2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Classifier Design for Decentralised Sensing with Digital Communication

Gauri Garg[1], Harshad Khadilkar[2], Ankur A. Kulkarni[1], and Aditya A. Paranjape[2]

***Abstract*— We consider the problem of classifying the operating mode of a plant, using distributed sensors and a digital channel. The abstract problem is formulated using simplifications, where the plant only has two modes, the sensors have independent and identically distributed (but possibly mode-dependent) measurement noise, and a noise-less digital communication channel. The objective is to design a combined distributed digitisation (quantisation) and centralised classification strategy that maximises accuracy while observing n messages, each of which can take k unique values. Even in this simplistic scenario, our analysis shows that (i) the optimal decision boundaries even in the fully observable (analog) case depend strongly on the assumptions about measurement noise, (ii) as a result, the classification strategy selection is non-trivial, and (iii) the distributed quantisation algorithm design also has a strong influence on the final classification accuracy. We support the analytical arguments by empirical simulation experiments.**

I. INTRODUCTION

An important class of decentralized estimation problems involves the classification of the operating mode of a remote plant using a swarm of sensors. For example, one may be interested in monitoring the state of a boiler by tracking temperature measurements from a set of sensors placed around it. We are concerned with the setting where the state of the plant could be in one of two classes but each sensor can only receive independent partial observations of it. Not unlike the men in the parable of the *blind men and the elephant*, the sensors must collectively estimate the true state of the plant from these partial and potentially noisy observations. Problems of this form have been studied in the distributed estimation literature, where results take system theoretic forms [1], [2], or take the form of multisensor fusion [3].

However, prior studies working with such scenarios do not seem to account for a fundamental property of such systems: the existence of limited-bandwidth digital channels for transmitting information from the sensors to the classifier. In this paper, we consider the problem of maximising classification accuracy for a centralised classifier using digital information transmitted by a set of sensors, each of which receives noisy measurements from the plant. We show that apart from the aggregation strategy, the mapping from continuous noisy measurements to digitally transmitted (quantised) values plays a key role in the final accuracy of the estimator.

[1]Gauri Garg and Ankur A. Kulkarni are with the Indian Institute of Technology Bombay, Mumbai, India [190010023,kulkarni.ankur]@iitb.ac.in

[2]Harshad Khadilkar and Aditya Paranjape are with TCS Research, India [harshad.khadilkar,aditya.paranjape]@tcs.com

We assume that the state (scalar) of the plant is distributed according to a Gaussian distribution whose mean and standard deviation depend on the operating mode. In the symmetric case, the standard deviations are both equal whereby the modes are distinguished by the corresponding means. The outputs of the individual sensors are combined by an aggregator. The end-to-end performance of the system depends on three factors: the channel bandwidth available (i.e., number of quantization levels), the encoding policy used by the sensors, and the aggregation policy.

In this symmetric case, the ideal decision boundary separating the two modes is a hyperplane in $\mathbb{R}^n$, where n is the number of sensors. However with digital communication or, equivalently, quantization we find that the problem takes on a more complicated combinatorial character. In general the optimal quantization depends on $\frac{\Delta\mu}{\sigma}$ where $\Delta\mu$ is the difference in the means of the modes and σ is the common standard deviation. This implies, in particular, that uniformly spaced quantization is not necessarily optimal.

In the asymmetric case, the ideal 'decision boundary' is the surface of a sphere in $\mathbb{R}^n$. The encoding policy of sensors now induces non-contiguous regions for each signal value. This implies that with quantization, the posterior distribution of the sensors is not a sufficient statistic for the policy of the sensors. The optimal signal sent by the sensors must depend on the posterior probability of the mode *and* the data itself.

One can think of quantization as a form of classification where the preimages of the quantization signals constitute classes. The non-contiguity described above implies that the classifier employed by these sensors must take a complex multimodal form. Simple function approximators such sigmoids used in logistic regression may not be rich enough for this task.

Learning the optimal combination of encoders and aggregators appears to be an interesting and hard problem. Quantized communication for decentralized estimation is also used in many other settings. For example e-commerce platforms seek quantized feedback (e.g., star ratings) from buyers about their experience with an item, and present to prospective buyers an aggregated rating of the item. One may think of this as a system of decentralized estimation of the quality of the item. What should be number of star rating levels, and should the feedback be aggregated? Our study is a beginning in this direction.

A. Literature review

Our work in this paper overlaps somewhat with the area of *multi-sensor data fusion*. The latter concerns methods for

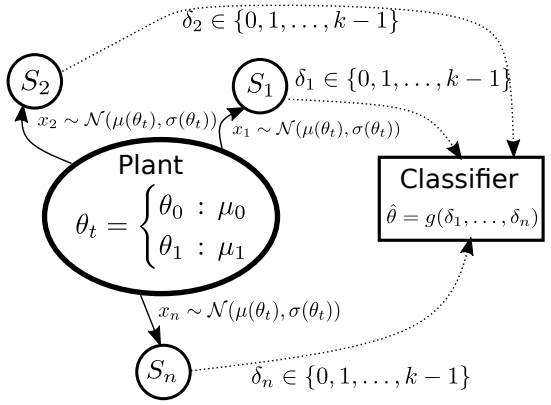

Fig. 1. Schematic of the problem setup. The plant operates in one of two modes θ_0 or θ_1 in every time step t. Each mode corresponds to a scalar state μ_0 or μ_1 respectively. Sensors located in the vicinity measure this state with an i.i.d. Gaussian noise of standard deviation σ_0 or σ_1 corresponding to the two modes. Each sensor has access to a digital communication channel with k unique values per message to a centralised classifier.

combining data obtained from multiple sources of information. These problems are usually handled within a statistical framework, particularly the Bayesian framework [3]. Several combination strategies such as average, majority, Borda count, and so on have been explored, and the right one is usually chosen heuristically. As mentioned in the introduction, the aspect of quantization or digital communication that forms the main point of interest in paper has, to the best of our knowledge, not been studied. Another closely related problem is that of *ensemble learning*. Here the problem is to classify an unknown object optimally to one of K classes, using an ensemble of m classifiers. Ensemble learning however does not involve the aspect of decentralization which is key to our paper.

Digital communication is of relevance in wireless sensor networks used in commercial and military applications [4]. The common issues addressed pertain to sensor management for energy efficiency – sensor tasking and control, tracking and localization, distributed databases, and communication protocols. There are also works such as [5] [6] [7] [8], which address quantization, but do so in a naive way. For example, [8] proposes an energy efficient counting rule for distributed estimation by ordering sensor transmissions in wireless sensor networks. Moreover, these works are not concerned with the classification setting we consider and are primarily concerned with communicating a fixed remote state. The incomplete information arising out of the overlap of two classes is a source of significant difficulty in our problem.

There are also works that concern malicious behaviour of sensors. For example [9] considers the problem of decentralized detection in wireless sensor networks in the presence of one or more classes of misbehaving nodes. Perhaps closest to our work is the lines of work in decentralized detection such as [10], [11], [12]. These works study sensor encoding and transmission strategies for distributed detection in wireless sensor networks and analyze the impact of the number of sensor measurements, and associated network power consumption, on detection performance. The communication channels here are usually assumed to be continuous, whereby once again, the complications arising out of quantization do not occur.

In the control theoretic literature, there is a significant literature on remote state estimation, e.g., [1], [13], some of which also involve quantization [2]. However, to the best of our knowledge, the decentralized classification problem we consider has not been studied.

B. Organization

The paper is organized as follows. In Section II we describe and formulate the problem. In Section III we present an analytical analysis of our problem. Section IV contains empirical results. We conclude in Section V.

II. PROBLEM FORMULATION

Consider the setup shown in Fig. 1. For simplicity, we assume that the plant has only two modes, each of which corresponds to a scalar measurable state μ_0 and μ_1 respectively. A set of n sensors $S_1, \ldots, S_n$ is placed in the vicinity of the plant. In each reading t, a sensor S_i receives a noisy measurement $x_i \sim \mathcal{N}(\mu_t, \sigma_t)$ (similar to the assumption used in Kalman filters [14]), where $\mu_t \in \{\mu_0, \mu_1\}$ is the mode of the plant during reading t, and σ_0, σ_1 are the standard deviations associated with the two modes (denoted by θ_0 and θ_1 respectively). We assume that the measurements x_i are i.i.d. conditioned on the mean μ_t. Each sensor has a access to a mapping function $f_i : \mathbb{R} \to \{0, 1, \ldots, k-1\}$, from the real-valued measurements x_i to a $k-$level digital signal. We define $\delta_i = f_i(x_i)$, and note that f_i may be preprogrammed or trained (learned), and could be unique to each sensor or identical for all sensors.

The digital signals δ_i are transmitted to a classifier, which applies an aggregator $g : \{0, \ldots, k-1\}^n \to \{0, 1\}$ to produce a binary class output. An output of $\hat{\theta} = 1$ corresponds to a decision that the plant is operating in mode θ_1, and vice versa. The goal is to find the combination of mapping f_i and aggregator g that maximises the balanced classification accuracy on test data, given by

$$J = \max_{f_i, g} \mathbb{E}\left[(\theta = \theta_0 \wedge \hat{\theta} = 0) \ \vee \ (\theta = \theta_1 \wedge \hat{\theta} = 1)\right].$$

III. ANALYSIS

Recalling that k denotes the number of bits conveyed by the sensors to the aggregator, the problem of aggregator design lies between two extremes: $k = 1$ (binary classifiers) and $k = \infty$ (classifiers which can transmit information with arbitrarily high fidelity). If the means μ_0, μ_1 are known, then aggregator design is akin to a hypothesis test. In particular, for finite k, the aggregator performs a hypothesis test on the probability distribution of the k-bit outputs y_j arising from the classifiers f_j. Thus, we surmise that the optimal aggregator must take the form of a likelihood ratio test. We begin with an analysis of a centralised aggregator that can access the raw readings, followed by application of the digital communication constraints.

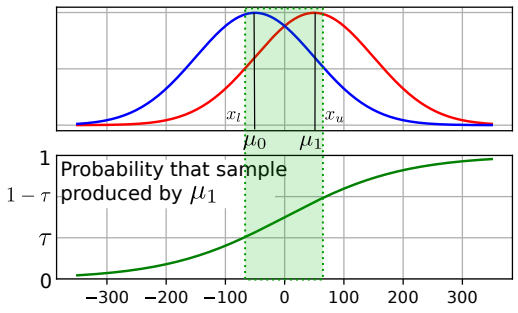

Fig. 2. Classification problem for the case with $n = 1$, $k = 3$, $\mu_0 = -50$, $\mu_1 = +50$, and $\sigma_0 = \sigma_1 = 100$. The relation described by (1) maps x to the posterior probability $P(\mu_1|x)$. In order to quantise the values for communication, one can choose thresholds τ and $(1-\tau)$ where the transmitted value switches from 0 to 1 and from 1 to 2 respectively (shown by the shaded region between x_l and x_u).

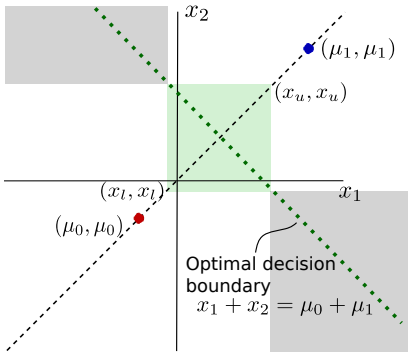

Fig. 3. Classification problem for the case with $n = 2$, $k = 3$ and $\sigma_0 = \sigma_1$. The optimal decision boundary is given by a line (a hyperplane for $n > 2$) as derived in the description below. The green shaded square in the middle corresponds to the region where both sensors observe samples in the $[\tau, 1-\tau]$ probability band. The two gray open-ended shaded regions correspond to a strong disagreement between the two sensors, with one sensor observing a sample in the $[0, \tau]$ probability band and other in the $[1-\tau, 1]$ band. Exactly half of the areas under the shaded regions correspond to a mismatch between the optimal decision and the quantised decision, while the unshaded areas see a match between the optimal and the quantised decisions.

A. *Optimal centralised classifier with full information*

Consider the generic case of n sensors with access to the raw sensor measurements $x_j \sim \mathcal{N}(\mu_t, \sigma_t)$, where $j \in \{1, \ldots, n\}$ and $t \in \{0, 1\}$. Assuming that the loss function is symmetric, the optimal classification policy will be to compare the *a posteriori* probabilities of the two hypotheses:

$$\hat{\theta} = \begin{cases} 1 & \text{if } P(\mu_1|x_1, \ldots, x_n) \geq P(\mu_0|x_1, \ldots, x_n) \\ 0 & \text{if } P(\mu_1|x_1, \ldots, x_n) < P(\mu_0|x_1, \ldots, x_n) \end{cases}$$

We apply Bayes rule, and note that both sides of the inequality contain the joint distribution $P(x_1, \ldots, x_n)$ in the denominator. Moreover, without prior knowledge about the operating modes, we assume a uniform prior $P(\mu_0) = P(\mu_1) = \frac{1}{2}$. This implies that the optimal aggregation policy is simply to compare the likelihoods of the sensor readings,

$$\hat{\theta} = \begin{cases} 1 & \text{if } P(x_1, \ldots, x_n|\mu_1) \geq P(x_1, \ldots, x_n|\mu_0) \\ 0 & \text{if } P(x_1, \ldots, x_n|\mu_1) < P(x_1, \ldots, x_n|\mu_0) \end{cases}$$

Since the likehoods are joint Gaussians in nature, we can obtain some intuition about the decision boundaries by considering small values of n. In the simplest case with $n = 1$ and $\sigma_0 = \sigma_1$, shown in Fig. 2, the decision boundary is the midpoint between μ_0 and μ_1. When $n = 2$, the likelihood is proportional to,

$$P(x_1, x_2|\mu_i) \propto \frac{1}{\sigma_i^2} \exp\left(-\sum_{j=1}^{2} \frac{(x_j - \mu_i)^2}{2\sigma_i^2}\right)$$

Remark 1: When $\sigma_0 = \sigma_1$, the interpretation in Fig. 3 shows that the likelihood test corresponds to a comparison between the Euclidean distance of (x_1, x_2) from the points (μ_0, μ_0) and (μ_1, μ_1) respectively. The decision boundary is thus a plane with the generic form $x_1 + x_2 = \mu_0 + \mu_1$. The same argument can be extended to the $n-$dimensional case, with the optimal binary classifier defined by a hyperplane with the equation $\sum_n x_j = \frac{n}{2}(\mu_0 + \mu_1)$. When $\sigma_1 \neq \sigma_0$, we will show presently that the decision boundary corresponds to a circle.

The difficulty with implementing this optimal classifier for the problem described in Sec. II is as follows:

- We do not have knowledge of μ_0 and μ_1
- We do not have direct access to x_j, and must rely on $y_j = f_j(x_j)$

We assume that the thresholds are set to τ and $(1-\tau)$, where $0 < \tau < 1/2$, for the 3-level, 2-sensor case. Then, at any given instant, the sensor $j \in \{1, 2\}$ outputs

$$y_j = \begin{cases} 0, & x_j < \tau \\ 1, & \tau \leq x_j \leq 1-\tau \text{ ('dead' or uncertain zone)} \\ 2, & \text{otherwise} \end{cases}$$

Of particular interest are the dead zones in the 2D real space spanned by x_1 and x_2. These correponds to regions where the sensors output $(y_1, y_2) = (1, 1)$, $(0, 2)$ or $(2, 0)$. For instance, the outputs $y_1 = y_2 = 1$ are obtained when

$$\tau \leq P(\mu_1|x_1)\,,\, P(\mu_1|x_2) \leq 1 - \tau$$

assuming that the logistic regression functions are correctly trained. We need to expand the posterior probability of μ_1 as follows:

$$P(\mu_1|x_1) = \frac{P(\mu_1)\, P(x_1|\mu_1)}{P(\mu_0)\, P(x_1|\mu_0) + P(\mu_1)\, P(x_1|\mu_1)}$$

Assuming the prior probabilities are equal and unknown, we can cancel $P(\mu_0)$ and $P(\mu_1)$ in the above equation. The posterior probability reduces to,

$$P(\mu_1|x_1) = \frac{\frac{1}{\sigma_1^2} \exp\left(\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sigma_1^2} \exp\left(\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1}{\sigma_0^2} \exp\left(\frac{-(x_1-\mu_0)^2}{2\sigma_0^2}\right)} \quad (1)$$

Since each state can fall into one of 3 outputs for each sensor, it follows that there are 9 possible combinations of outputs. We use (1) for x_1 as well as x_2 to compute the probability of falling within each of the 9 state zones, and the probability

of making an error. We consider two cases, depending on whether or not $\sigma_0 = \sigma_1$.

B. Case 1: $\sigma_0 = \sigma_1 = \sigma$

Sensor 1 outputs 1 (dead zone) if

$$\frac{\tau}{1-\tau} \leq \frac{\exp\left(\frac{-(x_1-\mu_1)^2}{2\sigma^2}\right)}{\exp\left(\frac{-(x_1-\mu_0)^2}{2\sigma^2}\right)} \leq \frac{1-\tau}{\tau}, \tag{2}$$

and a similar expression applies to sensor 2 producing 1 as the output. Let x_l and x_u, with $x_l < x_u$, denote the values of x_1 for which the two inequalities in (2) yield, one at a time, an equality. It can be shown readily that

$$x_{\{l,u\}} = \mu_m \mp \frac{\sigma^2}{\Delta\mu} \ln\left(\frac{1-\tau}{\tau}\right), \tag{3}$$
$$\mu_m = \frac{\mu_0+\mu_1}{2}, \quad \Delta\mu = \mu_1 - \mu_0$$

The probability that both sensors yield 1 as the output is given by

$$P([1,1]) = \frac{1}{2\pi\sigma^2}\left(\int_{x_l}^{x_u} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) dx\right)^2$$

Let $z = (x-\mu_1)/\sqrt{2}\sigma$, so that $dx = \sqrt{2}\sigma\, dz$ and

$$z_{\{l,u\}} = \frac{1}{\sqrt{2}\sigma}\left(-\frac{\Delta\mu}{2} \mp \frac{\sigma^2}{\Delta\mu}\ln\left(\frac{1-\tau}{\tau}\right)\right)$$

This gives

$$P([1,1]) = \frac{1}{\pi}\left(\int_{x_l}^{x_u} \exp(-z^2)\, dz\right)^2$$
$$= \left(\operatorname{erfn}\left(\frac{1}{\sqrt{2}}\left(-\frac{\Delta\mu}{2\sigma} + \frac{\sigma}{\Delta\mu}\ln\left(\frac{1-\tau}{\tau}\right)\right)\right)\right.$$
$$\left. -\operatorname{erfn}\left(\frac{1}{\sqrt{2}}\left(-\frac{\Delta\mu}{2\sigma} - \frac{\sigma}{\Delta\mu}\ln\left(\frac{1-\tau}{\tau}\right)\right)\right)\right)^2$$

where $\operatorname{erfn}(x) = (1/\sqrt{\pi})\int_{-\infty}^{z} \exp(-z^2)\, dz$. We define the following for brevity:

$$F_j = \operatorname{erfn}\left(\frac{1}{\sqrt{2}}\left(-\frac{\Delta\mu}{2\sigma} + (2j-1)\frac{\sigma}{\Delta\mu}\ln\left(\frac{1-\tau}{\tau}\right)\right)\right), \tag{4}$$
$$j \in \{0,1\}$$

so that

$$P([1,1]) = (F_1 - F_0)^2$$

Next, we derive an analytical expression for $P([0,2]|\mu_1)$. In order for sensors 1 and 2 to output 0 and 2, respectively, we have that

$$0 \leq P(\mu_1|x_1) < \tau, \quad 1-\tau < P(\mu_1|x_2) \leq 1$$

Using (1), we get

$$\frac{\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(x_1-\mu_0)^2}{2\sigma^2}\right)} < \frac{\tau}{1-\tau}, \quad \frac{\exp\left(-\frac{(x_2-\mu_1)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(x_2-\mu_0)^2}{2\sigma^2}\right)} > \frac{1-\tau}{\tau} \tag{5}$$

It follows that $x_1 < x_l$ and $x_2 > x_u$ in terms of the notation presented in (3), so that $P([0,2]) = F_0(1-F_1)$.

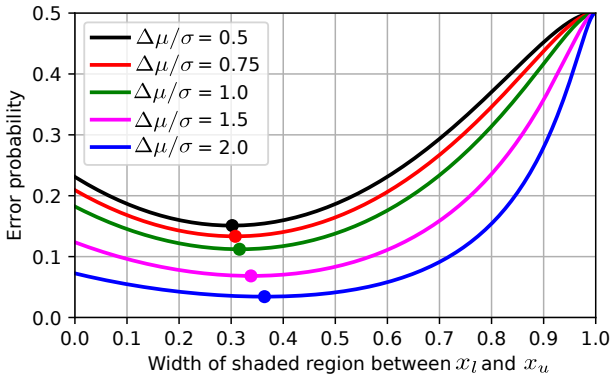

Fig. 4. The error probability from (6)

By symmetry, it follows that $P([2,0]) = F_0(1-F_1)$. Thus, we obtain the probability of error as

$$e(\tau) = \frac{1}{2}\left(P([1,1]) + P([0,2] + P([2,0])\right)$$
$$= \frac{1}{2}(F_1 - F_0)^2 + F_0(1-F_1), \tag{6}$$

which we plot in Fig. 4 as a function of τ for various values of $\Delta\mu/\sigma$. Notice that the error probability is minimized for values of τ between 0.3 and 0.35 across the entire range of $\Delta\mu/\sigma$ considered here.

C. Case 2: $\sigma_0 \neq \sigma_1$

We assume, w.l.o.g, that $\sigma_1 < \sigma_0$. Recall that

$$\frac{P(\mu_1|x)}{P(\mu_0|x)} = \frac{P(x|\mu_1)}{P(x|\mu_0)} = \frac{\frac{1}{\sigma_1^n}\prod_{j=1}^n \exp\left(-\frac{(x_j-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sigma_0^n}\prod_{j=1}^n \exp\left(-\frac{(x_j-\mu_0)^2}{2\sigma_0^2}\right)}$$

where n is the number of sensors. We will specialize our result for $n = 2$ later. Solving for the decision boundary $P(\mu_1|x) = P(\mu_0|x)$, we get

$$\sum_{j=1}^n \left(\frac{(x_j-\mu_1)^2}{2\sigma_1^2} - \frac{(x_j-\mu_0)^2}{2\sigma_0^2}\right) = \ln\left(\frac{\sigma_0^n}{\sigma_1^n}\right) \tag{7}$$

Let $\beta = \sigma_1/\sigma_0 < 1$ and, as before, $\Delta\mu = \mu_1 - \mu_0$. Let $z_j = x_j - \mu_1$ for all j, so that (7) can be written as

$$\sum_{j=1}^n \left(z_j^2 - \beta^2(z_j+\Delta\mu)^2\right) = -2n\sigma_1^2\ln(\beta) \tag{8}$$

Expanding the left hand side, completing the squares, and re-substituting for z_j in terms of x_j, we get

$$\sum_{j=1}^n \left(z_j^2 - 2\frac{\beta^2(\Delta\mu)}{1-\beta^2}z_j\right) = \frac{n\beta^2(\Delta\mu)^2}{1-\beta^2} - \frac{2n\sigma_1^2\ln(\beta)}{1-\beta^2}$$
$$\implies \sum_{j=1}^n \left(z_j - \frac{\beta^2}{1-\beta^2}\Delta\mu\right)^2 = \frac{n\beta^2(\Delta\mu)^2}{(1-\beta^2)^2} - \frac{2n\sigma_1^2\ln(\beta)}{1-\beta^2} \tag{9}$$
$$\implies \sum_{j=1}^n \left(x_j - \left(\mu_1 + \frac{\beta^2}{1-\beta^2}\Delta\mu\right)\right)^2 = \frac{n\beta^2(\Delta\mu)^2}{(1-\beta^2)^2}$$
$$-n\ln(\beta)\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right)^{-1}$$

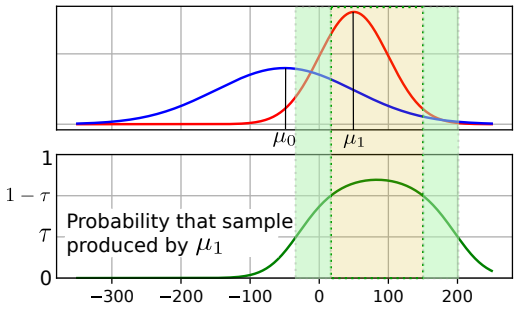

Fig. 5. Classification problem for the case with $n = 1$, $k = 3$, $\mu_0 = -50$, $\mu_1 = +50$, $\sigma_0 = 100$, and $\sigma_1 = 50$. The relation described by (1) maps x to the posterior probability $P(\mu_1|x)$. In order to quantise the values for communication, one can choose thresholds τ and $(1-\tau)$ as before. However, we now observe that the regions corresponding to probability ranges $[0, \tau]$ and $[\tau, 1-\tau]$ are not contiguous.

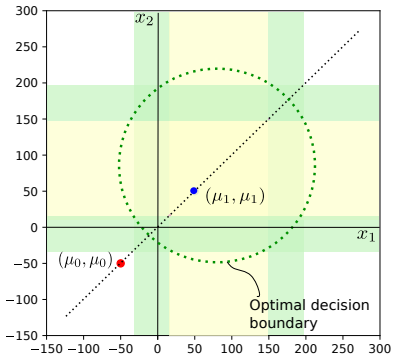

Fig. 6. Classification problem for the case with $n = 2$, $k = 3$, $\sigma_0 = 100$, and $\sigma_1 = 50$. The optimal decision boundary is now given by a circle (a spherical ball for $n > 2$) eccentric with respect to the narrower distribution. Extrapolating from the one-dimensional case, the decision regions for the quantised version are correspondingly more complex. Specifically for $k = 3$ and $n = 2$, there are 25 divisions of the $x_1 - x_2$ plane. The sensor outputs within each region are constant, and switch at the boundaries. The goal of optimising the threshold parameter τ would be to make these linear boundaries match up as closely as possible to the circular optimal boundary.

where the second term on the right hand side of the last equation is positive since $\beta < 1$. We are interested in $n = 2$, so that the decision boundary is a circle with its center "near" μ_1 (more generally, near the mean of the distribution with the smaller variance). For more general n, notice that the radius of the circle grows with $\sqrt{n}$. A formula similar to (6) can be derived for this case. We omit this derivation, and turn instead to a numerical study for insight.

IV. RESULTS

We present empirical studies that illustrate the preceding analysis, especially in terms of possible aggregation policies and their effect on the classification accuracy.

A. Aggregate classification policies

We use four aggregators, also called simple combiners [15, Chapter 5], in this section. Specifically, these are

1) Average: The centralised classifier calls $\hat{\theta} = 1$ iff $\frac{\sum \delta_i}{n} \geq \frac{k-1}{2}$, and $\hat{\theta} = 0$ otherwise.

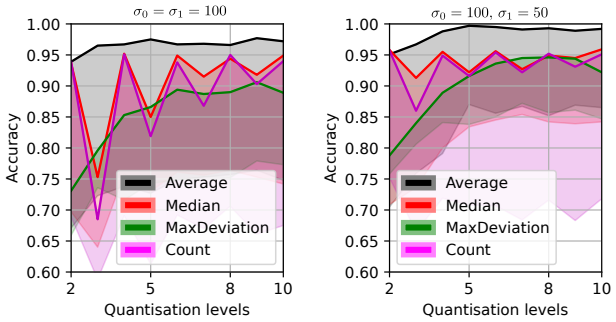

Fig. 7. Classification accuracy as a function of increasing quantisation levels k, for $\sigma_0 = \sigma_1 = 100$ (left) and $\sigma_0 = 100$, $\sigma_1 = 50$ (right), and all simple aggregation rules. The solid line corresponds to the highest number of sensors, $n = 15$, and the shaded regions cover all results for $n < 15$.

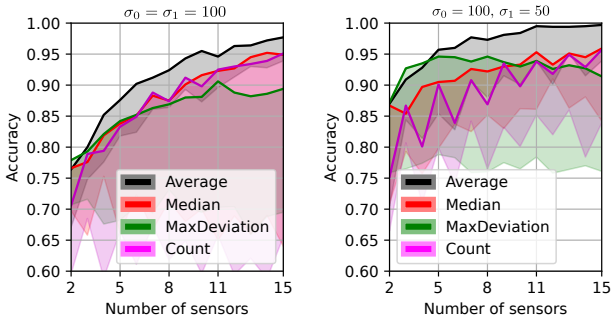

Fig. 8. Classification accuracy as a function of increasing number of sensors n, for $\sigma_0 = \sigma_1 = 100$ (left) and $\sigma_0 = 100$, $\sigma_1 = 50$ (right), and all simple aggregation rules. The solid line corresponds to the highest number of quantisation levels $k = 10$, and the shaded regions cover all results for $k < 10$.

2) Median: The centralised classifier calls $\hat{\theta} = 1$ iff $\text{median}(\delta_i) \geq \frac{k-1}{2}$, and $\hat{\theta} = 0$ otherwise.
3) MaxDeviation: The centralised classifier based on the sign of the largest deviation over all δ_i compared to $\frac{k-1}{2}$. Formally, it calls $\hat{\theta} = 1$ iff $|\max(\delta_i) - \frac{k-1}{2}| >= |\frac{k-1}{2} - \min(\delta_i)|$, and $\hat{\theta} = 0$ otherwise.
4) Count: The centralised classifier calls $\hat{\theta} = 1$ iff $\delta_i \geq \frac{k-1}{2}$ are at least as many as $\delta_i < \frac{k-1}{2}$, and $\hat{\theta} = 0$ otherwise.

Note that we break ties on the side of $\hat{\theta} = 1$ in all cases, without loss of generality. The expected error probability (conversely accuracy) remains the same if we break ties in favour of $\hat{\theta} = 1$ with any probability $\in [0, 1]$, due to the symmetry of the geometry as seen in Fig. 3 and Fig. 6.

B. Empirical results

In order to confirm the analytical results derived in the previous section, we performed empirical experiments for two scenarios:

1) Means $\mu_0 = -50$, $\mu_1 = +50$, equal standard deviations $\sigma_0 = \sigma_1 = 100$
2) Means $\mu_0 = -50$, $\mu_1 = +50$, unequal standard deviations $\sigma_0 = 100$ and $\sigma_1 = 50$

Each scenario is run with a range of quantisation levels $k \in [2, 10]$ and number of sensors $n \in [2, 15]$, and 1000 test

samples produced for each experiment. The $i = \{1, 2, \ldots, n\}$ readings obtained by sensors for each test sample are mapped to a posterior probability as per (1), and then to a quantised output δ_i. Given the complexity of optimising the switch thresholds for $k > 3$, we use a simple uniform division of the probability range $[0, 1]$ based on the available quantisation levels k. Formally, this means that a posterior probability of p_i as computed by sensor i is mapped to a quantised value $\delta_i = \lfloor p_i \, k \rfloor \in \{0, 1, \ldots, k-1\}$.

The empirically computed accuracy for all the aggregators is shown in Fig. 7 (left) for the case $\sigma_0 = \sigma_1 = 100$, as a function of the number of quantisation levels k. The solid lines correspond to the highest number of sensors $n = 15$ in the experiments, while the shaded regions show the range of accuracies obtained for $n < 15$. As expected, we note that the average aggregator dominates the other heuristics due to its affinity with the optimal decision boundary (see Fig. 3). The same empirical result is observed in Fig. 7 (right) for the case with $\sigma_0 = 100$ and $\sigma_1 = 50$, although we do not have a theoretical analysis of the same. Fig. 8 presents the same results but as a function of increasing number of sensors, for $\sigma_0 = \sigma_1$ (left) and $\sigma_0 \neq \sigma_1$ (right) respectively. The combined observations from these results can be summarised as follows:

- The jittery nature of the accuracy (especially for the Median and Count aggregators) is due to different behaviour in the case of even and odd number of quantisation levels or sensors. However, these effects are of opposite shape. In the case of quantisation levels, having an even number is beneficial because it maps each quantised level to a definite decision (the first $k/2$ values indicate $\hat{\theta} = 0$, while the remaining indicate $\hat{\theta} = 1$). On the other hand, allowing an odd number of quantisation levels puts the middle value on the fence, requiring a tie-breaker. The same effect applies to the number of sensors, but in this case having an odd number of sensors reduces the chance of requiring a tie-breaker, leading to higher accuracy.
- While the MaxDeviation rule is not competitive for any quantisation level as long as n is a high number, it is very competitive for small values of n (see Fig. 8). This is because when only a few readings are available, having an outlier value (corresponding to a posterior probability close to either 0 or 1) is a good indicator of the true distribution. However as n increases, the other statistical rules become much more reliable, while an outlier value is much more likely to be just that: a random outlier.
- For a quantisation level of $k = 2$ (see Fig. 7), the three statistical rules (Average, Median, Count) are identical since the sensors can only transmit a single binary value. This is why the three curves start from the same accuracy level. However, MaxDeviation is different because by definition (and including the tie-breaker), it decides on $\hat{\theta} = 1$ as long as at least one δ_i is equal to 1.

V. CONCLUSIONS

Our analysis of an apparently simple problem (noisy distributed measurements, a digital communication channel, and a centralised classifier) showed that there are several non-trivial design choices involved in the pursuit of classification accuracy. The intractable nature of the digitisation mapping optimisation problem for $k > 2$ suggests that a data-driven approach might be used for this step. Other interesting research directions include extensions to categorical or multi-dimensional states, non-standard measurement noise, communication channel noise, non-stationary distributions, and the inclusion of state dynamics.

REFERENCES

[1] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order lti systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2011.

[2] A. S. Leong, S. Dey, and G. N. Nair, "Quantized filtering schemes for multi-sensor linear state estimation: Stability and performance under high rate quantization," *IEEE Transactions on Signal Processing*, vol. 61, no. 15, pp. 3852–3865, 2013.

[3] H. B. Mitchell, *Multi-sensor data fusion: an introduction*. Springer Science & Business Media, 2007.

[4] R. Shorey, A. Ananda, M. C. Chan, and W. T. Ooi, "Detection, energy, and robustness in wireless sensor networks," in *Mobile, Wireless, and Sensor Networks: Technology, Applications, and Future Directions*, 2005, pp. 145–172.

[5] X. Luo and G. B. Giannakis, "Energy-constrained optimal quantization for wireless sensor networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 462930, 2007.

[6] D. Prajapati and H. Prajapati, "Data fusion in wireless sensor networks," 2020. [Online]. Available: https://doi.org/10.36227/techrxiv.12101373

[7] R. Niu, M. Moore, , and D. Klamer, "Decision fusion in a wireless sensor network with a large number of sensors," *Electrical Engineering and Computer Science*, vol. 82, 2004.

[8] N. Sriranga, K. G. Nagananda, R. S. Blum, A. Saucan, and P. K. Varshney, "Energy-efficient decision fusion for distributed detection in wireless sensor networks," in *2018 21st International Conference on Information Fusion (FUSION)*, 2018, pp. 1541–1547.

[9] E. Soltanmohammadi, M. Orooji, and M. Naraghi-Pour, "Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 205–215, 2013.

[10] K. Liu and A. M. Sayeed, "Type-based decentralized detection in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1899–1910, 2007.

[11] J.-F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, 2003.

[12] M. Bahrepour, N. Meratnia, and P. J. M. Havinga, "Sensor fusion-based event detection in wireless sensor networks," in *2009 6th Annual International Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous*, 2009, pp. 1–8.

[13] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2661–2675, 2015.

[14] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[15] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2014.