14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

LOCALLY OPTIMUM ESTIMATION IN WIRELESS SENSOR NETWORKS

Stefano Marano[a], *Vincenzo Matta*[a], *and Peter Willett*[b]

[a]DIIIE, University of Salerno,
via Ponte don Melillo, 84084, Fisciano (SA), Italy.
phone: + (39) 089-964269, fax: + (39) 089-964218
email: {marano,vmatta}@unisa.it

[b]ECE Department, U-157, University of Connecticut,
Storrs, CT 06269 USA.
phone: +(1) 860-486-2195, fax: +(1) 860-486-2447
email: willett@engr.uconn.edu.

ABSTRACT

A locally optimum approach for estimating a nonrandom parameter lying in some small neighborhood of a known nominal value is considered. Reference is made to a decentralized estimation problem in the context of wireless sensor networks, and particular attention is paid to the design of the quantizers used by the remote sensors.

1. INTRODUCTION

We consider a Wireless Sensor Network (WSN) engaged in the task of estimating a nonrandom parameter ξ, assuming that a nominal value of that (e.g., $\xi = 0$) is known in advance. Specifically, it is assumed that each node (i.e., sensor) of the network observes a random quantity x_i drawn from a family of probability density functions (pdfs) $f_\xi(x)$ parametrized by the unknown ξ. Observations are assumed iid (independent and identically distributed) across sensors, and these are transferred to a common fusion center (FC) that is devoted to produce the final estimate $\widehat{\xi}$. Actually, since x_i are real valued, we assume that the observations are first quantized and then sent to the FC. Focus is made just on the quantizers' design, assuming that these are scalar.

In Sect. 2, we draw parallels to the theory of Locally Optimum Detection (LOD), and exploit its paradigms and tools [1] in an inference context. As for the LOD, we end up with a simple *canonical form* of the estimator (referred to as the LOE, locally optimum estimator). Indeed, it turns out that the density $f_\xi(x)$ of the observations rules the behavior of a certain nonlinear function $g(x)$ by which the local observations are processed before being additively combined to form the estimate.

We then recognize that the structure of the LOE is particularly suited to WSN applications and, in fact, in Sect. 3 we exploit the LOE paradigm to design the quantizers at the remote nodes of the network. Furthermore, the LOE's additive structure matches well with the separate quantization of the x_i's, which is a design constraint in networks with sensors that do not communicate each other.

In Sect. 4 the convergence of the LOE to its asymptotic performance is checked, and a numerical investigation of the LOE rate distortion characteristic is provided. In addition, the quantization method is compared with standard compression techniques, where the final goal of the system is data *reproduction* rather than *estimation* of a parameter embedded in the observations. Concluding remarks are offered in Sect. 5.

2. LOCALLY OPTIMUM ESTIMATE

Let n be the number of sensors in the network and $f_\xi(x)$ the actual density of the observations x_i's. We want to estimate ξ from the observed $\{x_i\}_{i=1}^n$ and, as a distinct feature of our approach, it is known that ξ lies in some *small* neighborhood of a nominal value, say $\xi = 0$. Let us start from the logarithmic likelihood ratio between the actual density $f_\xi(x)$ and its nominal counterpart $f_0(x)$ that is obtained by setting $\xi = 0$:

$$L(\mathbf{x}) = \log \frac{f_\xi(\mathbf{x})}{f_0(\mathbf{x})} = \sum_{i=1}^n \log \frac{f_\xi(x_i)}{f_0(x_i)}.$$

Consider hence the behavior of $L(\mathbf{x})$ for large n and for $\xi \to 0$. Borrowing a standard approach from the theory of locally optimum detection [1], let us define $\gamma = \xi\sqrt{n}$, so that when n increases without bound and γ is arbitrary but fixed, ξ approaches 0 at a prescribed rate. Thus, for ξ approaching 0, we can expand the above expression in Taylor series about the nominal value $\xi = 0$; this, accounting for the definition of γ, yields

$$\begin{aligned} L(\mathbf{x}) &\approx \sum_{i=1}^n \left.\frac{\partial \log f_\xi(x_i)}{\partial \xi}\right|_{\xi=0} \frac{\gamma}{\sqrt{n}} \\ &+ \sum_{i=1}^n \left.\frac{\partial^2 \log f_\xi(x_i)}{\partial \xi^2}\right|_{\xi=0} \frac{\gamma^2}{2n}. \end{aligned} \quad (1)$$

Let us introduce now the *locally optimum estimator* (LOE), whose connection with eq. (1) will become clear shortly:

$$\widehat{\xi}_{LOE}(\mathbf{x}) = \frac{1}{nI(0)} \sum_{i=1}^n \left.\frac{\partial \log f_\xi(x_i)}{\partial \xi}\right|_{\xi=0} = \frac{1}{nI(0)} \sum_{i=1}^n g(x_i), \quad (2)$$

where the function

$$g(x) = \left.\frac{\partial \log f_\xi(x)}{\partial \xi}\right|_{\xi=0}, \quad (3)$$

is referred to as the score of the random variable x drawn from $f_\xi(x)$, computed at $\xi = 0$. The main property of the LOE is now stated.

Proposition 1. It results that

$$\sqrt{n}\left(\widehat{\xi}_{LOE}(\mathbf{x}) - \xi\right) \overset{f_\xi}{\sim} \mathscr{G}\left(0, \frac{1}{I(0)}\right), \quad (4)$$

where $\overset{f_\xi}{\sim} \mathscr{G}(a,b)$ *means that, under* $f_\xi(x)$*, the LHS converges in distribution to a Gaussian with mean* a *and variance* b*. In* *the above*

$$I(0) = \int f_0(x) \left[\left(\frac{\partial \log f_\xi(x)}{\partial \xi} \right)^2 \right] \Bigg|_{\xi=0} dx$$

is the Fisher information per sample computed under the nominal density $f_0(x)$.

Before proving the claim in eq. (4), let us stress its main implications. First, as detailed below, for large n we can confuse the second term in eq. (1) with its asymptotic value under $f_0(x)$. Expression (1) accordingly becomes

$$\sum_{i=1}^{n} \frac{\partial \log f_\xi(x_i)}{\partial \xi} \Bigg|_{\xi=0} \xi - \frac{\xi^2}{2} n I(0),$$

whose derivative with respect to ξ is zero at $\xi = \widehat{\xi}_{LOE}(\mathbf{x})$: the LOE is nothing but an ML estimator when the log-likelihood is approximated around the nominal value $\xi = 0$.

Second, asymptotically, $E_\xi[\widehat{\xi}_{LOE}(\mathbf{x}) - \xi] \approx 0$ and $\mathrm{VAR}_\xi[\widehat{\xi}_{LOE}(\mathbf{x})] \approx n^{-1} I^{-1}(0)$, where E_ξ and VAR_ξ represent statistical expectation and variance under distribution $f_\xi(x)$. Also, if the nominal Fisher information $I(0)$ in eq. (4) were replaced by the actual $I(\xi)$, computed under $f_\xi(x)$, then the above claim simply states that $\widehat{\xi}_{LOE}(\mathbf{x})$ is asymptotically optimal, in the usual Cramer-Rao sense [2]. From a practical viewpoint, it is clear that the Fisher information $I(\xi)$ is close to $I(0)$ as consequence of the fact that ξ vanishes to 0.

Further, we would like to emphasize that the *canonical structure* of the LOE is particularly appealing: eq. (2) reveals that the fusion center simply needs a transformed version $g(x_i)$ of the individual remote observations, in order to build up the estimator. This represents a simple and practical recipe for designing the data processing chain for estimation problems that can be modeled in the described asymptotic setting. Clearly, in the case of decentralized estimation, the transmission of the continuous valued quantities $g(x_i)$ is usually inhibited by the finite capacity of the communication channels between remote nodes and fusion center. Then, we are faced with the problem of quantizing the $g(x_i)$'s, before transmission; the topic is addressed in the next section.

Proof of Proposition 1. An outline of the proof of (4) goes as follows. Let $n \to \infty$.

a. Under $f_0(x)$, the term in the first line of eq. (1) is Locally Asymptotically Normal (LAN) with zero mean and variance $\gamma^2 I(0)$. That is to say

$$\sqrt{n}(\text{first line of eq. (1)}) \overset{f_0}{\sim} \mathscr{G}\left(0, \gamma^2 I(0)\right)$$

(see [3]).

b. Under $f_0(x)$, the term $\frac{1}{n} \partial^2 \log f_\xi(x_i) / \partial \xi^2 |_{\xi=0}$ in the second line of eq. (1) converges in probability to the constant $-I(0)$ (see, e.g., [4]).

c. *(Fundamental theorem on the use of contiguity)* Given two families of distribution functions $\{w^{(n)}(x)\}_{n=1}^\infty$ and $\{z^{(n)}(x)\}_{n=1}^\infty$, assume that the sequence of the associated likelihood ratios $\{w^{(n)}(x)/z^{(n)}(x)\}_{n=1}^\infty$ converges under $z^{(n)}(x)$ to a Gaussian distribution with mean μ and variance σ^2. Then *contiguity holds* and under $w^{(n)}(x)$ the likelihood sequence converges to a Gaussian with mean $\mu + \sigma^2$ and variance σ^2 (see [5]).

From *(a)* and *(b)*, it follows that $L(\mathbf{x})$ converges in distribution, under $f_0(x)$, to $\mathscr{G}\left(-\gamma^2 I(0)/2, \gamma^2 I(0)\right)$. Then, statement *(c)*, in view of Slutsky's theorem [6], immediately implies the following convergence in distribution:

$$L(\mathbf{x}) \overset{f_\xi}{\sim} \mathscr{G}\left(\frac{\gamma^2}{2} I(0), \gamma^2 I(0)\right).$$

Now, combining the definition of the LOE estimator with the likelihood in eq. (1) we easily get:

$$\sqrt{n}\left(\widehat{\xi}_{LOE}(\mathbf{x}) - \xi\right) \approx \frac{L(\mathbf{x})}{\gamma I(0)} - \sum_{i=1}^{n} \frac{\partial^2 \log f_\xi(x_i)}{\partial \xi^2} \Bigg|_{\xi=0} \frac{\gamma}{2nI(0)} - \gamma. \quad (5)$$

As is more or less obvious, in the above equation as well as in the initial likelihood (1), we have neglected any term which can be safely assumed vanishing (in probability) for large n. To this aim, we assume that all the conditions required for this to be rigorously true are fulfilled. In eq. (5), it is recognized that (i) the first term converges in distribution to a Gaussian with mean $\gamma/2$ and variance $I^{-1}(0)$, and (ii) the second term goes in probability to $\gamma/2$. The claim of Proposition 1 now follows as a direct application of Slutsky's theorem [6]. □

3. DESIGN OF THE QUANTIZERS

Let us consider a WSN with remote nodes that collect the x_i to be sent to the common FC. From the previous section, it follows that the remote nodes should send to the fusion center a transformed version $g(x_i)$ of their observation x_i, see eq. (2). However, sensors of the network send data to the fusion center by means of channels with finite capacity and, accordingly, the nodes must employ some kind of compression of their continuous observations. In some WSN applications the quantization may also be rather coarse, in the sense that very few bits must be used.

In the following we assume that each sensor employs a scalar quantizer $\mathscr{Q}$, and that these quantizers are identical across all the nodes (for symmetry). Let $q_i = \mathscr{Q}(x_i)$ be the discrete valued data to be delivered to the FC, and let $p_\xi(q)$ be the associated probability mass function, discrete counterpart of density $f_\xi(x)$:

$$p_\xi(q) = \int_{x \in \Re_q} f_\xi(x) dx,$$

with $\Re_q$ being the partition region yielding q as output. In the same asymptotic setting described in the previous section (i.e., $\gamma = \xi\sqrt{n}$, with $n \to \infty$, and γ held fixed), our goal is now to minimize the estimation error, measured in terms of the MSE, with a constraint on the rate of the quantizers, measured in terms of the number of quantization bits.

The arguments in the previous section can be repeated using as observables the q_i's in place of their continuous counterpart x_i's. Then, from eq. (4) we know that there exists a class of estimation problems (one for each n) for which the asymptotic performance can be measured by the Fisher information of these discrete valued observations, computed at the nominal value of the parameter. This supports the natural idea that one can optimize the quantizers working in the nominal case of $\xi = 0$. We formalize the issue in the following assertion.

Proposition 2. Let $\gamma = \xi\sqrt{n}$, and assume $n \to \infty$ with γ fixed. In the class of the LOE estimators, the optimization problem

$$\min_{\mathscr{Q}:\log_2 ||\mathscr{Q}|| \le R} \mathrm{MSE}_q, \qquad (6)$$

where MSE_q is the estimation mean square error using the quantizer $\mathscr{Q}$, reduces to

$$\min_{\mathscr{Q}:\log_2 ||\mathscr{Q}|| \le R} \varepsilon(g(x), c_q), \qquad (7)$$

where $g(x)$ is the optimal nonlinearity given in eq. (3), and

$$c_q = \left.\frac{\partial \log p_\xi(q)}{\partial \xi}\right|_{\xi=0} = \frac{\int_{x\in\Re_q} g(x) f_0(x)dx}{\int_{x\in\Re_q} f_0(x)dx}, \qquad (8)$$

where the function ε is defined as

$$\varepsilon(g(x), c_q) = \int f_0(x)(g(x)-c_q)^2\,dx.$$

Some comments are in order. Note that c_q in eq. (8) is the *centroid* (MMSE estimation, given q) of the optimal nonlinearity $g(x)$, computed with respect to pdf $f_0(x)$. In other words, c_q is a quantized version of the optimal (unquantized) nonlinearity. Thus, eq. (7) represents a classical optimization problem in the context of quantization for reproduction purposes: you have a continuous quantity $g(x)$ and c_q is its (scalar) quantized version, and you want to minimize the *(reproduction)* mean square error $E[(g(x)-c_q)^2]$ between the original $g(x)$ and the quantized counterpart c_q, where the expectation is with respect to $f_0(x)$. Such an optimization problem, hence, can be solved by means of a standard Lloyd & Max's algorithm [7], which provides us with the best quantizer $\mathscr{Q}$ achieving the minimum reproduction error, subject to a constraint on the number of bits.

Basically, we have reduced the problem of optimal quantizer design for *inference* purposes, to the standard problem of optimal quantizer design for *reconstruction* purposes. In this way, we are allowed to use the bulk of methods, tools, and skills, in the area of quantization. In fact, in the next section a well-known Lloyd & Max's algorithm is exploited in order to design the sensors' quantizers.

Proof of Proposition 2. The proof of the equivalence between (6) and (7) is now outlined. First, consider the LOE $\widehat{\xi}_q(\mathbf{x})$ computed using the quantized data q_i. As for the unquantized case, we can define

$$\widehat{\xi}_q(\mathbf{x}) = \frac{1}{nI_q(0)} \sum_{i=1}^{n} \left.\frac{\partial \log p_\xi(q_i)}{\partial \xi}\right|_{\xi=0},$$

where $I_q(0)$ is the Fisher information *per sample* for the quantized case. Following the same derivation steps as in the unquantized case, we get, in the asymptotic regime,

$$E_\xi[\widehat{\xi}_q(\mathbf{x}) - \xi] \approx 0, \quad \mathrm{VAR}_\xi[\widehat{\xi}_q(\mathbf{x})] \approx \frac{1}{nI_q(0)}.$$

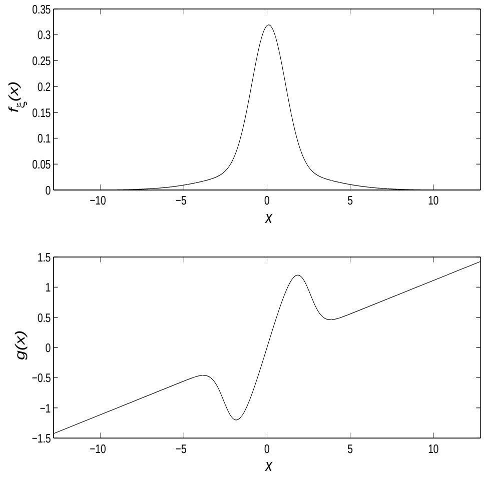

Figure 1: Mixture of Gaussians with the same mean and different standard deviations. Here we have $\xi = 0.1$, $\sigma_1 = 1$, $\sigma_2 = 3$ and $\alpha = 0.7$. *Top panel:* the density $f_\xi(x)$. *Bottom panel:* the optimal nonlinearity $g(x)$.

The implication is that, instead of minimizing the MSE_q, one can maximize $I_q(0)$, the Fisher information of the quantized observations under the nominal value $\xi = 0$: there exists a class of estimators that attain that asymptotic performance.

Further, the two alternative expressions of c_q, as given in eq. (8), reveal that c_q is both the score of the discrete valued random variable q and the centroid[1] of $g(x)$ with respect to $f_0(x)$. This implies (averages are computed under $f_0(x)$):

$$\begin{aligned} I(0) &= E[g^2(x)] = E[(g(x)-c_q+c_q)^2] \\ &= E[(g(x)-c_q)^2] + E[c_q^2] + 2E[(g(x)-c_q)c_q], \end{aligned}$$

and the last addend is zero, as c_q is an MMSE estimate of $g(x)$. Hence,

$$I(0) = E[(g(x)-c_q)^2] + E[c_q^2] = \varepsilon(g(x), c_q) + I_q(0).$$

We have thus shown that: minimize $\mathrm{MSE}_q \Leftrightarrow$ maximize $I_q(0) \Leftrightarrow$ minimize $\varepsilon(g(x), c_q)$. □

4. APPLICATIONS

We now consider examples of applications of the previous theory to a decentralized estimation problem in a WSN. Specifically, we are interested in the rate (No. of bits) distortion (estimation MSE) behavior in a WSN engaged in the task of estimating ξ, knowing that ξ lies in some small neighborhood of 0. Sensor i^{th} observes x_i and compute $g(x_i)$; this latter is quantized by the Lloyd & Max's algorithm [7], as described earlier. Finally, the fusion center receives the quantized sensors' outputs and provides the final estimate, according to the LOE approach.

4.1 Gaussian case

Consider first a Gaussian problem, in which the mean of Gaussian observations is to be estimated, assuming known

[1] The second expression in eq. (8) can be easily derived from the former; details are omitted. Recall also that the centroid is an MMSE estimate [7].

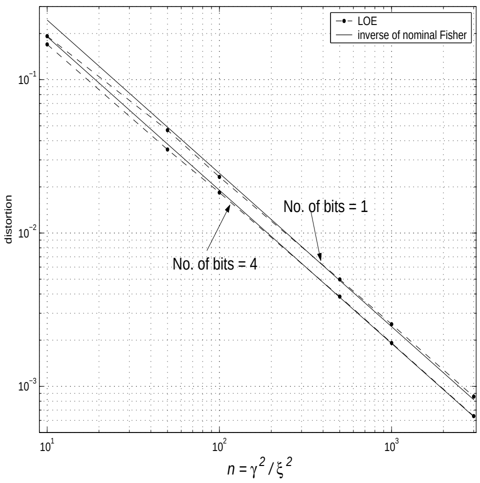

Figure 2: MSE of the LOE compared to its asymptotic value $1/I_q(0)$, for a mixture of Gaussians. As in the previous plot, $\sigma_1 = 1$, $\sigma_2 = 3$, $\alpha = 0.7$, and two rates (No. of bits) are considered. Note that $\gamma = 3$ is held fixed so that when n increases, ξ decreases as $n^{-1/2}$.

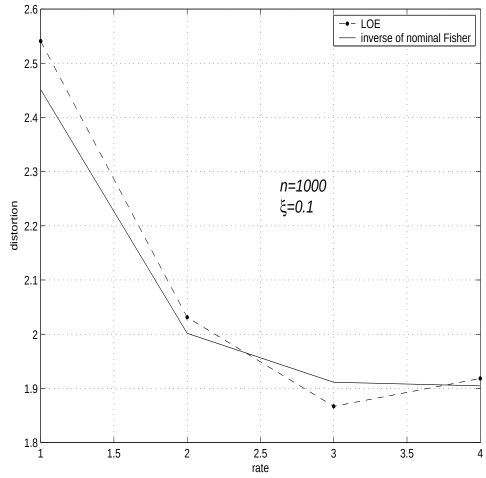

Figure 3: MSE of the LOE compared to its asymptotic value $1/I_q(0)$, for a mixture of Gaussians. Here we have $\sigma_1 = 1$, $\sigma_2 = 3$, $\alpha = 0.7$ (as before), $\xi = 0.1$, and $n = 10^3$.

variance. It is straightforward to show that the optimal nonlinearity in eq. (3) is $g(x) \propto x$: the optimal estimator fuses the original (untransformed) observations x_i and, as a consequence, the attempt here is to recover at the FC the $\{x_i\}_{i=1}^n$ with the best possible fidelity. The implication is that estimation-oriented quantization is the same of that reconstruction-oriented and no benefit can arise by the previous theory[2].

4.2 Mixture of Gaussians

Consider a mixture of two Gaussian densities with the same mean ξ and standard deviations σ_1 and σ_2, with relative probabilities (coefficients of the mixture) α and $1-\alpha$, respectively. Figure 1 depicts the pdf $f_\xi(x)$ and the optimal nonlinearity $g(x)$, see eq. (3).

Figure 2 shows the convergence of the LOE's MSE to the asymptotic value of $I_q^{-1}(0)$, assuming that ξ scales as $1/\sqrt{n}$, as prescribed by the asymptotic theory. Oppositely, in Fig. 3, ξ and n are given, and the rate (No. of bits) distortion (estimation MSE) characteristic is shown.

Now, we consider an estimation system in which the quantizers are designed as described in Sect. 3, but the fusion center does not employ the LOE. In other words, let us assume that the previous theory is used to design the remote quantizers, but it is not used at the FC to produce the final estimates. Then, it makes sense to compare the Fisher information contained in the quantized data $\{q_i\}_{i=1}^n$, and to use the inverse of this as distortion proxy. We consider four cases.

Case 1. The Fisher information per sample $I_q(\xi)$, corresponding to the quantized observations $q_i = \mathscr{Q}(x_i)$, is (numerically) computed under the *true* pmf p_ξ (namely when the x_i's come from f_ξ).

Case 2. We compute (again, numerically) the Fisher information $I_q(0)$ pertaining to the quantized q_i's, under the *nominal* pmf p_0 (we thus assume that the x_i's are drawn from f_0).

[2]We get similar insights in [8].

Reproduction-oriented quantization is aimed to recovering the x_i's as accurately as possible at the fusion center. This approach can be referred to as a *blind* method, in that the quantization stage completely ignores the fact that the final aim of the system is to estimate a parameter embedded in the observations, rather than recovering the observations themselves[3]. Not unexpectedly, the blind approach can lead, in some cases, to a large waste of system resources. We consider two blind cases.

Case 3. The Lloyd & Max's quantizer $\mathscr{S}$ that minimizes the *reproduction* error is designed using the nominal distribution $f_0(x)$. The data x_i's are generated according to the true pdf $f_\xi(x)$ and quantized as $s_i = \mathscr{S}(x_i)$; the related Fisher information is then computed numerically, and its inverse is taken as distortion measure.

Case 4. We also consider the ideal case where the Lloyd & Max's procedure is still designed to minimize the reproduction MS error, but the design algorithm is run using the true distribution $f_\xi(x)$ (which is actually unknown). The quantized variables, say $c_i = \mathscr{C}(x_i)$, are used to compute the Fisher information per sample that is used in the comparisons[4].

Figure 4 shows the rate distortion behavior of the system, considering the four cases described earlier. Note the presence of a curve labelled as "time sharing": this represents the actual rate distortion curve that can be obtained by using different bit rates for prescribed fractions of time. Two values of ξ are considered to investigate the effect of moving away from the nominal point $\xi = 0$.

5. SUMMARY

In the classical theory of locally optimum detection, to test for a vanishing small signal (e.g., a constant ξ) is studied under the assumption that the number n of (iid) observations becomes increasingly large. This can be conveniently for-

[3]To be explicit, in this case one still uses the Lloyd & Max algorithm, but this is run over x_i rather than over $g(x_i)$.

[4]Note that, as opposed to cases 1 and 2, the quantizers used in cases 3 and 4 are different.

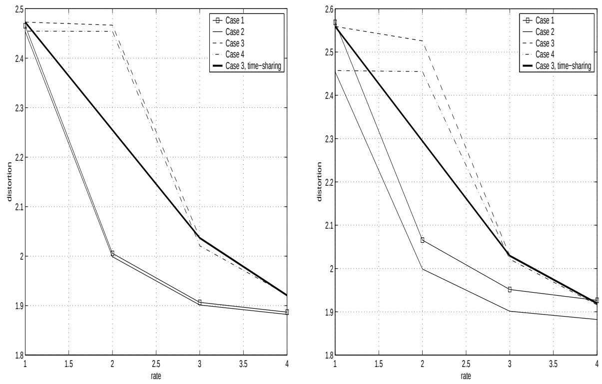

Figure 4: Rate distortion characteristics for a mixture of Gaussians. Here $\sigma_1 = 1$, $\sigma_2 = 3$, and $\alpha = 0.7$. *Left panel:* $\xi = 0.1$. *Right panel:* $\xi = 0.3$.

malized by assuming that ξ goes to zero at rate $1/\sqrt{n}$. We have investigated the similar problem of estimating a non-random parameter ξ, assuming that a nominal value of that (e.g., $\xi = 0$) is a-priori known, meaning that it is known that ξ lies in some *small*, not exactly specified, neighborhood of 0. Setting $\xi = \gamma/\sqrt{n}$ (γ constant) with n the number of iid observations, we formalize the problem as a sequence of estimation problems, one for each n, and consider the asymptotic behavior for $n \longrightarrow \infty$.

In this setting, the LOE (locally optimum estimator) is introduced and its properties are investigated; then, in the context of a WSN decentralized estimation, we exploit the LOE paradigm to design the quantizers employed at the remote nodes of the network. It turns out that (i) the LOE can be obtained by solving the maximum likelihood equation, with the log-likelihood expanded around the nominal point $\xi = 0$; and (ii) for the decentralized estimation scenario, a meaningful way to design the quantizers consists in optimizing the restitution levels by maximizing as objective function the Fisher information computed under the nominal value $\xi = 0$, thus yielding a simple and constructive way to solve the decentralized estimation problem.

A contribution of our work is to (re)derive these two intuitive design recipes in a suitable mathematical framework. However, this is our first investigation of the issue and, as such, we have only presented the basic ideas and tools. Part of our ongoing work is focused on better defining the range of applicability of the method and to assessing a comprehensive comparison with alternative schemes.

REFERENCES

[1] S. A. Kassam, *Signal Detection in Non-Gaussian Noise,* Springer-Verlag, 1987.

[2] A. M. Maras, “Threshold Parameter Estimation in Non-additive Non-Gaussian Noise,” *IEEE Transactions on Signal Processing,* Vol. 45, No. 7, pp. 1681-1688, July 1997.

[3] L. M. Lecam, *Théorie asymptotique de la décision statistique,* Les presses de l’université de Montréal, 1969.

[4] H. Cramér, *Mathematical Methods of Statistics,* Princeton University Press, Princeton USA, 1946.

[5] V. Genon-Catalot, D. Picard, *Eléments de statistique asymptotique,* Springer-Verlag, 1993.

[6] J. Shao, *Mathematical Statistics*, Springer Verlag, 1999.

[7] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[8] S. Marano, V. Matta, P. Willett, “Asymptotical Design of Quantizers for Decentralized MMSE Estimation,” submitted to *IEEE Transactions on Signal Processing,* September 2005.