Adaptive Quantizers for Estimation

Rodrigo Cabral Farias, Jean-Marc Brossier

Adaptive Quantizers for Estimation

Rodrigo Cabral Farias[a,*], Jean-Marc Brossier[a]

[a] *GIPSA-Lab, 11 rue des Mathématiques, BP46 - 38402 Saint-Martin d'Hères, France*

Abstract

This paper addresses a problem of location parameter estimation from multibit quantized measurements. An adaptive estimation algorithm using an adjustable quantizer is proposed. By using general results from adaptive algorithms theory, the asymptotic estimation performance is obtained and optimized through the quantizer parameters. Despite its very low complexity, it can be shown that the proposed algorithm is asymptotically optimal for estimating a constant parameter.

The asymptotic performance for optimal quantizer parameters is shown to rapidly reach real-valued based estimation performance as the number of bits increases. In practice, 4-bit quantization appears to be enough for estimation purposes. It is also shown that the performance gap between the quantized and continuous cases is even smaller when the parameter varies according to a random walk (Discrete Wiener process with or without drift).

Keywords: Parameter estimation, adaptive estimation, quantization.

1. Introduction

Continuous advances in the development of cheaper and smaller sensors and communication devices motivated the introduction of sensor networks in many different domains, *e.g.* military applications, infrastructure security, environment monitoring, industrial applications and traffic monitoring [1]. When designing a sensing system, one must account not only for the physical perturbations that can affect sensing performance, more specifically noise, but also for the inherent design constraints such as bandwidth and complexity limitations. Commonly, the effect of the noise in system performance is taken into account, but bandwidth and complexity constraints are neglected.

One simple way to respect bandwidth constraints is to compress sensor information using quantizers. The theory of quantizer design for reducing distortion in the measurement representation is well established in the literature [2], however much less results can be found when the quantities to be reconstructed are not directly the measurements but an underlying parameter embedded in noise.

*Corresponding author.
Email addresses:
rodrigo.cabral-farias@gipsa-lab.grenoble-inp.fr (Rodrigo Cabral Farias),
jean-marc.brossier@gipsa-lab.grenoble-inp.fr (Jean-Marc Brossier)

In [3], noisy samples of a constant are taken using a quantizer with adjustable input offset, the output samples of the quantizer are used to estimate the constant. Using this type of measurement system, results for different types of offset were obtained. The types of offset considered were known constant and variable offset, random offset and offset based on feedback of the output measurements. The comparison was performed based on the Cramér–Rao bound (CRB) ratio which is the worst case ratio between the CRB for quantized measurements and continuous measurements. It was shown that the last type of offset, based on feedback, is the most efficient one.

Another interesting result from [3] is that in the Gaussian noise case with one bit quantized measurements, the minimum CRB ratio that can be attained is $\frac{\pi}{2}$. This result was used as a motivation for [4] to study more in detail estimation under Gaussian noise and binary quantization. In [4], it was shown that the CRB for a fixed known threshold can be upper bounded by the exponential of the squared difference between the threshold and the constant to be estimated. This means that the closer the threshold is to the parameter to be estimated with binary measurements, the lower can be the estimation variance. It was also pointed out that an iterative algorithm could be used to adjust the threshold exactly to be the last estimate of the parameter.

An adaptive algorithm for placing the threshold was detailed in [5], where a sensor network extension was also proposed. At each time step, a sensor measures one bit, updates its threshold using a simple cumulative sum and broadcasts the new threshold to the other sensors and to a fusion center. Thus, the thresholds are placed around the parameter in an adaptive way and at the fusion center the broadcasted bits are used to obtain a more precise estimate of the parameter. Two other methods for updating the thresholds were presented in [6], one method used a more refined cumulative sum based on the last two measured bits, the other proposed method was to estimate the parameter using a maximum likelihood method and then set the threshold at the estimate of the parameter. It was shown that in the asymptotic case (large number of iterates) the CRB for the fusion center estimate using maximum likelihood threshold updates converges to the minimum possible CRB, which is the CRB when the threshold is placed exactly at the parameter. In a similar way, an adaptive estimation-quantization procedure based on binary observations is presented in [7], for each binary observation the parameter is estimated with a recursive algorithm and only one threshold update is done for a block of observations, as the size of the block tends to infinity, the estimator is shown to converge asymptotically to the true parameter value.

In the same line of the work mentioned above, algorithms for estimating a scalar parameter from multiple bit quantized noisy measurements are proposed in this paper. The algorithms developed in this work jointly estimate the parameter and set the quantizer thresholds. They are based on low complexity adaptive techniques that can be easily implemented in practice. The mean and mean squared error (MSE) are obtained for a general class of symmetrically distributed noise and three types of parameter evolution: constant, Wiener process and Wiener process with drift. As in related work [3], the loss of estimation performance due to quantization is also evaluated and the validity of the performance results is verified through simulation.

The main contributions of this work are

- *Design and analysis of adaptive estimation algorithms based on multiple bit quantized noisy measurements.* Differently from the algorithms in [5], [6] and [7] where only binary quantization is treated.

- *Explicit performance analysis for tracking of a varying parameter.* In [3–6] the parameter is set to be constant and all subsequent analysis is based on this hypothesis, whereas in this paper, we consider explicitly the performance analysis for tracking a varying parameter.

- *Low complexity algorithms.* The algorithms proposed here are based on simple recursive techniques that have lower complexity than the maximum likelihood methods used in [5] and [6].

The paper is structured in the following form: in Section II the problem is stated and the main assumptions are made, in Section III the general adaptive algorithm and results from adaptive algorithms theory are presented, then in Section IV the parameters of the adaptive algorithm are obtained. Section V contains theoretical performance results and also the simulation of the algorithm. Section VI concludes the paper.

2. Problem statement

Let $\mathbf{X}$ be a discrete-time stochastic process defined on the probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbb{P})$ with values on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, at each instant $k \in \mathbb{N}^{\star}$, the corresponding scalar random variable (r.v.) X_k will be given by the following model:

$$X_k = X_{k-1} + W_k, \tag{1}$$

where W_k is a sequence of independent Gaussian random variables with its mean given by a small amplitude deterministic unknown sequence u_k and small known standard deviation σ_w:

$$W_k \sim \mathcal{N}\left(u_k, \sigma_w^2\right), \tag{2}$$

W_k is independent of X_n for $n < k$ and the initial condition X_0 will be considered to be an unknown deterministic constant.

The model expressed in (1) is a compact form to describe three different evolution models for X_k:

- *Constant*: by taking $u_k = \sigma_w = 0$, then $X_k = X_0 = x$ is an unknown deterministic constant.

- *Wiener process*: if $u_k = 0$, $\sigma_w > 0$ and small , then X_k is a slowly varying Wiener process. This model is commonly used to describe a slowly varying parameter of a system when the model for its evolution is random but with unknown form.

- *Wiener process with drift*: in this case u_k and σ_w are non zero and with small amplitudes. The fact that u_k is nonzero makes the Wiener process to have a drift, thus representing a model with a deterministic component that is perturbed by small random fluctuations.

The process $\mathbf{X}$ is observed through $\mathbf{Y}$ and they are related as follows:

$$Y_k = X_k + V_k, \tag{3}$$

where the noise V_k is a sequence of additive independent and identically distributed (i.i.d.) r.v. which is also independent of W_k. The cumulative distribution function (CDF) of V_k will be denoted by F. Some assumptions on F are stated below.

Assumptions (on the noise distribution):

A1. F is locally Lipschitz continuous.

A2. F admits a probability density function (PDF) f with respect to (w.r.t.) the standard Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

A3. The PDF $f(x)$ is an even function and it strictly decreases w.r.t. $|x|$.

The first assumption is required by the method of analysis that will be used to assess the performance of the proposed algorithms. Most noise CDFs considered in practice are Lipschitz continuous, thus the first assumption is generally satisfied. Assumption 2 is a commonly used assumption that in practice will be used when the derivative of F w.r.t. its arguments is needed. Assumption 3 will be used to prove the asymptotic convergence of the algorithms and it is also commonly satisfied in practice.

The observations are quantized using an adjustable quantizer whose output is given by

$$i_k = Q\left(\frac{Y_k - b_k}{\Delta_k}\right), \tag{4}$$

where i_k is an integer defined on a finite set of N_I integers, N_I being the number of quantization intervals. The quantizer parameters b_k and $\frac{1}{\Delta_k}$ are sequences of adjustable offsets and gains respectively. The function Q represents a static normalized quantizer and it is characterized by $N_I + 1$ thresholds. For simplification purposes some assumptions on the quantizer will be used.

Assumptions (on the quantizer):

A4. N_I will be considered to be an even natural number and

$$i_k \in I = \left\{-\frac{N_I}{2}, \ldots, -1, +1, \ldots, +\frac{N_I}{2}\right\}.$$

A5. It will be assumed that the static quantizer is symmetric and centered at zero. This means that the

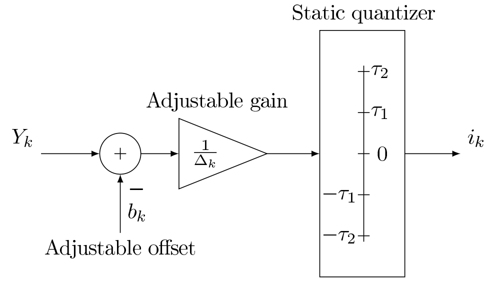

Figure 1: Scheme representing the adjustable quantizer. The offset and gain can be adjusted dynamically while the quantizer thresholds are fixed.

vector of thresholds[1]

$$\boldsymbol{\tau} = \left[\tau_{-\frac{N_I}{2}} \ \ldots \ \tau_{-1} \ \tau_0 \ \tau_1 \ \ldots \ \tau_{\frac{N_I}{2}}\right]^\top$$

has elements given by the following expressions

$$\begin{cases} \tau_0 & = \quad 0, \\ \tau_i & = \quad -\tau_{-i}, \ \forall i \in \left\{1, \cdots, \frac{N_I}{2}\right\}, \\ \tau_{\frac{N_I}{2}} & = \quad +\infty. \end{cases} \tag{5}$$

These assumptions will be used later to simplify the choice of parameters of the algorithms.

For $\frac{|Y_k - b_k|}{\Delta_k} \in [\tau_{i-1}, \tau_i)$, the adjustable quantizer output is given by

$$i_k = Q\left(\frac{Y_k - b_k}{\Delta_k}\right) = i \operatorname{sign}(Y_k - b_k). \tag{6}$$

A scheme representing the quantizer is given in Fig. 1. Note that even if the quantizer is not uniform (with constant distance between thresholds), it can be implemented using a uniform quantizer with a compander approach [2].

Based on the quantizer outputs the main objective is, according to a minimum MSE criterion, to estimate X_k, which is a location parameter of the distribution of Y_k (the mean of the distribution whenever it exists). A secondary objective is to adjust the parameters b_k and Δ_k to enhance estimation performance. As the estimate $\hat{X}_k$ of X_k will be possibly used in real time applications, X_k might be estimated online, which means that $\hat{X}_k$ will only depend on past and present i_k. To simplify it will be considered that the offset is set to be $\hat{X}_{k-1}$ and that the gain is set to be a constant Δ. For the adaptive algorithm presented later, the fact that the offset is set to

[1] Infinite thresholds are used to have the same notation for the probabilities of the granular and overload regions.

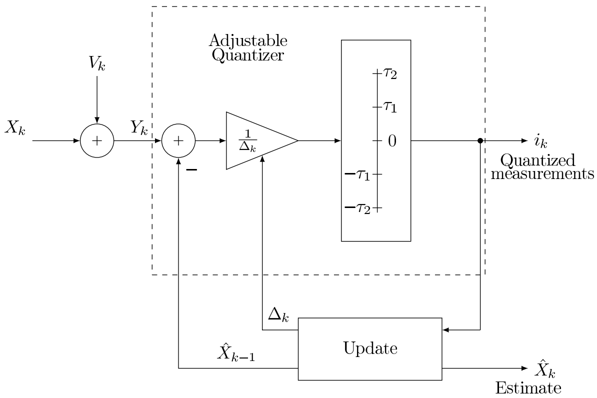

Figure 2: Block representation of the estimation scheme. The estimation algorithm and the procedures to set the offset and the gain are represented by the *Update* block. In this paper, Δ_k is considered to be a constant Δ.

$\hat{X}_{k-1}$ will have, as a consequence, an asymptotic performance that does not depend on the mean of X_k, thus simplifying the analysis. The choice of Δ is discussed in Section IV.

The general scheme for the estimation of X_k is depicted in Fig. 2 and the main objective will be to find a low complexity algorithm that will be placed in the block named *Update*.

3. General algorithm

A simple and general form for the estimation algorithm that respects the constraints defined above (low complexity and online) is the following adaptive algorithm:

$$\hat{X}_k = \hat{X}_{k-1} + \gamma_k \eta \left[Q\left(\frac{Y_k - \hat{X}_{k-1}}{\Delta} \right) \right]. \tag{7}$$

In the expression above, γ_k is a sequence of positive real gains and $\eta[\cdot]$ is a mapping from I to $\mathbb{R}$ that is defined as a sequence of N_I coefficients $\left\{\eta_{-\frac{N_I}{2}}, \ldots, \eta_{-1}, \eta_1, \ldots, \eta_{\frac{N_I}{2}}\right\}$, these coefficients are equivalent to the output quantization levels used in quantization theory. The use of this algorithm is also motivated by the following observations:

- when estimating a constant, the maximum likelihood estimator can be approximated by a simpler online algorithm using a stochastic gradient ascent algorithm, which has the same form as (7). It is shown in Section IV that for the optimal choice of η_i, (7) is equivalent to a stochastic gradient ascent method to maximize the log-likelihood.
- To estimate a Wiener process, a simple choice of estimator is a Kalman filter like method based on the quantized innovation, which is also (7).

Due to the symmetry of the noise distribution, when $\hat{X}_k$ is close to X_k, it seems reasonable to suppose that the corrections given by the output quantizer levels have odd symmetry with positive values for positive i, this symmetry will be useful later for simplification purposes. Thus, one assumption will be added to A1-A5.

Assumption (on the quantizer output levels):

A6. The quantizer output levels have odd symmetry w.r.t. i:

$$\eta_i = -\eta_{-i}, \tag{8}$$

with $\eta_i > 0$ for $i > 0$.

The non-differentiable nonlinearity in (7) makes it difficult to be analyzed. Fortunately, an analysis based on mean approximations was developed in [8] for a wide class of adaptive algorithms, within this framework, the function $\eta\left[Q\left(\cdot\right)\right]$ could be a general nonlinear non-differentiable function of Y_k and $\hat{X}_k$ and it was shown that the gains γ_k that optimize the estimation (minimize the estimation MSE) of X_k should be as follows:

- $\gamma_k \propto \frac{1}{k}$ when X_k is constant.
- γ_k is constant for a Wiener process X_k.
- $\gamma_k \propto u_k^{\frac{2}{3}}$ when X_k is a Wiener process with drift.

In the following parts of this section the results of [8] will be applied for the analysis of (7) in the three evolution models of X_k.

3.1. Constant X_k

In this case $X_k = x$. To obtain convergence of $\hat{x}_k$ to a constant, the gains must be [8]:

$$\gamma_k = \frac{\gamma}{k}. \tag{9}$$

For large k, the mean trajectory of $\hat{X}_k$ can be approximated using the ordinary differential equation (ODE) method. The ODE method approximates the expectation of the estimator $\mathbb{E}\left[\hat{X}_k\right]$ by $\hat{x}\left(t_k\right)$, where $\hat{x}\left(t\right)$ is the solution of

$$\frac{d\hat{x}}{dt} = h\left(\hat{x}\right), \tag{10}$$

the correspondence between continuous and discrete time is given by $t_k = \gamma \sum_{j=1}^{k} \frac{1}{j}$ and $h(\hat{x})$ is the following:

$$h(\hat{x}) = \mathbb{E}\left[\eta\left(Q\left(\frac{x-\hat{x}+V}{\Delta}\right)\right)\right], \qquad (11)$$

where the expectation is evaluated w.r.t. $F(v)$.

For the solution of (10) to be valid as an approximation of $\mathbb{E}\left[\hat{X}_k\right]$, $h(\hat{x})$ has to be a locally Lipschitz continuous function of $\hat{x}$. Using the assumptions on the quantizer thresholds and output levels, the expectation in (11) can be written as:

$$h(\hat{x}) = \sum_{i=1}^{\frac{N_I}{2}} \left[\eta_i F_d(i,\hat{x},x) - \eta_i F_d(-i,\hat{x},x)\right], \qquad (12)$$

where F_d is a difference of CDFs:

$$F_d = \begin{cases} F(\tau_i\Delta + \hat{x} - x) - F(\tau_{i-1}\Delta + \hat{x} - x), \\ \qquad \text{if } i \in \left\{1,\cdots,\frac{N_I}{2}\right\}, \\ F(\tau_{i+1}\Delta + \hat{x} - x) - F(\tau_i\Delta + \hat{x} - x), \\ \qquad \text{if } i \in \left\{-1,\cdots,-\frac{N_I}{2}\right\}. \end{cases} \qquad (13)$$

From assumption A1, the function h is a linear combination of locally Lipschitz continuous functions, which implies that h is also locally Lipschitz continuous, thus the ODE method can be applied.

If $\hat{x} \to x$ when $t \to \infty$ for all x and all $\hat{x}(0)$, the adaptive algorithm is asymptotically unbiased, and in this case it can also be shown, using a central limit theorem, that the estimation error is asymptotically distributed as a Gaussian r.v. [8, pp. 109]:

$$\sqrt{k}\left(\hat{X}_k - x\right) \underset{k\to\infty}{\rightsquigarrow} \mathcal{N}\left(0,\sigma_\infty^2\right), \qquad (14)$$

where the asymptotic variance σ_∞^2 is given by:

$$\sigma_\infty^2 = \frac{\gamma^2 R(x)}{-2\gamma h_{\hat{x}}(x) - 1}, \qquad (15)$$

- The term denoted R in the numerator is the variance of the adaptive algorithm normalized increments $\left(\frac{\hat{X}_k - \hat{X}_{k-1}}{\gamma_k}\right)$ when $\hat{x}$ is equal to x. From A3 and A6, $h(\hat{x}) = 0$ when $\hat{x} = x$ and this variance can be written as the second order moment of the quantizer output levels:

$$\begin{aligned} R(x) &= \mathbb{V}\mathrm{ar}\left[\eta\left(Q\left(\frac{x-\hat{x}+V}{\Delta}\right)\right)\right]\Bigg|_{\hat{x}=x} \\ &= \sum_{i=1}^{\frac{N_I}{2}} \left[\begin{array}{c} \eta_i^2 F_d(i,x,x) \\ +\ \eta_{-i}^2 F_d(-i,x,x) \end{array}\right] \\ &= 2\sum_{i=1}^{\frac{N_I}{2}} \eta_i^2 F_d(i,x,x), \end{aligned} \qquad (16)$$

where the last equality comes from the symmetry assumptions.

- The term in the denominator is the derivative of h when $\hat{x}$ is equal to x:

$$\begin{aligned} h_{\hat{x}}(x) &= \frac{dh}{d\hat{x}}\bigg|_{\hat{x}=x} \\ &= -\sum_{i=1}^{\frac{N_I}{2}} \left[\begin{array}{c} \eta_i f_d(i,x,x) \\ -\eta_i f_d(-i,x,x) \end{array}\right] \end{aligned} \qquad (17)$$

with

$$f_d = \begin{cases} f(\tau_{i-1}\Delta + \hat{x} - x) - f(\tau_i\Delta + \hat{x} - x), \\ \qquad \text{if } i \in \left\{1,\cdots,\frac{N_I}{2}\right\}, \\ f(\tau_i\Delta + \hat{x} - x) - f(\tau_{i+1}\Delta + \hat{x} - x), \\ \qquad \text{if } i \in \left\{-1,\cdots,-\frac{N_I}{2}\right\}. \end{cases} \qquad (18)$$

From the symmetry assumptions, $f_d(i,x,x)$ is odd w.r.t. i, thus (17) can be rewritten as

$$h_{\hat{x}}(x) = -2\sum_{i=1}^{\frac{N_I}{2}} \eta_i f_d(i,x,x). \qquad (19)$$

Minimizing σ_∞^2 w.r.t. the positive gain γ gives

$$\gamma^\star = -\frac{1}{h_{\hat{x}}(x)} \qquad (20)$$

$$\sigma_\infty^2 = \frac{R(x)}{h_{\hat{x}}^2(x)}. \qquad (21)$$

When $\hat{x} = x$, the functions $F_d(i,\hat{x},x)$ and $f_d(i,\hat{x},x)$ do not depend on x anymore, thus from now on they will be denoted $F_d[i]$ and $f_d[i]$. The functions $R(x)$ and $h_{\hat{x}}(x)$ do not depend on x either, thus they will be denoted by the constants R and $h_{\hat{x}}$ respectively.

To specify completely the adaptive algorithm, the quantizer parameters η_i, $\boldsymbol{\tau}$ and Δ can be chosen to minimize (21).

3.2. *Wiener process*

If X_k is a Wiener process, the mean of W_k is $u_k = 0$ and the variance is a known constant $\mathbb{V}\text{ar}[W_k] = \sigma_w^2$. The algorithm gain can be chosen to be a constant $\gamma_k = \gamma$. For small σ_w^2, the mean trajectory of $\hat{X}_k$ is also approximated by (10), x being the initial condition x_0 of the Wiener process, which is equal to its mean for every k. Thus, if $\hat{x}$ converges to x, the algorithm is asymptotically unbiased and, in this case, it can be shown [8, pp. 130-131] under the additional condition

$$h_{\hat{x}} < 0, \tag{22}$$

that the estimation error process ξ_k is asymptotically stationary with a marginal Gaussian density $\mathcal{N}\left(0, \sigma_\xi^2\right)$, where the variance is

$$\sigma_\xi^2 = \frac{\gamma^2 R + \sigma_w^2}{-2\gamma h_{\hat{x}}}. \tag{23}$$

Thus, the asymptotic estimation MSE, MSE_∞, can be approximated by σ_ξ^2. Minimizing MSE_∞ w.r.t. γ gives the optimal γ:

$$\gamma^\star = \frac{\sigma_w}{\sqrt{R}}. \tag{24}$$

The MSE for $\gamma^\star$ is

$$\text{MSE}_\infty = \frac{\sigma_w \sqrt{R}}{-h_{\hat{x}}}. \tag{25}$$

Using (25) and (21) the MSE can be rewritten as

$$\text{MSE}_\infty = \sigma_w \sigma_\infty. \tag{26}$$

Both the asymptotic MSE for estimating a Wiener process and the asymptotic variance for estimating a constant depend on the quantizer parameters through σ_∞, therefore the optimal quantizer parameters will be the same in both cases. The only difference in the adaptive algorithms for these two cases is the sequence of gains γ_k.

3.3. *Wiener process with drift*

In this case the mean of W_k is nonzero and given by a small amplitude sequence u_k, the variance σ_w^2 is constant. The gain γ_k will be considered to be variable in time and under the assumption of asymptotic unbiasedness for constant X_k, the MSE can be approximated by the term due to the estimation bias which is given by [8, pp. 136]:

$$\text{MSE}_k = \mathbb{E}\left[\hat{X}_k - X_k\right]^2 \approx \frac{u_k^2}{\gamma_k^2 h_{\hat{x}}^2} - \gamma_k \frac{R}{2h_{\hat{x}}}. \tag{27}$$

Minimization w.r.t. γ_k leads to

$$\gamma_k^\star = \left[\frac{4u_k^2}{-h_{\hat{x}} R}\right]^{\frac{1}{3}} \tag{28}$$

$$\text{MSE}_k \approx 3\left[\frac{|u_k|}{4}\frac{R}{h_{\hat{x}}^2}\right]^{\frac{2}{3}}. \tag{29}$$

Note that in practice, u_k may be unknown and it will be necessary to replace its value in $\gamma_k^\star$ by an estimate of it $\hat{U}_k$, which can be also obtained adaptively, for example by calculating a recursive mean on $\hat{X}_k - \hat{X}_{k-1}$.

The MSE can also be rewritten as a function of σ_∞^2 with a dependence on u_k

$$\text{MSE}_k \approx 3\left[\frac{|u_k|}{4}\sigma_\infty^2\right]^{\frac{2}{3}}. \tag{30}$$

Also in this case the MSE is an increasing function of σ_∞. From the three cases it is possible to see that the quantizer design will depend on the following:

1. *Asymptotic unbiasedness*: it is necessary to prove asymptotic unbiasedness of the algorithm when X_k is constant for the MSE results given above to be valid. This can be done by proving the asymptotic global stability of the ODE (10) for an arbitrary $X_k = x$ and $\hat{X}_0 = \hat{x}(0)$ in $\mathbb{R}$.
2. *Minimization of* σ_∞^2: the quantizer parameters can be chosen to minimize σ_∞^2 and, as a consequence, they will maximize the performance for the three evolution models of X_k.

4. Asymptotic unbiasedness and adaptive algorithm design

In this section, first it will be shown that the algorithm is asymptotically unbiased. Then, optimization of the algorithm asymptotic performance will be done by minimizing σ_∞^2 ,which depends on η_i, Δ and $\boldsymbol{\tau}$. The optimal coefficients η_i will be found and then the choice for the parameters Δ and $\boldsymbol{\tau}$ will be discussed.

4.1. *Asymptotic unbiasedness*

For the asymptotic performance results to be valid, it is necessary to prove that the estimation procedure when $X_k = x$ is asymptotically unbiased. For doing so, one needs to prove that the solution of (10) for any $\hat{x}(0)$ and x tends to x as $t \to \infty$.

The approximation for the mean error can be written as

$$\epsilon = \hat{x} - x \tag{31}$$

and the ODE for the mean error is

$$\frac{d\epsilon}{dt} = \tilde{h}(\epsilon), \tag{32}$$

where $\tilde{h}(\epsilon) = h(\epsilon + x)$ is a function that does not depend on x.

It is necessary to prove that $\epsilon \to 0$ as $t \to \infty$ for every $\epsilon(0) \in \mathbb{R}$, which means that $\epsilon = 0$ is a globally asymptotically stable point [9]. Global asymptotic stability of $\epsilon = 0$ can be shown using an asymptotic stability theorem for nonlinear ODEs. This will require the definition of an unbounded Lyapunov function of the error. To simplify, a quadratic function will be used:

$$\mathcal{L}(\epsilon) = \epsilon^2, \tag{33}$$

which is a positive definite function and tends to infinity when ϵ tends to infinity.

If $\tilde{h}(\epsilon) = 0$ for $\epsilon = 0$ and $\frac{d\mathcal{L}}{dt} < 0$ for $\epsilon \neq 0$ then by the Barbashin–Krasovskii theorem [9, Ch. 4], $\epsilon = 0$ is a globally asymptotically stable point.

To show that both conditions are met, expression (12) can be rewritten using A6:

$$h(\epsilon) = \sum_{i=1}^{\frac{N_I}{2}} \eta_i \left[\tilde{F}_d(i,\epsilon) - \tilde{F}_d(-i,\epsilon)\right], \tag{34}$$

where $\tilde{F}_d(i,\epsilon) = F_d(i,\epsilon + x, x)$ is also a function that does not depend on x.

When $\epsilon = 0$, the differences between $\tilde{F}_d$ in the sum are differences between probabilities on symmetric intervals, the symmetry of the noise PDF stated in A3 and the symmetry of the quantizer stated in A5 imply that $\tilde{h}(0) = 0$, fulfilling the first condition.

The second condition can be written in more detail by using the chain rule for the derivative:

$$\frac{d\mathcal{L}}{dt} = \frac{d\mathcal{L}}{d\epsilon}\frac{d\epsilon}{dt} = 2\epsilon\tilde{h}(\epsilon) < 0, \quad \text{for} \quad \epsilon \neq 0. \tag{35}$$

The function $\tilde{h}(\epsilon)$ has to respect the following constraints:

$$\tilde{h}(\epsilon) > 0, \text{for } \epsilon < 0 \text{ and } \tilde{h}(\epsilon) < 0, \text{for } \epsilon > 0. \tag{36}$$

When $\epsilon \neq 0$, the terms in the sum that gives $\tilde{h}(\epsilon)$ are the difference between integrals of the noise PDF under the same interval size but with asymmetric interval centers. Using the symmetry assumptions, for $\epsilon > 0$, $\tilde{F}_d(i,\epsilon)$ is the integration of f over an interval more distant to zero than for $\tilde{F}_d(-i,\epsilon)$, then by the decreasing assumption on f, $\tilde{F}_d(i,\epsilon) < \tilde{F}_d(-i,\epsilon)$ and consequently $\tilde{h}(\epsilon) < 0$. Using the same reasoning for $\epsilon < 0$ one can show that $\tilde{h}(\epsilon) > 0$. Therefore, the inequalities in (36) are verified and $\frac{d\mathcal{L}}{dt} < 0$ for $\epsilon \neq 0$.

Finally, as both conditions are satisfied one can say that $\epsilon = 0$ is globally asymptotically stable, which means that the estimator is asymptotically unbiased and that all the performance results obtained are valid.

Note that from A3 and A5, $h_{\hat{x}}(x) < 0$, thus the supplementary condition for stationarity (22) is also respected.

4.2. Optimal quantizer parameters

The performance of the adaptive algorithm can be maximized by minimizing σ_∞^2 w.r.t. the quantizer levels η_i. Using (16) and (19) in (21) gives the following minimization problem:

$$\underset{\boldsymbol{\eta}}{\arg\min}\left\{\frac{R}{h_{\hat{x}}^2}\right\} = \underset{\boldsymbol{\eta}}{\arg\min}\left\{\frac{\boldsymbol{\eta}^\top \mathbf{F_d}\boldsymbol{\eta}}{2\left[\boldsymbol{\eta}^\top \mathbf{f_d}\right]^2}\right\}, \tag{37}$$

where $\boldsymbol{\eta}$ is a vector with the coefficients

$$\boldsymbol{\eta} = \left[\eta_1 \, \ldots \, \eta_{\frac{N_I}{2}}\right]^\top. \tag{38}$$

$\mathbf{F_d}$ is a diagonal matrix given by

$$\mathbf{F_d} = \operatorname{diag}\left[F_d[1], \cdots, F_d\left[\frac{N_I}{2}\right]\right] \tag{39}$$

and $\mathbf{f_d}$ is the following vector

$$\mathbf{f_d} = \left[f_d[1] \cdots f_d\left[\frac{N_I}{2}\right]\right]^\top. \tag{40}$$

The minimization problem is equivalent to the following maximization problem:

$$\underset{\boldsymbol{\eta}}{\arg\max}\left\{\frac{\left[\boldsymbol{\eta}^\top \mathbf{f_d}\right]^2}{\boldsymbol{\eta}^\top \mathbf{F_d}\boldsymbol{\eta}}\right\}. \tag{41}$$

Using the fact that $\mathbf{F_d}$ is diagonal with non zero diagonal elements, (41) becomes

$$\underset{\boldsymbol{\eta}}{\arg\max}\left\{\frac{\left[\left(\mathbf{F_d}^{\frac{1}{2}}\boldsymbol{\eta}\right)^\top\left(\mathbf{F_d}^{-\frac{1}{2}}\mathbf{f_d}\right)\right]^2}{\left(\mathbf{F_d}^{\frac{1}{2}}\boldsymbol{\eta}\right)^\top\left(\mathbf{F_d}^{\frac{1}{2}}\boldsymbol{\eta}\right)}\right\}, \tag{42}$$

the matrices $\mathbf{F_d}^{\frac{1}{2}}$ and $\mathbf{F_d}^{-\frac{1}{2}}$ are obtained by taking the square root and the inverse of the square root of the diagonal elements in $\mathbf{F_d}$. Using the Cauchy–Schwarz inequality on the expression in the numerator gives

$$\left\{\frac{\left[\left(\mathbf{F_d}^{\frac{1}{2}}\boldsymbol{\eta}\right)^\top\left(\mathbf{F_d}^{-\frac{1}{2}}\mathbf{f_d}\right)\right]^2}{\left(\mathbf{F_d}^{\frac{1}{2}}\boldsymbol{\eta}\right)^\top\left(\mathbf{F_d}^{\frac{1}{2}}\boldsymbol{\eta}\right)}\right\} \leq \mathbf{f_d}^\top\mathbf{F_d}^{-1}\mathbf{f_d} \tag{43}$$

and the equality happens for

$$\mathbf{F_d}^{\frac{1}{2}}\boldsymbol{\eta} \propto \mathbf{F_d}^{-\frac{1}{2}}\mathbf{f_d}. \tag{44}$$

Therefore, the optimal $\boldsymbol{\eta}$ can be chosen to be

$$\boldsymbol{\eta}^{\star} = \mathbf{F_d}^{-1}\mathbf{f_d}. \tag{45}$$

It is possible to see that the coefficients chosen in this way still depends on Δ and $\boldsymbol{\tau}$. The minimum σ_∞^2 is

$$\sigma_\infty^2 = \frac{1}{2\left(\mathbf{f_d}^\top \mathbf{F_d}^{-1}\mathbf{f_d}\right)} = \left(2\sum_{i=1}^{\frac{N_I}{2}} \frac{f_d^2\left[i\right]}{F_d\left[i\right]}\right)^{-1}. \tag{46}$$

To simplify the choice of the constant Δ, it will be considered that the noise CDF is parametrized by a known scale parameter δ, which means that

$$F\left(x\right) = F_n\left(\frac{x}{\delta}\right), \tag{47}$$

where F_n is the noise CDF for $\delta = 1$. Thus, the evaluation of the quantizer output levels can be simplified by setting:

$$\Delta = c_\Delta \delta. \tag{48}$$

Since the coefficients $\boldsymbol{\eta}^\star$ do not depend on x anymore, for a given c_Δ and noise CDF, they can be pre-calculated and stored in a table. For $i > 0$, these coefficients are given by

$$\eta_i^\star = \frac{f_d\left[i\right]}{F_d\left[i\right]}. \tag{49}$$

Note that for Δ given by (48), η_i depends on δ only through a $\frac{1}{\delta}$ multiplicative factor, the other factor can be written as a function of normalized PDFs and CDFs, thus this factor can be pre-calculated based only on the normalized distribution. Note also that the $\eta_i^\star$ are given by the score function for estimating a constant location parameter when considering that the offset is fixed and placed exactly at x, therefore this algorithm is equivalent to a gradient ascent technique to maximize the log-likelihood that iterates only one time per observation and sets the offset each time at the last estimate.

Using the η_i from (49), the adaptive estimator can be written as

$$\hat{X}_k = \hat{X}_{k-1} + \gamma_k \text{sign}\left(i_k\right)\eta_{|i_k|}, \tag{50}$$

with $i_k = Q\left(\frac{Y_k - \hat{X}_{k-1}}{\Delta}\right)$.

The sum in (46) is the Fisher information I_q for estimating a constant x from the output of the adjustable quantizer with an offset exactly placed at x:

$$I_q = 2\sum_{i=1}^{\frac{N_I}{2}} \frac{f_d^2\left[i\right]}{F_d\left[i\right]}, \tag{51}$$

this quantity can be maximized w.r.t. $\boldsymbol{\tau}$, thus leading to the following optimization problem:

$$\boldsymbol{\tau}^\star = \underset{\boldsymbol{\tau}}{\arg\max} \quad I_q. \tag{52}$$

Problem (52) without constraints on the thresholds seems to be very difficult to solve analytically and no simple solutions for this problem were found in the literature. Therefore, general solutions for (52) will not be treated here, for the results that will be presented in Section V it will be considered that the quantizer is uniform, with $\boldsymbol{\tau}$ defined as follows

$$\boldsymbol{\tau} = \left[\tau_1 = 1 \cdots \tau_{\frac{N_I}{2}-1} = \frac{N_I}{2} - 1\, \tau_{\frac{N_I}{2}} = \infty\right]^\top, \tag{53}$$

then in this case, only c_Δ need to be set and consequently a grid method can be used.

In the next section the results for each case using the choice of parameters obtained above will be detailed and discussed.

5. Results and simulation

It will be supposed that the noise CDF and δ are known and also the type of evolution model for X_k. Thus for a given N_I, c_δ and $\boldsymbol{\tau}$, the coefficients η_i used in the estimation algorithm (50) can be calculated using (49).

There are two quantities that still need to be determined, $h_{\hat{x}}$ and R. Using (49) in (16) and (19) gives

$$h_{\hat{x}} = -2\sum_{i=1}^{\frac{N_I}{2}} \frac{f_d^2\left[i\right]}{F_d\left[i\right]} = -I_q \tag{54}$$

$$R = 2\sum_{i=1}^{\frac{N_I}{2}} \frac{f_d^2\left[i\right]}{F_d\left[i\right]} = I_q. \tag{55}$$

The specific gain γ_k and the performance of the algorithm for each model will now be determined.

5.1. Constant X_k

Replacing $h_{\hat{x}}$ given by (54) in (20) and the result in (9) gives the following gains:

$$\gamma_k = \frac{1}{kI_q} \tag{56}$$

and by replacing (54) and (55) in (21), σ^2_∞ is obtained:

$$\sigma^2_\infty = \frac{1}{I_q}. \tag{57}$$

In practice this means that for large k, the estimation variance will be (*cf.* (14))

$$\mathbb{V}\text{ar}\left[\hat{X}_k\right] \approx \frac{1}{kI_q}. \tag{58}$$

The right hand side of (58) is the inverse of the Fisher information for estimating $X_k = x$ based on i_k when the offset is fixed to be x. The inverse of the Fisher information is known as the Cramér–Rao bound and it is a lower bound on the variance of unbiased estimators [10, Ch. 3]. This means that for large k, the estimator has the lowest possible variance within the class of unbiased estimators using quantized observations with offset $b_k = x$.

In the continuous measurement case (infinite number of quantization intervals) the CRB for k observations is given by

$$\text{CRB}_c = \frac{1}{kI_c}, \tag{59}$$

where I_c is the Fisher information given by

$$I_c = \int_{\mathbb{R}} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\,\mathrm{d}x \tag{60}$$

and $f'(x) = \frac{df(x)}{dx}$. In the cases where I_c exists and for large k, one can calculate the loss of estimation performance L_q in decibels (dB) in the following way:

$$\begin{aligned} L_q &= 10\log_{10}\left(\frac{\mathbb{V}\text{ar}\left[\hat{X}_k\right]}{\text{CRB}_c}\right) \\ &= -10\log_{10}\left(\frac{I_q}{I_c}\right). \end{aligned} \tag{61}$$

5.2. Wiener process

Using (55) in (24), the following constant gain is obtained:

$$\gamma^\star = \frac{\sigma_w}{\sqrt{I_q}} \tag{62}$$

and for this gain, the asymptotic MSE is obtained by substituting (57) in (26):

$$\text{MSE}_\infty = \frac{\sigma_w}{\sqrt{I_q}}. \tag{63}$$

The comparison with the continuous case can be done also using a lower bound on the variance. In this case as X_k is random the Bayesian Cramér–Rao bound (BCRB) can be used, this bound is defined as the inverse of the Bayesian information for time k [11, Ch. 1]:

$$\text{BCRB}_k = \frac{1}{J_k}. \tag{64}$$

For a Wiener process, the Bayesian information can be calculated recursively. The recursive expression, given in its general form in [12], for a scalar Wiener process observed with additive noise is

$$J_k = I_c + \frac{1}{\sigma_w^2} - \frac{1}{\sigma_w^4\left(J_{k-1} + \frac{1}{\sigma_w^2}\right)}. \tag{65}$$

The comparison must be done for $k \to \infty$. After calculating the fixed point J_∞ of (65), the asymptotic BCRB obtained is

$$\text{BCRB}_\infty = \frac{2}{I_c + \sqrt{I_c^2 + 4\frac{I_c}{\sigma_w^2}}}. \tag{66}$$

Expression (63) is only valid for small σ_w, in this case (66) can be approximated by

$$\text{BCRB}_\infty \approx \frac{\sigma_w}{\sqrt{I_c}} \tag{67}$$

and the loss in asymptotic performance L_q^W for the estimation of the Wiener process can be approximated by a function of L_q:

$$L_q^W \approx \frac{1}{2}L_q. \tag{68}$$

5.3. Wiener process with drift

The varying optimal gain and the MSE are obtained by replacing (54) and (55) in (28) and (29):

$$\gamma_k^\star = \left[\frac{4u_k^2}{I_q^2}\right]^{\frac{1}{3}} \tag{69}$$

$$\text{MSE}_k \approx 3\left[\frac{|u_k|}{4I_q}\right]^{\frac{2}{3}}. \tag{70}$$

As u_k is unknown, it might be estimated. For slowly varying u_k it can be estimated by smoothing the differences between successive estimates:

$$\hat{U}_k = \hat{U}_{k-1} + \gamma_k^u \left[\left(\hat{X}_k - \hat{X}_{k-1} \right) - \hat{U}_{k-1} \right]. \quad (71)$$

Then, $\hat{U}_k$ can replace u_k in the evaluation of the gain and the MSE. If more information about the evolution of u_k is known, it might be incorporated in (71) to have more precise estimates and get closer to the optimal adaptive gain.

As it is hard to have a bound on performance for the estimation of a deterministic signal under non Gaussian noise, the comparison with the continuous observation case will be done using the approximate performance for a nonlinear adaptive algorithm using continuous observations. The algorithm has the following form:

$$\hat{X}_k = \hat{X}_{k-1} + \gamma_k^c \eta_c \left(Y_k - \hat{X}_{k-1} \right), \quad (72)$$

where γ_k^c and the non linearity $\eta_c(x)$ are optimized to minimize the MSE.

Using the same theory described for the quantized case it is possible to show that the optimal γ_k^c and $\eta_c(x)$ are

$$\gamma_k^c = \left[\frac{4u_k^2}{I_c^2} \right]^{\frac{1}{3}} \quad (73)$$

$$\eta_c(x) = \frac{f'(x)}{f(x)}, \quad (74)$$

which exist under the constraint that I_c converges and is not zero and that $f'(x)$ exists for every x.

The MSE can be approximated in a similar way as before:

$$\text{MSE}_k \approx 3 \left[\frac{|u_k|}{4I_c} \right]^{\frac{2}{3}}. \quad (75)$$

Therefore, the loss in performance incurred by quantizing the observations in the estimation of the Wiener process with drift L_q^{WD} can be approximated by

$$L_q^{WD} \approx \frac{2}{3} L_q. \quad (76)$$

The losses for the three models of X_k depend directly on L_q, thus L_q allows to approximate how much of performance is lost for a specific type of noise and threshold set comparing to the optimal (possibly suboptimal in the case with drift) estimator based on continuous measurements. In the next subsection the loss will be evaluated for two different classes of noise considering that the quantization is uniform, then the adaptive algorithm will be simulated in the three cases and the simulated loss will be compared to the results given above to check their validity.

5.4. Simulation

The thresholds are considered to be uniform and given by (53). For a given type of noise, supposing that δ is known and for fixed N_I, I_q can be evaluated by replacing (53) and (48) in the expressions for f_d and F_d. As I_q is now a function of c_Δ only, it can be maximized by adjusting this parameter. Being a scalar maximization problem this can be done by using grid optimization (searching for the maximum in a fine grid of possible c_Δ). After finding the optimal c_Δ and I_q, the coefficients η_i, the optimal gains γ_k and the quantizer input gain $\frac{1}{\Delta}$ can be evaluated and then all the parameters are defined.

Note that it is supposed that the model for X_k is known as setting γ_k depends on it. As a consequence of this assumption, in a real application the choice between the three models must be clear. When this choice is not clear from the application, it is always simpler to choose X_k to be a Wiener process, first, because the complexity of the algorithm is lower and second, because supposing that the increments are Gaussian and i.i.d. does not impose too much information on the evolution of X_k. Still, σ_w must be known, in practice it can be set based on prior knowledge on the possible variation of X_k or by accepting a slower convergence and a small loss of asymptotic performance, it can be estimated jointly with X_k using an extra adaptive estimator for it. In the last case, when it is known that the increments of X_k have a deterministic component, the fact the γ_k depends on u_k is not very useful and prior information on the variations of X_k are not normally as detailed as knowing u_k itself, making it necessary to accept a small loss of performance to estimate u_k jointly. The estimation of u_k can be done using (71) where prior knowledge on the variations of u_k can be integrated in the gain γ_k^u. If precise knowledge on the evolution of u_k is known through dynamical models, then it might be more useful to use other forms of adaptive estimators known as multi-step algorithms [8, Ch. 4].

The evaluation of the loss and the verification of the results will be done considering two different classes of noise that verify assumptions A1 to A3, namely, generalized Gaussian (GG) noise and Student's-t (ST) noise. The motivation for the use of these two densities comes from signal processing, statistics and information theory.

In signal processing, when additive noise is not constrained to be Gaussian a common assumption is that the noise follows a GG distribution [13]. This distribution not only contains the Gaussian case as an specific example, but also by changing one of its parameters, one can represent from the impulsive Laplacian case to distributions close to the uniform case. In robust statistics, when the additive noise is considered to be impulsive, a general class for the distribution of the noise is the ST distribution [14]. ST distribution includes as a specific case the Cauchy distribution, known to be heavy tailed and thus normally used in robust statistics, also by changing a parameter of the distribution an entire class of heavy tailed distributions can be represented. When looking from an information point of view, if no priors on the noise distributions are given, noise models must be as random as possible to ensure that the noise is an uninformative part of the observation, thus noise models must maximize some criterium of randomness. Commonly used criteria for randomness are entropy measures and both distributions considered above are entropy maximizers. GG distributions maximize the Shannon entropy under constraints on the moments [15, Ch. 12] and ST distributions maximize the Rényi entropy under constraints on the second order moment [16].

Both distributions are parametrized by a shape parameter $\beta \in \mathbb{R}^+$ and their PDFs and CDFs for $\delta = 1$ are

$$f_{GG}(x) = \frac{\beta}{2\Gamma\left(\frac{1}{\beta}\right)} \mathrm{e}^{-|x|^\beta}, \tag{77}$$

$$F_{GG}(x) = \frac{1}{2}\left[1 + \mathrm{sign}(x)\frac{\gamma\left(\frac{1}{\beta}, |x|^\beta\right)}{\Gamma\left(\frac{1}{\beta}\right)}\right] \tag{78}$$

for the GG distribution, where $\gamma(\cdot,\cdot)$ is the incomplete gamma function and $\Gamma(\cdot)$ is the gamma function,

$$f_{ST}(x) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\sqrt{\beta\pi}\,\Gamma\left(\frac{\beta}{2}\right)}\left(1 + \frac{1}{\beta}x^2\right)^{-\frac{\beta+1}{2}} \tag{79}$$

$$F_{ST}(x) = \frac{1}{2}\left\{1 + \mathrm{sign}(x)\left[1 - I_{\frac{\beta}{x^2+\beta}}\left(\frac{\beta}{2}, \frac{1}{2}\right)\right]\right\} \tag{80}$$

for the ST distribution, where $I_{\frac{\beta}{x^2+\beta}}(\cdot,\cdot)$ is the incomplete beta function.

5.4.1. *Performance loss* - L_q

The first quantity to be evaluated will be the loss L_q. To evaluate L_q, after evaluating I_q based on f and F defined above, it is also needed to evaluate I_c. Evaluating the integral on (60), one obtains for the GG and ST distributions respectively:

$$I_{GG} = \frac{\beta(\beta-1)\Gamma\left(1-\frac{1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} \tag{81}$$

$$I_{ST} = \frac{\beta+1}{\beta+3}. \tag{82}$$

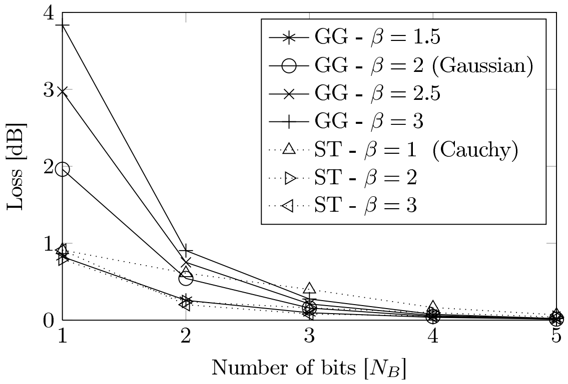

Figure 3: Loss of performance due to quantization of measurements for different types of noise and number of quantization bits.

The loss was evaluated for $N_I = \{2, 4, 8, 16, 32\}$, which corresponds to numbers of bits $N_B = \{1, 2, 3, 4, 5\}$, and for the shape parameters $\beta = \{1.5, 2, 2.5, 3\}$ for GG noise and $\beta = \{1, 2, 3\}$ for ST noise. The results are shown in Fig. 3. As it was expected, the loss reduces with increasing N_B. It is interesting to note that the maximum loss, observed for $N_B = 1$, goes from approximately 1dB to 4dB, which represents factors less than 3 in MSE increase for estimating a constant with 1 bit quantization. Also interesting is the fact that the loss decreases rapidly with N_B, for 2 bits quantization all the tested types of noise produce losses below 1dB, resulting in linear increases in MSE not larger than 1.3. This indicates that when using the adaptive estimators developed here, it is not very useful to use more than 4 or 5 bits for quantization.

The performance for 2 bits seems to be related to the noise tail, note that smaller losses were obtained for distributions with heavier tail (ST distributions and GG distribution with $\beta = 1.5$), this is due to the fact that for large tail distributions a small region around the median of the distribution is very informative, thus as most of the information is contained there, when the only threshold available is placed there, the relative gain of information is greater than in the other cases, leading to smaller losses. This can also be the reason for the slow

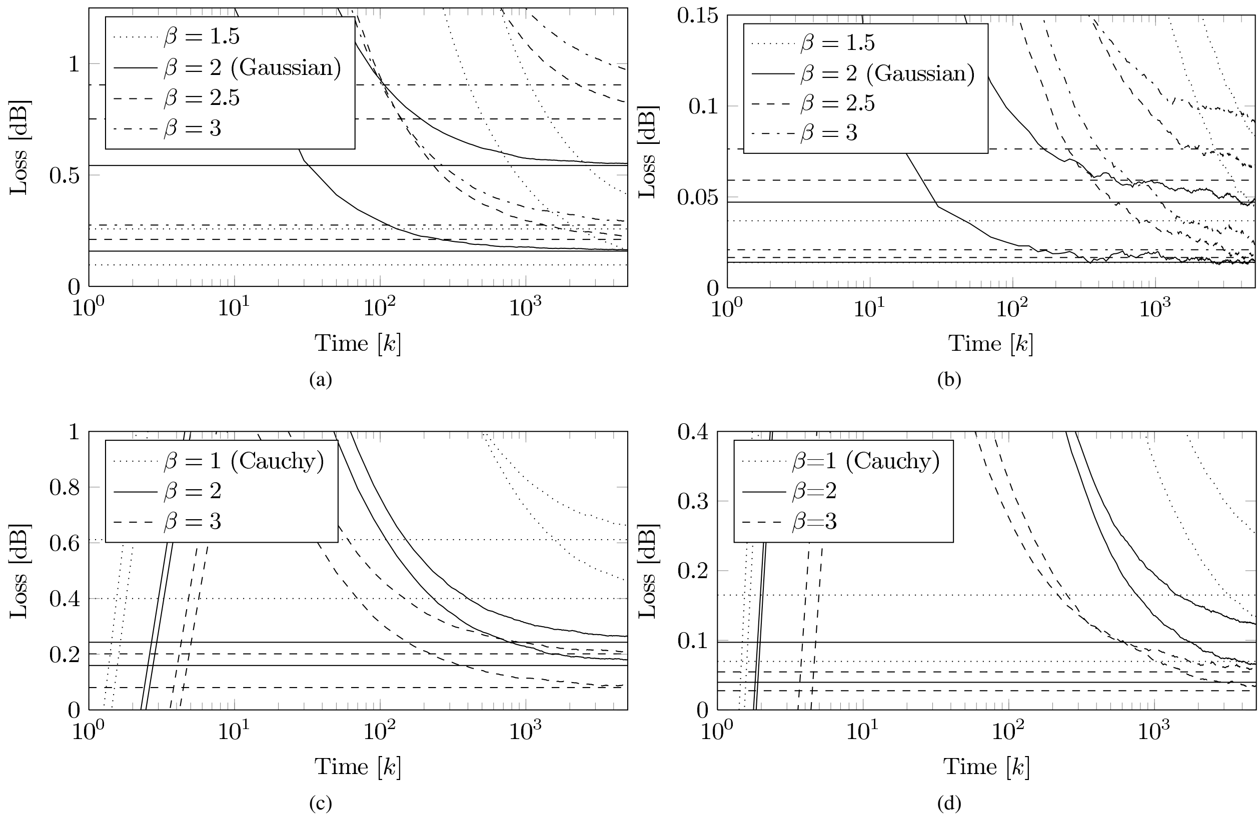

Figure 4: **Constant.** Quantization loss of performance for GG and ST noises and $N_B = \{2, 3, 4, 5\}$ when X_k is constant. For each type of noise there are 4 curves, the constant losses are the theoretical results and the decreasing losses are the simulated results, thus producing pairs of curves of the same type, for each pair the higher results represent lower number of quantization bits. In (a) results for GG noise and $N_B = 2$ and 3, in (b) the results for GG noise and $N_B = 4$ and 5 are shown. The figures (c) and (d) are the results for ST noise, in (c) $N_B = 2$ and 3 are considered while in (d) $N_B = 4$ and 5.

decrease of the loss for these distributions, as the quantizer thresholds are placed uniformly, some of them will be placed in the non informative amplitude region and consequently the decrease in loss will be not as sharp as in the other cases.

Laplacian distribution was not tested, because for this distribution the optimal adaptive estimator in the continuous case is already an adaptive estimator with a binary quantizer. This can be seen easily if one evaluates I_q as a function of the thresholds, the result will be a constant for all possible sets of thresholds meaning that they are unimportant, moreover, if η_i are evaluated one will find that they are all equal, therefore only the sign of the difference between the observations and the last estimate is important. Consequently, the loss found in this case would be a constant for all N_B.

To validate the results, the adaptive algorithms will be simulated and the loss obtained will be compared to the approximations given above. The simulation results will be presented in the same order as before, first the constant case, then the Wiener process case and finally the case with drift. All the simulation were done considering $N_B = \{2, 3, 4, 5\}$.

5.4.2. Simulated loss - ***Constant***

In the constant case, the 7 types of noise with evaluated L_q were tested, the value of $X_0 = x$ was set to be zero and the initial condition of the adaptive algorithm was set with a small error ($\hat{X}_0 \in \{0, 10\}$), the number of samples was set to be 5000 to have sufficient points for convergence, the algorithm was simulated 2.5×10^6 times and the error results were averaged to produce a simulated MSE. Based on the simulated MSE a simulated loss was calculated, and it is shown in Fig. 4.

The simulated results seems to converge to the theoretical approximations of L_q, thus validating these approximations. This also means that the variance of estimation tends in simulation to the CRB for quantized

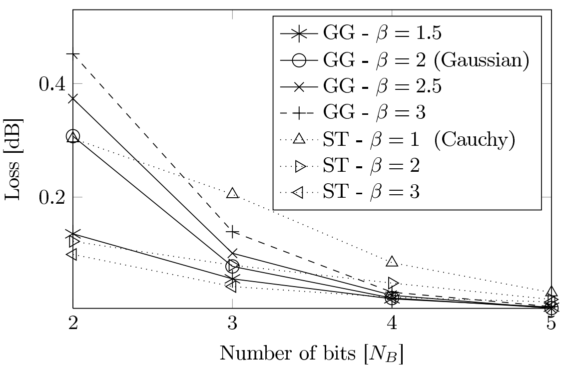

Figure 5: **Wiener process.** Simulated quantization performance loss for a Wiener process X_k with $\sigma_w = 0.001$, different types of noise and number of quantization bits.

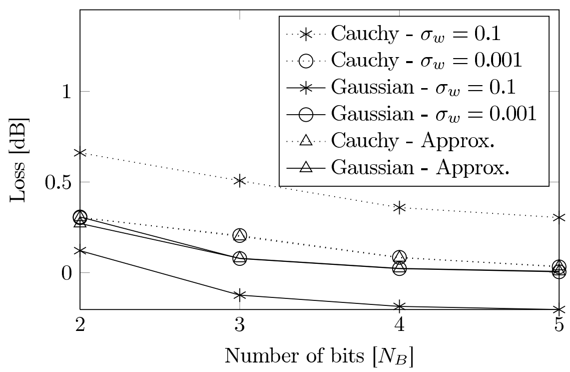

Figure 6: **Wiener process.** Comparison of simulated and approximate (Approx.) theoretical losses (68) in the Gaussian and Cauchy noise cases when estimating a Wiener process with $\sigma_w = 0.1$ or $\sigma_w = 0.001$.

observations, validating the fact that the algorithm is asymptotically optimal. The convergence time looks to be related to N_B, when N_B increases the time to get closer to the optimal performance decreases.

In practice, if increased initial complexity can be accepted, a robust maximum likelihood estimator that uses blocks of measurements with different thresholds [17] can be used to generate an initial estimate $\hat{X}_0$ close to x, allowing a faster convergence of the algorithm.

*5.4.3. Simulated loss -**Wiener process***

For a Wiener process, L_q was evaluated by setting $\hat{X}_0$ randomly around 0 and $X_0 = 0$, then 10^4 realizations with 10^5 samples were simulated and the MSE was estimated by averaging the realizations of the squared error for each instant, then as it was observed that the error was approximately stationary after $k = 1000$, the sample mean squared error was also averaged resulting in an estimate of the asymptotic MSE. Based on the obtained values of the MSE a simulated loss was evaluated. The results for the 7 types of noise and $\sigma_w = 0.001$ are shown in Fig. 5.

As expected, the results have the same form of the theoretical loss given in Fig. 3. To verify the results for different σ_w, the loss was evaluated through simulation also for $\sigma_w = 0.1$ in the Gaussian (GG with $\beta = 2$) and Cauchy cases (ST with $\beta = 1$). The results are shown in Fig. 6, where the approximate theoretical losses (68) for these cases are also shown. It is clear from the results that X_k might move slowly to give a performance close to the theoretical results, but it is also interesting that the simulated loss seems to have the same decreasing rate as a function of N_B when compared to the theoretical results. This means that the dependence on I_q of the MSE seems to still be correct and it indicates that even

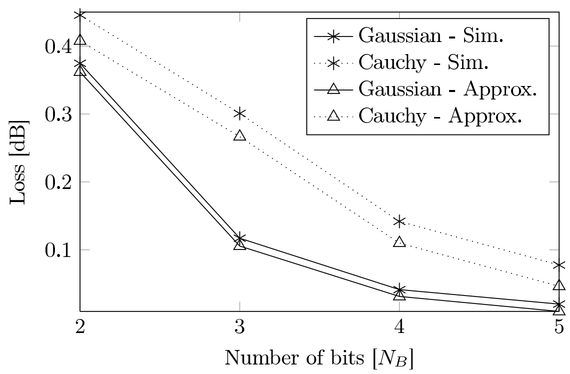

Figure 7: **Wiener process with drift.** Comparison of simulated and approximate (Approx.) theoretical losses (76) in the Gaussian and Cauchy noise cases for estimating a Wiener process with constant mean drift $u_k = 10^{-4}$ and standard deviation $\sigma_w = 10^{-4}$.

in a faster regime for X_k, the thresholds can be set by maximizing I_q.

*5.4.4. Simulated loss - **Wiener process with drift***

For X_k with drift, W_k was simulated with mean and standard deviations $u_k = \sigma_w = 10^{-4}$, which represents a slow linear drift with small random fluctuations, the initial conditions were set to be $X_0 = \hat{X}_0 = 0$ and the drift estimator was set with constant gain $\gamma_k^u = 10^{-5}$. Its initial condition was set to the true u_k to reduce the transient time and consequently the simulation time. As u_k is constant, the loss evaluation was done in the same form as for X_k without drift, based on averaging through realizations and time. The results for the Gaussian and Cauchy cases are shown in Fig. 7.

The small offset between simulated and theoretical results is produced by the joint estimation of u_k. Note that keeping γ_k^u to a small constant allows to adaptively follow slow variations in u_k. The convergence to the simulated loss in Fig. 7 was also obtained for simulations with errors in the initial conditions but in this case the transient regime was very long, as the estimator of u_k took a very long time to converge, indicating that other schemes might be considered when the theoretical performance is needed in a short period of time. Multi-step adaptive algorithms could be used for faster convergence to the theoretical performance but they would need a precise model for the evolution of the drift which is not considered here.

In the ideal case of known drift, under the simulation conditions described above, no offset was observed between the simulated asymptotic loss and the approximate theoretical results.

6. Conclusions

In this work an adaptive estimation algorithm based on quantized observations was proposed. Based on observations with additive noise and quantized with adjustable offset and gain, the objective was to estimate with a low complexity online adaptive algorithm a scalar parameter that could follow one of three models: constant, Wiener process and Wiener process with drift. Under the hypothesis that the noise PDF is symmetric and strictly decreasing, and that quantizer is also symmetric, by using Lyapunov theory it was shown that for the optimal quantizer output coefficients, the algorithm is asymptotically stable. It was also shown that the asymptotic performance in terms of mean squared error could be optimized by using static update coefficients that depend only on the shape of the observation noise and on the quantizer thresholds.

Performance results were obtained based on the optimal choice of the quantizer output levels. It was observed that the effect of quantization on performance could be quantified by the Fisher information of the quantized observations. Thus, this clearly indicates that the quantizer thresholds must be placed to maximize the Fisher information. It was also observed that for the three models, the loss of performance of the algorithm w.r.t. the optimal continuous measurement is given by a function of the ratio of the corresponding Fisher informations.

For testing the results, two different families of noise were considered, generalized Gaussian noise and Student's-t noise, both under uniform quantization. First, the theoretical loss was evaluated for different numbers of quantization intervals. The results indicate that with only a few quantization bits (4 and 5) the adaptive algorithm performance is very close to the continuous observation case and it was observed that uniform quantization seems to penalize more estimation performance under heavy tailed distributions.

Estimation in the three possible scenarios was simulated and the results validated the accuracy of the theoretical approximations. In the constant case it was observed that the algorithm performance was very close to the Cramér–Rao bound, in the Wiener process case it was observed that the theoretical results are very accurate for small increments of the Wiener process and in the drift case it was seen that by accepting a small increase in the mean squared error it is possible to estimate jointly the drift.

Another interesting result is that a varying parameter has a loss of performance smaller than a constant parameter, thus a type of dithering effect seems to be present. In this case, the variations of the input signal makes the tracking performance of the estimator to get close to the continuous measurement performance.

The fact that the number of quantization bits does not influence much the performance of estimation leads to conclude that it seems more reasonable to focus on using more sensors than using high resolution quantizers for increasing performance. Consequently, this motivates the use of sensor network approaches.

As the Fisher information for quantized measurements plays a central role in the performance of the algorithms, the study of its properties as a function of the noise type and quantizer thresholds seems to be a subject for future work. A possible approach for the study of its general behavior would be to consider high resolution approximations.

Finally, as in practice sensor noise scale parameter and Wiener process increment standard deviation can be unknown and slowly variable, it would be also interesting to study how the algorithm design and performance would change by estimating all these parameters jointly.

Acknowledgment

The authors would like to thank Eric Moisan, Steeve Zozor and Olivier J. J. Michel for their helpful comments and the Erasmus Mundus EBWII program for funding this study.

[1] C. Chong, S. Kumar, Sensor networks: Evolution, opportunities, and challenges, Proceedings of the IEEE 91 (8) (2003) 1247–1256.

[2] A. Gersho, R. Gray, Vector quantization and signal compression, Springer, 1992.

[3] H. Papadopoulos, G. Wornell, A. Oppenheim, Sequential signal encoding from noisy measurements using quantizers with dynamic bias control, IEEE Trans. Inf. Theory 47 (3) (2001) 978–1002.

[4] A. Ribeiro, G. Giannakis, Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case, IEEE Trans. Signal Process. 54 (3) (2006) 1131–1143.

[5] H. Li, J. Fang, Distributed adaptive quantization and estimation for wireless sensor networks, IEEE Signal Process. Lett. 14 (10) (2007) 669–672.

[6] J. Fang, H. Li, Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood, IEEE Trans. Signal Process. 56 (10) (2008) 5246–5257.

[7] L. Wang, G. Yin, J. Zhang, Y. Zhao, System identification with quantized observations, Birkhauser, 2010.

[8] A. Benveniste, M. Métivier, P. Priouret, Adaptive algorithms and stochastic approximations, Springer-Verlag New York, Inc., 1990.

[9] H. Khalil, J. Grizzle, Nonlinear systems, Macmillan Publishing Company New York, 1992.

[10] S. Kay, Fundamentals of statistical signal processing, Volume 1: Estimation theory, PTR Prentice Hall, 1993.

[11] H. L. Van Trees, K. L. Bell, Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking, Wiley-IEEE Press, 2007.

[12] P. Tichavsky, C. Muravchik, A. Nehorai, Posterior Cramér–Rao bounds for discrete-time nonlinear filtering, IEEE Trans. Signal Process. 46 (5) (1998) 1386 –1396.

[13] M. Varanasi, B. Aazhang, Parametric generalized Gaussian density estimation, The Journal of the Acoustical Society of America 86 (1989) 1404–1415.

[14] K. Lange, R. Little, J. Taylor, Robust statistical modeling using the t distribution, Journal of the American Statistical Association (1989) 881–896.

[15] T. M. Cover, J. A. Thomas, Elements of Information Theory 2nd Edition, Wiley-Interscience, 2006.

[16] J. Costa, A. Hero, C. Vignat, On solutions to multivariate maximum α-entropy problems, in: A. Rangarajan, M. Figueiredo, J. Zerubia (Eds.), Energy Minimization Methods in Computer Vision and Pattern Recognition, Vol. 2683 of Lecture Notes in Computer Science, Springer Berlin/Heidelberg, 2003, pp. 211–226.

[17] X. Shen, P. K. Varshney, Y. Zhu, Robust distributed maximum likelihood estimation with quantized data, arXiv:1208.4161v2 (December 2012).