big/\left(\frac{\beta-\alpha}{\beta(1-\beta)}\right),$$

where $\alpha = P_0\left[p_1(y)/p_0(y) > \lambda\right]$ *and* $\beta = P_1\left[p_1(y)/p_0(y) > \lambda\right]$. *Thus* α, β, *and* λ *are coupled with each other.*

Due to limited space we omit the details of the proof. The key observation is that the KL distance is convex in $p(v|y)$ and so we can apply Proposition II.1. In addition, because of the constraint $p(v=1|y) + p(v=0|y) = 1$, the derivative of $K[x,v] \equiv K^{\text{YX}}$ requires the differentiation rule

$$\partial p(v|y)/\partial p(v'|y') = (-1)^{v-v'}\delta_{yy'}. \tag{20}$$

which holds for binary decisions. Notice that the decision region defined in (19) and the threshold of the likelihood ratio of y are coupled with each other. Iterative process is thus needed for finding the optimal λ and the associated α and β.

The maximum KL distance is given by

$$K_{\max}^{\mathrm{YX}} = K[x] + \alpha^* \log \frac{\alpha^*}{\beta^*} + (1-\alpha^*) \log \frac{1-\alpha^*}{1-\beta^*}, \tag{21}$$

where $K[x] = D(p_0(x)\|p_1(x))$ and α^* and β^* are the values of α and β at convergence.

Consider again the hypothesis test described in (14). By Theorem IV.1, the optimal decision region at Y is $R_{v=1} = \{y:\ y > t = \sigma_y^2 \log \lambda(t) + \frac{1}{2}\}$, where

$$\lambda(t) = \left(\log \frac{Q\left(\frac{t-1}{\sigma_y}\right)\left(1-Q\left(\frac{t}{\sigma_y}\right)\right)}{Q\left(\frac{t}{\sigma_y}\right)\left(1-Q\left(\frac{t-1}{\sigma_y}\right)\right)} \right) \Bigg/ \left(\frac{Q\left(\frac{t-1}{\sigma_y}\right) - Q\left(\frac{t}{\sigma_y}\right)}{Q\left(\frac{t-1}{\sigma_y}\right)\left(1-Q\left(\frac{t-1}{\sigma_y}\right)\right)} \right).$$

The corresponding maximum KL distance $K_{\max}[x,v]$ is

$$\frac{1}{2\sigma_x^2} + Q\left(\frac{t^*}{\sigma_y}\right) \log \frac{Q\left(\frac{t^*}{\sigma_y}\right)}{Q\left(\frac{t^*-1}{\sigma_y}\right)} + \left(1-Q\left(\frac{t^*}{\sigma_y}\right)\right) \log \frac{1-Q\left(\frac{t^*}{\sigma_y}\right)}{1-Q\left(\frac{t^*-1}{\sigma_y}\right)},$$

where t^* is the threshold t at convergence.

C. Two-step XYX process

Let two independent sensors X and Y make observations x and y as in Section IV-B. X then makes a decision $u = \gamma(x)$ and passes u onto Y. Y further makes a decision $v = \delta(y,u)$ and sends it back to X (See Fig. 1. (b)). The optimal decisions u and v are chosen so as to maximize the KL distance $K[x,v] \equiv K^{\mathrm{XYX}}$ in the final step at X. The KL distance can be written as

$$\begin{aligned} K^{\mathrm{XYX}} &= D\big(p_0(x,v)\|p_1(x,v)\big) \\ &= D(p_0(x)\|p_1(x)) + \sum_x p_0(x) \sum_v p_0(v|x)\ \log \frac{p_0(v|x)}{p_1(v|x)}, \end{aligned}$$

where $p_i(v|x) = \sum_u p(u|x) \sum_y p(v|y,u) p_i(y)$.

Theorem IV.2. *The optimal decision region at sensor X is given by*

$$R_{u=1} = \left\{x : \tfrac{\partial K[x,v]}{\partial p(u=1|x)} > 0\right\} = \left\{x : \textstyle\sum_u I_{R_u}(x) A_u B_u > 0\right\} \tag{22}$$

$$A_u = \tfrac{\beta_u^{(2)} - \alpha_u^{(2)}}{\beta_u^{(2)}(1-\beta_u^{(2)})}, \qquad B_u = \tfrac{\beta_1^{(2)} - \beta_0^{(2)}}{\alpha_1^{(2)} - \alpha_0^{(2)}} - \lambda_u^{(2)},$$

and the optimal decision regions at sensor Y are given by

$$R_{v=1|u} = \left\{y:\ \tfrac{\partial K[x,v]}{\partial p(v=1|y,u)} > 0\right\} = \left\{y : \tfrac{p_1(y)}{p_0(y)} > \lambda_u^{(2)}\right\}, \tag{23}$$

$$\lambda_u^{(2)} = \left(\log \tfrac{\beta_u^{(2)}(1-\alpha_u^{(2)})}{\alpha_u^{(2)}(1-\beta_u^{(2)})} \right) \Big/ \left(\tfrac{\beta_u^{(2)} - \alpha_u^{(2)}}{\beta_u^{(2)}(1-\beta_u^{(2)})} \right),$$

where $\alpha_u^{(2)} = P_0(R_{v=1|u})$, and $\beta_u^{(2)} = P_1(R_{v=1|u})$.

We again omit the details of the proof but note that the proof requires Proposition II.1, the modified differentiation rule (20), and the following lemma whose proof follows immediately from the properties of a partition of a given set.

Lemma IV.1. *Let $\{R_i : i = 1,...,m\}$ be any partition of the data space $\mathcal{X} = \{x\}$. Then for any continuous multivariate function f, and for each $x \in \mathcal{X}$,*

$$f\left(\sum_{i=1}^m I_{R_i}(x) a_{i1}, \sum_{i=1}^m I_{R_i}(x) a_{i2}, ...\right) = \sum_{i=0}^m I_{R_i}(x)\ f(a_{i1}, a_{i2}, ...), \tag{24}$$

where a_{ij} for all i and j are numbers.

Using (24), the KL distance can be expressed as

$$\begin{aligned} K^{\mathrm{XYX}} &= K[x] + \sum_{u,v} P_0(R_u)\ P_0(R_{v|u}) \log \frac{P_0(R_{v|u})}{P_1(R_{v|u})} \\ &= K[x] + \alpha^{(1)} f(\alpha_1^{(2)}, \beta_1^{(2)}) + (1-\alpha^{(1)}) f(\alpha_0^{(2)}, \beta_0^{(2)}), \end{aligned}$$

where $K[x] = D(p_0(x)\|p_1(x))$, $\alpha^{(1)}$ is a constant independent of the thresholds, and $f(\alpha,\beta) = \alpha \log \frac{\alpha}{\beta} + (1-\alpha) \log \frac{1-\alpha}{1-\beta}$.

Since the two threshold-dependent terms in K^{XYX} are decoupled with respect to threshold dependence, at convergence $\lambda_0^{(2)} = \lambda_1^{(2)}$. In other words, the maximization of these two terms can be carried out independently as each corresponds to different values of u in the initial feedback from X. Therefore $K_{\max}^{\mathrm{XYX}} = K_{\max}^{\mathrm{YX}}$, and consequently, XYX achieves the same performance as YX. This observation holds for any probability distribution. The results for the constant signal in WGN under hypotheses (14) are shown in Fig. 4, where the KL distances of YX and XYX processes coincide with each other. Similarly, the KL distances of XY and YXY also coincide with other. An interesting observation from the plot is that the two sets of curves, each corresponding to making final decision at different nodes, intercept each other at the point when $\sigma_x = \sigma_y = 1$. Thus for this example, it is always better to make the final decision at the sensor with better signal to noise ratio.

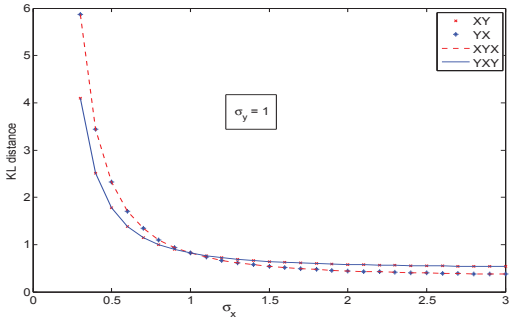

Fig. 4. Comparison of KL distances of one-way tandem fusion and interactive fusion with different communication directions. For this plot, we fix $\sigma_y = 1$ throughout while varying σ_x.

D. Discussion

We have established that two-way feedback is not asymptotically better than one-way tandem fusion. The result can be generalized to interactive fusion networks where multiple iterations between the two sensors may be involved. Careful analysis of the N-step feedback process shows that whenever a sensor's data is explicitly summed over in the KL distance, the decision process becomes independent of that particular sensor's data. Since repetition of the decision process involving only one sensor's data cannot improve performance, it follows that iterative feedback processing will not improve performance with respect to the KL distance. Notice, however, that the N-step iteration process assumes that the sensors have limited memory. Therefore at the sth step, sensor $Z \in \{X,Y\}$ uses only its observation $z \in \{x,y\}$ and the preceding decision u_{s-1} from the previous step at the other sensor, to make its current decision $u_s = \gamma_s(z, u_{s-1})$. The result is no longer valid if all previous feedback bits can be collectively processed.

V. Conclusion

We have considered a class of decision theory problems that involve convex objective functions and used it to study two-sensor tandem fusion networks with conditionally independent observations. It was established that while feedback improves the performance of fixed sample size NP test, it does not affect the asymptotic performance as characterized by the error exponent of type II error.

REFERENCES

[1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1997.

[2] J. D. Papastavrow and M. Athans, "On optimal distributed decision architectures in a hypothesis testing environment," *IEEE Trans. Autom. Control*, vol. 37, no. 8, pp. 1154-1169, Aug. 1992.

[3] S. Zhu, E. Akofor, and B. Chen, "Interactive distributed detection with conditionally independent observations," *IEEE WCNC, Shanghai China,* April 2013.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, Wiley, 2nd edition, 2006.

[5] S. Alhakeem and P. K. Varshney, "Decentralized Bayesian detection with feedback," *IEEE Trans. Sys., Man, Cyb. - A: Sys. Hum.*, Vol. 26, No. 4, pp. 503 - 513, July 1996.

[6] W. P. Tay and J. N. Tsitsiklis. "The value of feedback for decentralized detection in large sensor networks." *Proceedings of the 6th International Symposium on Wireless and Pervasive Computing (ISWPC)*, 2011. 1-6.

[7] H. Shalaby and A. Papamarcou, "A note on the asymptotics of distributed detection with feedback," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 633-640, Mar. 1993.

[8] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Comm.*, vol. 41, no. 4, pp. 550-558, April. 1993.