DATA VERSUS DECISION FUSION IN WIRELESS SENSOR NETWORKS

Ashwin D'Costa and Akbar M. Sayeed

Electrical and Computer Engineering
University of Wisconsin-Madison
dcosta@cae.wisc.edu, akbar@engr.wisc.edu

ABSTRACT

Sensor networks provide virtual snapshots of the physical world via distributed wireless nodes that can sense in different modalities, such as acoustic and seismic. Classification of objects moving through the sensor field is an important application that requires collaborative signal processing (CSP) between nodes. Given the limited resources of nodes, a key constraint is to exchange the least amount of information between them to achieve desired performance. Two main forms of CSP are possible. Data fusion – exchange of low dimensional feature vectors – is needed between correlated nodes, in general, for optimal performance. Decision fusion – exchange of likelihood values – is sufficient between independent nodes. Decision fusion is generally preferable due to its lower communication burden. We study CSP of multiple node measurements, each modeled as a Gaussian signal vector (corresponding to the target class) corrupted by additive white Gaussian noise. The measurements are partitioned into groups. The signal components within each group are perfectly correlated whereas they vary independently between groups. Three classifiers are compared: the optimal maximum likelihood classifier, a data averaging classifier that treats all measurements as correlated, and a decision fusion classifier that treats them all as independent. Analytical and numerical results based on real data are provided to compare the performance of the three CSP classifiers. Our results indicate that the sub-optimal decision fusion classifier, that is most attractive in the context of sensor networks, is also a robust choice from a decision theoretic viewpoint.

1. INTRODUCTION

Wireless sensor networks are an emerging technology for monitoring the physical world with a densely distributed network of wireless nodes. Each node has limited communication and computation ability and can sense the environment in a variety of modalities, such as acoustic, seismic, and infra red [1, 2, 3]. A wide variety of applications are being envisioned for sensor networks, including disaster relief, border monitoring, condition-based machine monitoring, and surveillance in battlefield scenarios. Detection and classification of objects moving through the sensor field is an important task in many envisioned applications. Exchange of sensor information between different nodes in the vicinity of the object is necessary for reliable execution of such tasks due to a variety of reasons, including limited (local) information gathered by each node, variability in operating conditions, and node failure. Consequently, development of theory and methods for collaborative signal processing (CSP) of the data collected by different nodes is a key research area for realizing the vision of sensor networks.

The CSP algorithms have to be developed under the constraints imposed by the limited communication and computational abilities of the nodes as well as their finite battery life. A key goal of CSP algorithms in sensor networks is to exchange the least amount of data between nodes to attain a desired level of performance. In this paper, with the above goal in mind, we investigate CSP algorithms for single target classification based on multiple acoustic measurements at different nodes. The numerical results presented here are based on real data collected in the DARPA SensIT program.

There are two main forms of information exchange between nodes dictated by the statistics of measured signals. If two nodes yield correlated measurements, *data fusion* is needed, in general, for optimal performance – exchange of (low-dimensional) feature vectors that yield sufficient information for desired classification performance. On the other hand, if two nodes yield independent measurements, *decision fusion* is sufficient – exchange of likelihood values (scalars) computed from individual measurements. In general, the measurements would exhibit a mixture of correlated and independent components and would require a combination of data and decision fusion between nodes. In the context of sensor networks, decision fusion is clearly the more attractive choice. First, it imposes a significantly lower communication burden on the network, compared to data fusion, since only scalars are transmitted to a manager node (where the final processing is done) [3]. Second, it also imposes a lower computational burden compared to data fusion since lower dimensional data has to be processed by the joint classifier at the manager node.

In this paper, we investigate the design of CSP classifiers and assess their performance in an idealized abstraction of measurements from multiple nodes. We consider $K = Gn_G$ measurements corresponding to a particular event. The K measurements are split into G groups with n_G measurements in each group. The signal component in the n_G measurements in a particular group is identical (perfectly correlated), but it varies independently from group to group. We compare the performance of three classifiers: 1) the optimal maximum likelihood (ML) classifier, 2) a sub-optimal (decision fusion) classifier that treats all the measurements as independent, and 3) a sub-optimal (data averaging) classifier that treats all the measurements as perfectly correlated. Our results indicate that the decision fusion classifier is remarkably robust to the true statistical correlation between measurements. Thus, the decision fusion classifier, that is the most attractive choice in view of the computational and communication constraints imposed by the network, is also a robust choice from a decision theoretic viewpoint.

This work was supported in part by DARPA SensIT program under Grant F30602-00-2-0555.

2. CSP CLASSIFIERS FOR MULTIPLE MEASUREMENTS

In this section, we briefly describe the structure of classifiers based on multiple measurements. We consider Gaussian classifiers which assume that the underlying data has complex circular Gaussian statistics. The notation $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\mathrm{E}[\boldsymbol{x}] = \boldsymbol{\mu}$ and $\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^H] = \boldsymbol{\Sigma}$ and $\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^T] = \boldsymbol{0}$ (circular assumption). We first discuss the classifier structure for a single measurement and then generalize it to multiple measurements.

2.1. Single Measurement Classifier

Consider M target classes. Let $\boldsymbol{x}$ denote a complex-valued N-dimensional feature vector corresponding to a detected event. Under hypothesis $j = 1, \cdots, M$ (corresponding to j-th target class), $\boldsymbol{x}$ is modeled as

$$H_j \; : \; \boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n} \;\; , j = 1, \cdots, M, \tag{1}$$

where $\boldsymbol{s} \sim \mathcal{CN}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ denotes the Gaussian signal component corresponding to the j-th class, and $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ denotes additive white Gaussian noise. A classifier C maps the event feature vector $\boldsymbol{x}$ to one of the target classes. We assume that all classes are equally likely. Thus, the optimal classifier is the maximum likelihood (ML) classifier which takes the form

$$C(\boldsymbol{x}) = \arg \max_{j \in \{1, \cdots, M\}} p_j(\boldsymbol{x}) \tag{2}$$

where $p_j(\boldsymbol{x})$ denotes the likelihood function for j-th class which takes the following form under the complex Gaussian assumption

$$p_j(\boldsymbol{x}) = \frac{1}{\pi^N |\boldsymbol{\Sigma}_j + \boldsymbol{I}|} e^{-(\boldsymbol{x} - \boldsymbol{\mu}_j)^H (\boldsymbol{\Sigma}_j + \boldsymbol{I})^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_j)}. \tag{3}$$

In this paper, we assume zero-mean signals so that $\boldsymbol{\mu}_j = \boldsymbol{0}$ for all j and, thus, all information about the targets is contained in the covariance matrices $\boldsymbol{\Sigma}_j$. In practice, $\{\boldsymbol{\Sigma}_j\}$ have to be estimated from available training data. We assume that $\mathrm{tr}(\boldsymbol{\Sigma}_j)$ (signal energy) is the same for all j.

2.2. Multiple Measurement Classifiers

Suppose that we have K measurements (in a given modality), $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K\}$, from different nodes available to us. We are interested in combining these measurements to achieve improved classification performance. Consider the concatenated NK-dimensional feature vector

$$\boldsymbol{x}^{cT} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_K^T] \tag{4}$$

which has the same form as (1) under different hypotheses except for the larger number of dimensions. The noise is still white but the signal correlation matrix under H_j can be partitioned as

$$\boldsymbol{\Sigma}_j^c = \begin{bmatrix} \boldsymbol{\Sigma}_{j,11} & \boldsymbol{\Sigma}_{j,12} & \cdots & \boldsymbol{\Sigma}_{j,1K} \\ \boldsymbol{\Sigma}_{j,21} & \boldsymbol{\Sigma}_{j,22} & \cdots & \boldsymbol{\Sigma}_{j,2K} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{j,K1} & \boldsymbol{\Sigma}_{j,K2} & \cdots & \boldsymbol{\Sigma}_{j,KK} \end{bmatrix} \tag{5}$$

where $\boldsymbol{\Sigma}_{j,kk'} = \mathrm{E}[\boldsymbol{x}_k \boldsymbol{x}_{k'}^H]$ denotes the cross-covariance between the k-th and k'-th measurements. The optimal classifier operates on $\boldsymbol{x}^c$ and takes the form (2) with $p_j(\boldsymbol{x}^c)$ given by (3) by replacing $\boldsymbol{x}$ with $\boldsymbol{x}^c$ and $\boldsymbol{\Sigma}_j$ with $\boldsymbol{\Sigma}_j^c$.

2.3. A Simple Measurement Model

Let $K = G n_G$. Suppose that the signal component of $\boldsymbol{x}^c$ can be partitioned into G groups of n_G measurements each as

$$\boldsymbol{s}^{cT} = [\boldsymbol{s}_1^T, \cdots, \boldsymbol{s}_1^T, \boldsymbol{s}_2^T, \cdots, \boldsymbol{s}_2^T, \cdots, \boldsymbol{s}_G^T, \cdots, \boldsymbol{s}_G^T] \tag{6}$$

where the signal component of the n_G measurements in each group is identical and it varies independently from group to group. That is, $\{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_G\}$ are i.i.d. according to $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_j)$ under H_j. The noise measurements, on the other hand, are independent across all measurements. The above signal model can capture a range of correlation between measurements. For $K = G$ ($n_G = 1$), all the measurements have independent signal components (no correlation), whereas for $K = n_G$ ($G = 1$), all the measurements have identical signal components (maximum correlation).

2.4. Optimum Classifier

There are two sources of classification error: background noise and the inherent statistical variability in the signals captured by $\boldsymbol{\Sigma}_j$'s. The optimal classifier performs signal averaging within each group to reduce the noise variance and statistical averaging over the groups to reduce the inherent signal variations. The optimum classifier operates on the NG dimensional vector

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_G \end{bmatrix} = \begin{bmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_G \end{bmatrix} + \begin{bmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_G \end{bmatrix} = \boldsymbol{s} + \boldsymbol{w} \tag{7}$$

where $\boldsymbol{y}_i$, $i = 1, \cdots, G$, are obtained by averaging the measurements in each group

$$\boldsymbol{y}_i = \frac{1}{n_G} \sum_{j=1}^{n_G} \boldsymbol{x}_{(i-1)G+j} = \boldsymbol{s}_i + \boldsymbol{w}_i. \tag{8}$$

Note that $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}/n_G)$ due to signal averaging and $\boldsymbol{s} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_{G,j})$ under H_j where $\boldsymbol{\Sigma}_{G,j} = \mathrm{diag}(\boldsymbol{\Sigma}_j, \boldsymbol{\Sigma}_j, \cdots, \boldsymbol{\Sigma}_j)$ is an $NG \times NG$ block diagonal matrix with $\{\boldsymbol{\Sigma}_j\}$ on the G diagonal blocks. It can be shown that the optimal classifier takes the form

$$C_{opt}(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_G) = \arg \min_{j = 1, \cdots, M} l_{opt,j}(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_G) \tag{9}$$

where the (negative) log-likelihood function $l_{opt,j}(\boldsymbol{y})$ is given by

$$\begin{aligned} l_{opt,j}(\boldsymbol{y}) &= \log |\boldsymbol{\Sigma}_j + \boldsymbol{I}/n_G| + \frac{1}{G} \sum_{i=1}^{G} \boldsymbol{y}_i^H (\boldsymbol{\Sigma}_j + \boldsymbol{I}/n_G)^{-1} \boldsymbol{y}_i \\ &= \log |\boldsymbol{\Sigma}_j + \boldsymbol{I}/n_G| + \mathrm{tr}((\boldsymbol{\Sigma}_j + \boldsymbol{I}/n_G)^{-1} \hat{\boldsymbol{\Sigma}}_y) \end{aligned} \tag{10}$$

and $\hat{\boldsymbol{\Sigma}}_y = \frac{1}{G} \sum_{i=1}^{G} \boldsymbol{y}_i \boldsymbol{y}_i^H$ is the estimated data correlation matrix of $\{\boldsymbol{y}_i\}$.

It is insightful to consider two limiting cases. First, suppose that $K = n_G$ ($G = 1$) so that all measurements are perfectly correlated. Then, in the limit of large K

$$\lim_{K \to \infty} l_{opt,j}(\boldsymbol{y}) = \log |\boldsymbol{\Sigma}_j| + \boldsymbol{y}_1^H \boldsymbol{\Sigma}_j^{-1} \boldsymbol{y}_1 \tag{11}$$

which shows that noise is completely eliminated and the only remaining source of error is the inherent statistical variation in the

signal component. Now, suppose that $K = G$ ($n_G = 1$) so that all measurements are i.i.d. In the limit of large K we have

$$\lim_{K\to\infty} l_{opt,j}(\boldsymbol{y}) = \log|\boldsymbol{\Sigma}_j + \boldsymbol{I}| + \mathrm{tr}((\boldsymbol{\Sigma}_j + \boldsymbol{I})^{-1}\boldsymbol{\Sigma}_y) \quad (12)$$

where $\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_m + \boldsymbol{I}$ under H_m. In this case, all statistical variation in the signal is removed due to ensemble averaging. However, there is a bias in the estimated data correlation (relative to $\boldsymbol{\Sigma}_j$) due to noise. In this case, it can be shown that perfect classification can be attained in the limit of large K as long as the differential entropies satisfy $D(p_j\|p_m) > 0$ for all $j \neq m$ (p_j is defined in (3)). Note that both data averaging (correlated measurements) and ensemble (decision) averaging (uncorrelated measurements) contribute to improved classifier performance. However, as we will see, ensemble averaging is more critical in the case of stochastic signals.

2.5. Data-Averaging Classifier

As mentioned earlier, for the measurement model of the previous section, we compare the optimal classifier to two sub-optimal classifiers. In this section, we describe the first sub-optimal classifier that treats all measurements as correlated. It operates on the average of all measurements

$$\boldsymbol{y}_{da} = \frac{1}{K}\sum_{i=1}^{K}\boldsymbol{x}_i = \frac{1}{G}\sum_{i=1}^{G}\boldsymbol{y}_i = \boldsymbol{s}_{da} + \boldsymbol{w}_{da} \quad (13)$$

where $\boldsymbol{s}_{da} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_j/G)$ under H_j and $\boldsymbol{w}_{da} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}/K)$ in the measurement model. The data-averaging classifier takes the form

$$\begin{aligned} C_{da}(\boldsymbol{y}_{da}) &= \arg\min_{j=1,\cdots,M} l_{da,j}(\boldsymbol{y}_{da}) \\ l_{da,j}(\boldsymbol{y}_{da}) &= \log|\boldsymbol{\Sigma}_j + \boldsymbol{I}/K| + \boldsymbol{y}_{da}^H(\boldsymbol{\Sigma}_j + \boldsymbol{I}/K)^{-1}\boldsymbol{y}_{da} \end{aligned} \quad (14)$$

Note that C_{opt} and C_{da} are identical for $K = n_G$ in the measurement model. Furthermore, all measurements $\{\boldsymbol{x}_i\}$ have to be communicated to the manager node for the computation of C_{opt} and C_{da}. However, the computational complexity of C_{da} is lower than that of C_{opt}.

2.6. Decision-Fusion Classifier

The decision-fusion classifier treats all measurements as independent. It takes the form

$$\begin{aligned} C_{df}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K) &= \arg\min_{j=1,\cdots,M} l_{df,j}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K) \\ l_{df,j}(\boldsymbol{x}) &= \log|\boldsymbol{\Sigma}_j + \boldsymbol{I}| + \frac{1}{K}\sum_{i=1}^{K}\boldsymbol{x}_i^H(\boldsymbol{\Sigma}_j + \boldsymbol{I})^{-1}\boldsymbol{x}_i \\ &= \log|\boldsymbol{\Sigma}_j + \boldsymbol{I}| + \mathrm{tr}((\boldsymbol{\Sigma}_j + \boldsymbol{I})^{-1}\hat{\boldsymbol{\Sigma}}_x) \end{aligned} \quad (15)$$

where $\hat{\boldsymbol{\Sigma}}_x = \frac{1}{K}\sum_{i=1}^{K}\boldsymbol{x}_i\boldsymbol{x}_i^H$ is the estimated data correlation matrix of $\{\boldsymbol{x}_i\}$. Note that C_{opt} and C_{df} are identical for $K = G$ in the measurement model. Note also from (15) that the M scalars $\{\boldsymbol{x}_i^H(\boldsymbol{\Sigma}_j + \boldsymbol{I})^{-1}\boldsymbol{x}_i\}$ for $j = 1,\cdots,M$ need to be transmitted from the K nodes to the manager node. Thus, C_{df} imposes a much smaller communication (and computational) burden on the network.

2.7. Performance of the Three Classifiers

The optimal classifier does data averaging over the correlated measurements in each group to improve the effective SNR and combines the likelihoods across independent groups to stabilize the inherent variability in the stochastic signal. Comparing (10) and (15) we can note that the decision fusion classifier is very similar to the optimal classifier: the main difference is that since C_{df} does not do data averaging within each group, it encounters a lower effective SNR compared to C_{opt} (evident from $\boldsymbol{x}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_j + \boldsymbol{I})$ in (15) compared to $\boldsymbol{y}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_j + \boldsymbol{I}/n_G)$ in (10)). In fact, it can be shown that the performance of C_{df} can be conservatively approximated by that of C_{opt} by modeling $\boldsymbol{y}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_j + \boldsymbol{I})$ in (10). The data-averaging classifier, on the other hand, operates on $\boldsymbol{y}_{da} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_j/G + \boldsymbol{I}/K)$ and thus captures SNR gain of n_G afforded by correlated measurements within each group but does not exploit the statistical independence across groups to reduce the statistical variation in the signal. Thus, in the limit of large number of uncorrelated measurements, we expect both C_{opt} and C_{df} to exhibit improved performance (perfect classification under certain conditions), but the performance of C_{da} will always be limited by the inherent statistical variation in the signal.

We note that for $M = 2$ the average probability of detection (PD) and the average probability of false alarm (PFA) can be exactly computed for C_{opt} and C_{da} and those for C_{df} can be approximated via C_{opt} as discussed above [4]. For $M > 2$ bounds for PD and PFA can be obtained [4]. This is because both $l_{opt,j}$ and $l_{df,j}$ are weighted sums of NG χ_2^2 random variables, whereas $l_{da,j}$ is a weighted sum of N χ_2^2 random variables. The combining weights are determined in all cases by G, n_G and the eigenvalues of $\boldsymbol{\Sigma}_j$. Thus, standard analysis for performance of diversity signaling over fading channels can be adapted to this problem [5].

3. SIMULATION RESULTS BASED ON REAL DATA

In this section, we first consider and motivate a special scenario in which the covariance matrices for different classes are simultaneously diagonalizable. We then present some numerical results on real data collected as part of the DARPA SensIT program.

3.1. Simultaneously Diagonalizable Classes

We assume that all the covariance matrices share the same eigenfunctions, that is

$$\boldsymbol{\Sigma}_j = \boldsymbol{U}\boldsymbol{\Lambda}_j\boldsymbol{U}^H, j = 1,\cdots,M \quad (16)$$

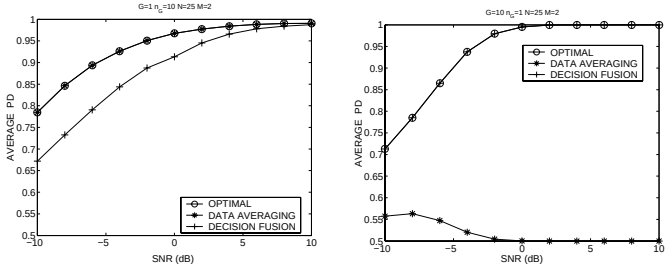
where $\boldsymbol{U}$ represents the matrix of common (orthonormal) eigenvectors for all the classes – the different classes are characterized by the diagonal matrix of eigenvalues $\boldsymbol{\Lambda}_j$. One scenario in which this assumption is approximately valid is when the source signals for different targets can be modeled as stationary processes over the duration of the detected event. In such a case, choosing $\boldsymbol{U}$ as a discrete Fourier transform (DFT) matrix would serve as an approximate set of eigenfunctions. The eigenvalues will then correspond to samples of the associated power spectral densities (PSD's). The numerical results in the next section are based on this assumption and rely on experimental data collected in the SensIT program. Note that given the knowledge of $\boldsymbol{\Lambda}_j$, a realization for the signal from j-th class can be generated as

$$\boldsymbol{s} = \boldsymbol{U}\boldsymbol{\Lambda}_j^{1/2}\boldsymbol{z}\,,\ \boldsymbol{z} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}). \quad (17)$$

3.2. Numerical Results for M=2 Classes

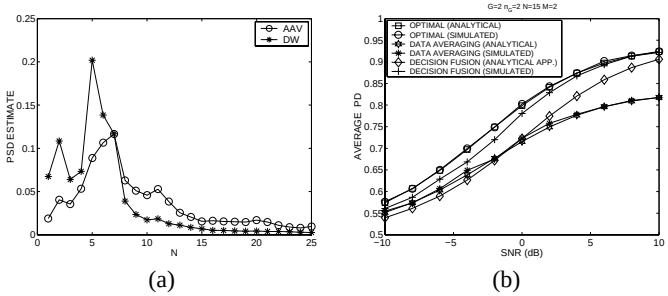
In this section, we present results for classifying two vehicles (Amphibious Assault Vehicle (AAV) and Dragon Wagon (DW)) based on simulated $N = 25$ dimensional measurements from $K = Gn_G = 10$ nodes according to the model in Section 2.3. The eigenvalues (PSD samples) for the two vehicles were estimated from experimental data. The measurements at different nodes were generated using (17) according to the model in Section 2.3. The PD and PFA were estimated using Monte Carlo simulation over 5000 independent events.

Figure 1 plots the PD as a function of SNR for the three classifiers for the two extreme cases: $K = n_G = 10$ (correlated) and $K = G = 10$ (independent). The PFA is simply given by $1 - PD$ for $M = 2$. As expected, C_{opt} and C_{da} perform identically in the first case, whereas C_{opt} and C_{df} perform identically in the second case. However, C_{df} incurs a small loss in performance in the correlated case which diminishes at high SNRs. The performance loss in C_{da} in the independent case is very significant and does not improve with SNR. This is due the classifier mismatch and due to averaging over independent realizations of a zero mean signal – y_{da} is converging to the zero vector due to the law of large numbers. In fact, at high SNR, all events are classified as DW by C_{da} since that class has a larger largest eigenvalue compared to AAV, as evident from Figure 2(a). Figure 2(b) compares the PD of the three classifiers for an intermediate case ($G = n_G = 2$) with $K = 4$ measurements ($N = 15$). Analytically computed PD for C_{opt} and C_{da}, and the conservative approximation for PD of C_{df} are also plotted and agree well with the simulated results.



(a) (b)

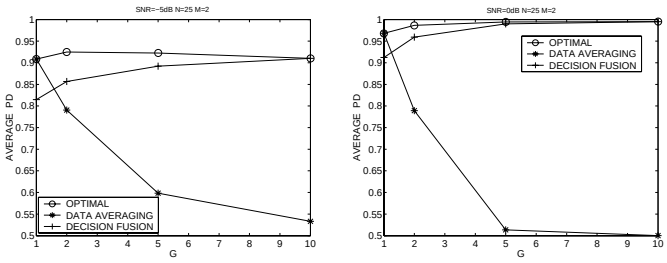
Fig. 1. PD of the three classifiers versus SNR. (a) $K = n_G = 10$ (perfectly correlated measurements). (b). $K = G = 10$ (independent measurements).



(a) (b)

Fig. 2. (a) Covariance matrix eigenvalues (PSD estimates) for AAV and DW. (b) Comparison of simulated and analytically computed PD for $K = 4$, $G = 2$, $n_G = 2$ and $N = 15$.

Figure 3 plots the PD for the three classifiers as function of G ($K = 10$) for two different SNRs. It is evident that C_{df} closely approximates C_{opt} whereas C_{da} incurs a large loss when $K \neq n_G$.



(a) (b)

Fig. 3. Comparison of PD of the three classifiers for varying values of G ($K = 10$). (a) SNR = -5dB. (b) SNR = 0dB.

4. CONCLUSIONS

We have taken a first step in addressing the problem of how much information should be exchanged between nodes for distributed decision making in sensor networks. Our analysis is based on modeling the source signal as a stationary Gaussian process. In general, measurements from multiple nodes will provide a mixture of correlated and uncorrelated information about the source signal. The optimal classifier exploits the correlated measurements to improve the SNR and the independent measurements to stablize the inherent statistical variability in the signal. Both effects are important for improving classifier performance. However, for stochastic signals, the fusion of independent measurements is most significant. In this context, our results demonstrate that the simple suboptimal decision fusion classifier, that treats all measurements as independent, is an attractive choice given the computational and communications constraints in a sensor network. Compared to the optimal classifier, the decision fusion classifier fully exploits the independent measurements but incurs an effective SNR loss depending on the fraction of the measurements that are correlated. However, if the source signal exhibits fewer degrees of freedom (lower-rank covariance matrix), data averaging to improve SNR might become more important. In ongoing research we are developing a framework for near-optimal fusion techniques that are best matched to the constraints of sensor networks.

5. REFERENCES

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor network," in *Proc. ICASSP'01*, 2001, pp. 2675–2678.

[2] "Special issue on collaborative signal and information processing in microsensor networks," in *IEEE Signal Processing Magazine*. (S. Kumar and F. Zhao and D. Shepherd (eds.), March 2002.

[3] D. Li, K. Wong, Y. Hu, and A. Sayeed, "Detection, classification, tracking of targets in micro-sensor networks," in *IEEE Signal Processing Magazine*, March 2002, pp. 17–29.

[4] A. D'Costa and A. Sayeed, "Collaborative signal processing for distributed decision making in sensor networks," *in preparation*, 2003.

[5] J. G. Proakis, *Digitial Communications*, McGraw Hill, New York, 3rd edition, 1995.