



**HAL**  
open science

# Estimation sous contraintes de communication : algorithmes et performances asymptotiques

Rodrigo Cabral Farias

► **To cite this version:**

Rodrigo Cabral Farias. Estimation sous contraintes de communication : algorithmes et performances asymptotiques. Autre. Université de Grenoble, 2013. Français. NNT : 2013GRENT024 . tel-00877073v2

**HAL Id: tel-00877073**

**<https://theses.hal.science/tel-00877073v2>**

Submitted on 23 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **signal, image, parole, télécommunications (SIPT)**

Arrêté ministériel : 7 août 2006

Présentée par

**M. Rodrigo CABRAL FARIAS**

Thèse dirigée par **M. Jean-Marc BROSSIER**

préparée au sein du

**laboratoire Grenoble, images, parole, signal, automatique (GIPSA-lab)**

dans l'école doctorale d'électronique, électrotechnique, automatique et traitement du signal (EEATS)

## Estimation sous contraintes de communication : algorithmes et performances asymptotiques

Thèse soutenue publiquement le **17/07/2013**,  
devant le jury composé de :

**M. Eric MOULINES**

Professeur Télécom ParisTech, Rapporteur

**M. Jean-Yves TOURNERET**

Professeur ENSEEIHT, Rapporteur

**M. Josep VIDAL**

Professeur Universitat Politècnica de Catalunya, Examineur

**M. Jean-François BERCHER**

Professeur associé ESIEE, Examineur

**M. Eric MOREAU**

Professeur Université du Sud Toulon-Var, Examineur

**M. Jean-Marc BROSSIER**

Professeur Grenoble-INP, Directeur de thèse





# Acknowledgements

I would like to thank the Erasmus program Euro Brazilian Windows II for funding this thesis and the director of the GIPSA-lab, Jean-Marc Thiriet, for welcoming me in his laboratory. I would like to express my gratitude to my thesis director, Jean-Marc Brossier, for allowing me to be a free researcher during my thesis, pointing carefully the mistakes in some of my strange ideas and motivating me to look deeper when the ideas were not that strange. Special thanks are extended to the professors Eric Moisan, Laurent Ros, Olivier Michel and Steeve Zozor, who helped me countless times.

Also, I would like to thank the members of my jury Eric Moreau, Eric Moulines, Jean-Yves Tournieret, Josep Vidal and Jean-François Bercher for their precious remarks on my work and for all the insights on how I can extend it.

During these three years, I bothered a non negligible quantity of PhD students in the laboratory. I would like to acknowledge those who have survived to this torture for their patience. Thanks to the survivors: Aude, Damien, Douglas, Gailene, Humberto, Jonathan, Robin, Wei, Xuan Vu and Zhong Yang.

I am particularly grateful for the assistance given by Vio, who was able to encourage me to move forward, even when I felt the pressure was excessive.

To my family, mãe, Dea e vó, I can only express that it was very difficult to stay such a long time far from you.



# Contents

<b>Notations</b>	<b>11</b>
<b>Abbreviations and acronyms</b>	<b>13</b>
<b>Assumptions</b>	<b>15</b>
<b>Introduction</b>	<b>17</b>
<b>I Estimation based on quantized measurements: algorithms and performance</b>	<b>25</b>
<b>1 Estimation of a constant parameter</b>	<b>31</b>
1.1 Measurement model . . . . .	34
1.2 Maximum likelihood, Cramér–Rao bound and Fisher information . . . . .	37
1.3 Binary quantization . . . . .	44
1.4 Multibit quantization . . . . .	54
1.5 Adaptive quantizers: the high complexity fusion center approach . . . . .	60
1.6 Chapter summary and directions . . . . .	73
<b>2 Estimation of a varying parameter</b>	<b>75</b>
2.1 Parameter and measurement model . . . . .	77
2.2 Optimal estimator . . . . .	78
2.3 Particle Filtering . . . . .	81
2.4 Evaluation of the estimation performance . . . . .	87
2.5 Quantized innovations . . . . .	90
2.6 Chapter summary and directions . . . . .	102
<b>3 Adaptive quantizers for estimation</b>	<b>105</b>
3.1 Parameter model and measurement model . . . . .	108
3.2 General estimation algorithm . . . . .	111
3.3 Estimation performance . . . . .	113
3.4 Optimal algorithm parameters and performance . . . . .	125

3.5	Simulations . . . . .	135
3.6	Adaptive quantizers for estimation: extensions . . . . .	149
3.7	Chapter summary and directions . . . . .	164
<b>Conclusions of Part I</b>		<b>169</b>
<b>II Estimation based on quantized measurements: high-rate approximations</b>		<b>171</b>
<b>4</b>	<b>High-rate approximations of the FI</b>	<b>177</b>
4.1	Asymptotic approximation . . . . .	180
4.2	Bit allocation for scalar location parameter estimation . . . . .	200
4.3	Generalization with the $f$ -divergence . . . . .	207
4.4	Chapter summary and directions . . . . .	213
<b>Conclusions of Part II</b>		<b>217</b>
<b>Conclusions</b>		<b>219</b>
	Main conclusions . . . . .	219
	Perspectives . . . . .	220
<b>A Appendices</b>		<b>223</b>
A.1	Why? - Proofs . . . . .	223
A.2	More? - Further details . . . . .	235
A.3	How? - Algorithms and implementation issues . . . . .	248
<b>B Résumé détaillé en français (extended abstract in French)</b>		<b>253</b>
B.1	Introduction . . . . .	254
B.2	Estimation et quantification : algorithmes et performances . . . . .	261
B.3	Estimation et quantification : approximations à haute résolution . . . . .	286
B.4	Conclusions . . . . .	293
<b>Bibliography</b>		<b>295</b>

# List of Figures

1	Estimation using a sensing system. . . . .	20
2	Scalar remote sensing problem. . . . .	21
3	Estimation based on quantized measurements. . . . .	21
1.1	Quantizer function $Q(Y_k)$ with $N_I$ quantization intervals and uniform threshold spacing with length $\Delta$ . . . . .	35
1.2	Scheme representing the general measurement/estimation system. . . . .	42
1.3	Quantity related to the CRB for quantized measurements $B$ and its upper bound $\bar{B}$ . . . . .	45
1.4	Function $M \times \delta^2$ . . . . .	47
1.5	PDF for the uniform/Gaussian distribution. . . . .	49
1.6	CRB $_q^B$ and simulated MLE MSE for uniform/Gaussian noise. . . . .	50
1.7	CRB $_q^B$ and simulated MLE MSE for GGD noise. . . . .	52
1.8	FI as a function of the normalized difference between the central threshold and the true parameter. . . . .	57
3.1	Scheme representing the adjustable quantizer. . . . .	110
3.2	Block representation of the estimation scheme. . . . .	111
3.3	ODE bias approximation and simulated bias for the estimation of a Wiener process with the adaptive algorithm. . . . .	118
3.4	Adaptive algorithm loss of estimation performance due to quantization of measurements. . . . .	138
3.5	Quantization loss of performance for GGD noise and $N_B \in \{2, 3, 4, 5\}$ when $X_k$ is constant. . . . .	139
3.6	Quantization loss of performance for STD noise and $N_B \in \{2, 3, 4, 5\}$ when $X_k$ is constant. . . . .	140
3.7	Simulated quantization performance loss for a Wiener process $X_k$ with $\sigma_w = 0.001$ . . . . .	141
3.8	Comparison of simulated and theoretical losses in the Gaussian and Cauchy noise cases when estimating a wiener process with $\sigma_w = 0.1$ or $\sigma_w = 0.001$ . . . . .	142
3.9	Comparison of simulated and theoretical losses in the Gaussian and Cauchy noise cases for estimating a Wiener process with constant mean drift. . . . .	143
3.10	Minimum CRB and simulated MSE for the adaptive algorithm with decreasing gain and for the adaptive algorithm based on the MLE. . . . .	144

---

3.11	Asymptotic MSE for the optimal estimator of a Wiener process with small $\sigma_w$ and simulated MSE for the adaptive algorithm with constant gain and for the PF with dynamic central threshold. . . . .	146
3.12	Scheme representing the adjustable quantizer. The offset and gain are adjusted dynamically. . . . .	151
3.13	CRB for estimating a location parameter of Gaussian and Cauchy distributions based on quantized and continuous measurements and simulated MSE for the estimation of the location parameter with the adaptive location-scale parameter estimator. . . . .	156
3.14	Scheme representing the sensor network with a fusion center. . . . .	158
3.15	Cramér–Rao bound and simulated MSE for the adaptive algorithm in the fusion center approach with different numbers of sensors and 4 quantization intervals. . . . .	162
3.16	Cramér–Rao bound and simulated MSE for the adaptive algorithm with different numbers of sensors and fixed total number of bits. . . . .	163
4.1	Interval densities for the estimation of a GGD location parameter. . . . .	192
4.2	Simulated MSE for the adaptive algorithm with nonuniform thresholds considering Gaussian and Cauchy measurement distributions. . . . .	199
4.3	Water-filling solutions for multicarrier modulation power allocation and for rate constrained sensing system bit allocation. . . . .	206
A.1	Geometric scheme to show that the probability of the interval $A_0 + A_1$ is less than the probability of the exterior region of the left quarter circle $C_1$ . . . . .	225
A.2	Log-likelihood function for Cauchy noise distribution. . . . .	235
A.3	An iteration of the binary threshold update in a finite grid. . . . .	238

# List of Tables

4.1	FI for the estimation of Gaussian and Cauchy location parameters based on quantized measurements. . . . .	198
4.2	Functions characterizing the GFD for different inference problems and interval densities maximizing the inference performance based on quantized measurements. . . . .	212



# Notations

## Sets

$\mathbb{N}$	Natural numbers
$\mathbb{R}$	Real numbers
<b>Sub</b> <sub>+</sub>	Set with only positive elements
<b>Sup</b> <sup>*</sup>	Set including zero

## Vector and sequences

<b>X (boldface)</b>	Vector or matrix
<b>Sup</b> <sup>⊤</sup>	Transposition
<b>diag</b> ( <b>X</b> )	Diagonal matrix from <b>X</b>
$X_{1:N}$	Sequence $X_1, X_2, \dots, X_N$

## Probability

$X$ ( <b>uppercase</b> )	Random variable	$\mathbb{V}\mathbf{ar}(X)$	Variance
$x$ ( <b>lowercase</b> )	Realization or parameter	$\mathbb{V}\mathbf{ar}_{X  \cdot}(X)$	Conditional variance
$\mathbb{P}(\cdot)$	Probability measure	$f(\cdot)$ <b>or</b> $p(\cdot)$	Probability density function
$\mathbb{P}(\cdot \cdot)$	Conditional probability	$f(\cdot \cdot)$	Conditional density
$\mathbb{E}(X)$	Expectation	$F(\cdot)$	Cumulative distribution function
$\mathbb{E}_{X  \cdot}(X)$	Conditional expectation	$f(\cdot; x)$	Parametrization by $x$

Whenever the random variables related to a function are not clear from the context, they are indicated implicitly by the variable in the argument of the function.

## Main variables and parameters

$X$	Unknown parameter	$L$	Likelihood
$Y$	Continuous measurement	$S$	Score function
$i$	Quantized measurement	$I$	Fisher information
$\hat{X}$	Estimator	<b>CRB</b>	Cramér–Rao bound
$V$	Measurement noise	<b>BCRB</b>	Bayesian Cramér–Rao bound
$W$	Parameter variation	<b>MSE</b>	Mean squared error

## Main variables and parameters

$\sigma_w$	Increments variance	$N_I$	Number of quantization intervals
$\delta$	Noise scale	$N_B$	Number of quantization bits
$\varepsilon$	Estimation error	$N$	Number of samples
	Quantizer shift	$N_s$	Number of sensors
$k$	Sample/time index	$\tau$	Quantization threshold
$L_q$	Loss due to quantization	$\tau'$	Threshold variation
$u$	Deterministic drift	$q$	Quantization interval
$\gamma(\mathbf{scalar})$	Adaptive algorithm step	$\Delta$	Interval length
$\Gamma(\mathbf{matrix})$	Adaptive algorithm step		Quantizer input parameter
$\eta$	Adapt. algorithm correction	$\lambda(\cdot)$	Interval density

The subscript  $k$  can be used either to indicate one specific sample of a sequence or to make explicit that a quantity is a sequence.

Most quantities related to continuous measurements have a subscript  $c$  and most quantities related to quantized measurements have a subscript  $q$ .

## Functions

$(\cdot)^{-1}$	Inverse of a function
$\mathbf{sign}(\cdot)$	Sign function
$\Gamma(\cdot)$	Gamma function
$\gamma(\cdot, \cdot)$	Incomplete gamma function
$B$	Beta function
$I_\cdot(\cdot, \cdot)$	Incomplete beta function
$(\cdot)^+$	Ramp function

# Abbreviations and acronyms

<b>AUV</b>	<i>Autonomous underwater vehicle</i>
<b>BCRB</b>	<i>Bayesian Cramér–Rao bound</i>
<b>BI</b>	<i>Bayesian information</i>
<b>CRB</b>	<i>Cramér–Rao bound</i>
<b>CDF</b>	<i>Cumulative distribution function</i>
<b>DSP</b>	<i>Digital signal processing</i>
<b>FI</b>	<i>Fisher information</i>
<b>GFD</b>	<i>Generalized <math>f</math>-divergence</i>
<b>GGD</b>	<i>Generalized Gaussian distribution</i>
<b>i.i.d.</b>	<i>Independent and identically distributed</i>
<b>KLD</b>	<i>Kullback–Leibler divergence</i>
<b>MAP</b>	<i>Maximum a posteriori estimator</i>
<b>MLE</b>	<i>Maximum likelihood estimator</i>
<b>MSE</b>	<i>Mean squared error</i>
<b>MMSE</b>	<i>Minimum mean squared error</i>
<b>ODE</b>	<i>Ordinary differential equation</i>
<b>PF</b>	<i>Particle filter</i>
<b>PDF</b>	<i>Probability density function</i>
<b>r.v.</b>	<i>Random variable</i>
<b>RHS</b>	<i>Right-hand side</i>
<b>STD</b>	<i>Student’s-<math>t</math> distribution</i>
<b>w.r.t.</b>	<i>With respect to</i>



# Assumptions

## Assumptions (on the noise distribution):

**AN1** The marginal CDF of the noise, denoted  $F$ , admits a PDF  $f$  w.r.t. the standard Lebesgue measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

**AN2** The PDF  $f(v)$  is a strictly positive even function and it strictly decreases w.r.t.  $|v|$ .

**AN3**  $F$  is locally Lipschitz continuous.

## Assumptions (on the quantizer):

**AQ1**  $N_I$  is considered to be an even natural number and the set  $\mathcal{I}$  where  $i_k$  is defined is

$$\mathcal{I} = \left\{ -\frac{N_I}{2}, \dots, -1, 1, \dots, \frac{N_I}{2} \right\}.$$

**AQ2** The quantizer is symmetric around the central threshold. This means that the vector of thresholds  $\boldsymbol{\tau}$  is given by

$$\boldsymbol{\tau} = \left[ \tau_{-\frac{N_I}{2}} = \tau_0 - \tau'_{\frac{N_I}{2}} \cdots \tau_{-1} = \tau_0 - \tau'_1 \quad \tau_0 \quad \tau_1 = \tau_0 + \tau'_1 \cdots \tau_{\frac{N_I}{2}} = \tau_0 + \tau'_{\frac{N_I}{2}} \right]^\top$$

with the threshold vector elements forming a strictly increasing sequence and the non-negative vector of threshold variations w.r.t. the central threshold given by

$$\boldsymbol{\tau}' = \left[ \tau'_0 = 0 \quad \tau'_1 \quad \cdots \quad \tau'_{\frac{N_I}{2}} = +\infty \right]^\top.$$

**AQ3** The quantizer output levels have odd symmetry w.r.t.  $i$ :

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

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

## Modified assumptions (on the quantizer):

**AQ2'** The quantizer is symmetric around the central threshold which is equal to zero. This means that the vector of thresholds  $\boldsymbol{\tau}$  is given by the vector of threshold variations

$$\boldsymbol{\tau} = \left[ -\tau'_{\frac{N_I}{2}} \cdots -\tau'_1 \quad 0 \quad +\tau'_1 \cdots +\tau'_{\frac{N_I}{2}} \right]^\top,$$

where the threshold variations  $\tau'_i$  form an increasing sequence.

**AQ3'** The quantizer output levels  $\eta_x[i]$  are odd and the output levels  $\eta_\delta[i]$  are even.

$$\eta_x[i] = -\eta_x[-i], \quad \eta_\delta[i] = \eta_\delta[-i],$$

with  $\eta_x[i] > 0$  for  $i > 0$  and  $\eta_\delta[1] < 0$ .

**Assumptions on  $I_q$  for the MLE update to have asymptotically optimal performance:**

**A1.MLE**  $I_q(\varepsilon)$  is maximum for  $\varepsilon = 0$ .

**A2.MLE**  $I_q(\varepsilon)$  is locally decreasing around zero.

**A3.MLE** The function  $I_q(\varepsilon)$  has bounded  $I_q(0)$ ,  $\left. \frac{dI_q(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = 0$ , bounded  $\left. \frac{d^2 I_q(\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0}$ , therefore accepting a Taylor approximation around zero (for small  $\varepsilon'$ ):

$$I_q(\varepsilon') = I_q(0) + \frac{\varepsilon'^2}{2} \left. \frac{d^2 I_q(\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0} + o(\varepsilon'^2),$$

where the  $o(\varepsilon'^2)$  here is equivalent to say that the quantity  $\frac{o(\varepsilon'^2)}{\varepsilon'^2}$  tends to zero when  $\varepsilon'$  tends to zero.

# Introduction

## Quantization: the stranger in the room

Open a book, any basic book on **digital signal processing (DSP)**, and count the number of pages dedicated to the sampling theorem and discrete-time signal processing: FFT, Z-transform, FIR and IIR filtering. Now, count the number of pages dedicated to quantization. Even if half of the "digital world" comes from quantization, by reading some basic books on DSP, we have the feeling that it is a completely unimportant subject.<sup>1</sup>

A curious person might think: is it really unimportant? Maybe it is simply so difficult to be treated and explained in an easy way, that most DSP books skip a detailed description of quantization. We think this explanation is the reason most of the texts presenting DSP assume that signals are quantized with a very high resolution, so they have the possibility of explaining quantization almost in a footnote. As a consequence, quantization seems to be the stranger that comes to the "DSP party" and almost nobody wants to speak with (even if it is one of the party organizers). Some signal processing domains find useful (and in some circumstances they are not wrong) to refuse "contact" with quantization. Whenever they need to address quantization issues they always call it in a derogatory way – "quantization noise".

In this thesis we expect to make one of the subjects in the signal processing party to "talk" with quantization in a polite way, without detracting terms. The subject we chose is estimation.

In the following, we will explain the motivation and the main points of their "conversation".

## Sensor networks and quantization: the welcome guest

Although we do not explicitly design estimation algorithms using a sensor network architecture, this thesis is intended to contribute in the development of estimation techniques that can be applied or extended to sensor networks.

**Sensor network emergence.** With the reduction in cost and size of electronic devices such as sensors and transceivers, a whole new field emerged under the name Sensor Networks. This term, in general, means any set of sensors capable of communication and processing used for a specific task, *e.g.* estimation, detection, tracking, classification, etc.

Sensor networks are attractive for many reasons [Akyildiz 2002], [Intanagonwiwat 2000], [Zhao 2004, pp. 7–8]:

---

<sup>1</sup>Note that the real problem of digitizing a signal by considering sampling and quantization as a joint operation is simply a non-issue in signal processing literature. We do not study this problem in this thesis either, but it is an interesting problem.

- *fault-tolerance and flexibility.* By using multiple sensors to realize a sensing task, even if one of them is unable to measure, the other sensors guarantee that the sensing system is still working. By proper design, the sensor network can reconfigure the way it operates, so that if a failure occurs in a sensor or a small set of sensors, the performance of the sensing system is not strongly affected.
- *Easy deployment.* The decreased cost of the sensors makes it possible to deploy large quantities of sensors in a given area without detailed placement of the sensors. This simplifies the deployment of sensing systems in difficult access and hostile environments.
- *Risky environment sensing.* By allowing the sensors to communicate wirelessly, remote sensing can be done in areas where human activity is impossible or cannot be sustained for long periods of time.
- *No maintenance sensing.* The fault tolerance capabilities of sensor networks allows it to be applied in applications where maintenance of the sensing system is difficult.
- *Multi-hop communication.* By using the communication capabilities of the sensors to allow multi-hop communication, the total energy used in communication for the sensing task may decrease, as the attenuation of transmitted signals is smaller for smaller distances.
- *Enhanced signal-to-noise ratio.* In tracking or detection applications, the performance of the task is normally dependent on the signal-to-noise ratio of the measurements. If we consider that the signal we measure attenuates with distance, then in a sensor network, as the density of sensors can be high, it is expected that at least a few sensors will measure the signal with high signal-to-noise ratio, enhancing in this way the final performance.

**Sensor network applications.** Based on the advantages of sensor networks presented above, a plethora of applications can be developed in many different domains [Arampatzis 2005], [Chong 2003], [Durisic 2012], [Puccinelli 2005]:

- *environmental monitoring.* Habitat monitoring, bio-complexity mapping, weather forecasting and disaster prevention (volcanic eruptions, floods, earthquakes).
- *Agricultural monitoring.* Precision irrigation, fertilization and pest control.
- *Civil engineering.* Building automation, building emergency systems and structural health monitoring.
- *Urban monitoring.* Pollution monitoring, video surveillance and traffic control.
- *Health applications.* Monitoring of human physiological data, tracking of doctors and patients in a hospital.
- *Commercial applications.* Support for logistics, production surveillance and automation.

- *Military applications.* Self-healing landmines, soldier detection and tracking, shot origin information, perimeter protection, chemical, biological and explosive vapor detection, missile canister monitoring and blast localization.

**The need for quantization.** Even if progress in sensor and communication technologies motivates the use of a large number of communicating sensors, practical considerations such as the use of non replenishable energy sources (sensors are self-powered with batteries) and maximum size constraints impose three design constraints:

- *energy constraint:* which comes directly from the choice that the sensors use a non replenishable energy source.
- *Rate constraint:* this constraint is related to the fact that the communication channel bandwidth must be shared by a large quantity of sensors and that the energy is also constrained.

The energy spent in a sensor network can be divided mainly in three activities, sensing, communication and processing. It is known that the major energy consumer of these activities is communication [Akyildiz 2002]. As bandwidth is constrained, the simplest solution to have reduced energy consumption is to find a way to achieve the same or similar goal by communicating with a lower rate (number of bits per unit of time).

- *Complexity constraint:* although much less important in energy consumption, complexity both in terms of processing and memory must be small to keep the cost and size of the sensors small.

One way to treat these problems is to consider that the sensors quantize their measurements before the realization of any other operations<sup>2</sup>. This allows to

- reduce complexity by using pre-stored tables for the computations and also by bounding memory requirements.
- Reduce directly the rate by controlling the number of quantization intervals.
- Reduce energy requirements, as a consequence of the reduction in complexity and rate.

These are the main reasons for studying quantization in this thesis.

## Different objectives and the scope of the thesis

In a sensing system the main task is to infer some information that is embedded in the measurements. The two main classes of inference problems studied in signal processing are detection and estimation. The literature on the joint subjects, detection based on quantized

---

<sup>2</sup>We do not claim here that imposing quantization of the measurements is the optimal solution. In some cases, it can be shown that a complete analog scheme is optimal [Gastpar 2008].

measurements and estimation based on quantized measurements, is not expressive if compared with the literature on the separated subjects, however, as a consequence of the emergence of sensor networks its size is increasing. Some references on these subjects are the following:

- *Detection*: [Benitz 1989], [Gupta 2003], [Kassam 1977], [Longo 1990], [Picinbono 1988], [Poor 1977], [Poor 1988], [Tsitsiklis 1993], [Villard 2010], [Villard 2011].
- *Estimation*: [Aysal 2008], [Fang 2008], [Gubner 1993], [Luo 2005], [Marano 2007], [Papadopoulos 2001], [Poor 1988], [Ribeiro 2006a], [Ribeiro 2006b], [Ribeiro 2006c], [Wang 2010].

**Estimation based on quantized measurements.** As mentioned before, in this thesis we will study the second of the subjects mentioned above, namely, estimation based on quantized measurements. We will start by explaining the general estimation problem in a sensing system. By making a sequence of simplifications in the general problem, we will get to the main scope of this thesis.

In the general scheme, each sensor measures a continuous amplitude quantity  $X^{(i)}$ , processes locally its measurement and sends it to the point where the estimate will be evaluated. The point of evaluation can be either a fusion center, one of the sensors or all sensors. In the last case, all sensors broadcast their processed measurements. This scheme is shown in Fig. 1. The quantity in this case can be a sequence of vectors, a sequence of scalars, a constant vector or a constant scalar.

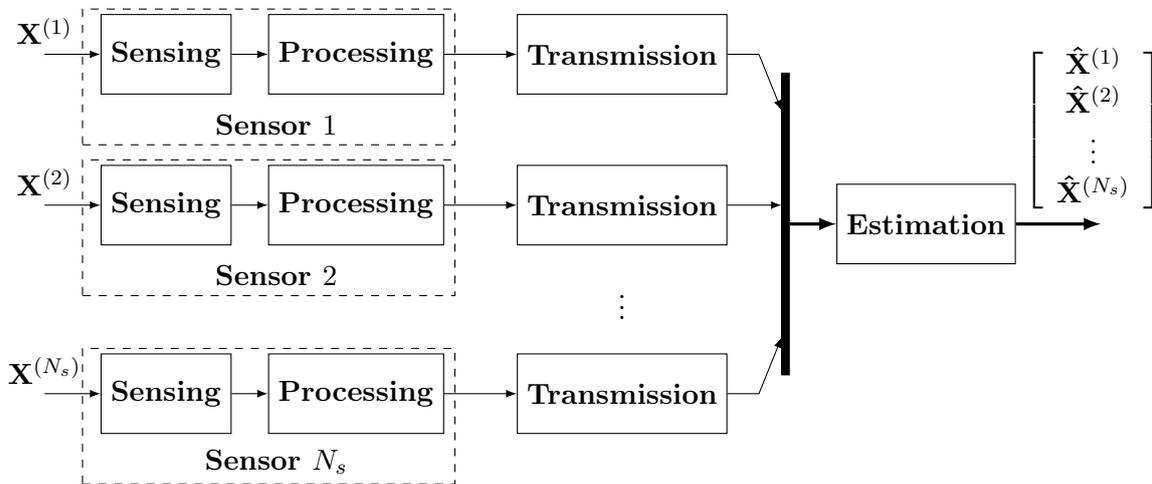


Figure 1: Estimation problem using a sensing system. Multiple sensors send preprocessed information to the final estimator that must recover the quantities of interest.

The first simplification that we will make is to consider only one of the terminals (sensors) in the sensing system, eventually, we might consider the problem with multiple terminals but with the same quantity being measured by all sensors. We will also consider that the quantity to be estimated is either a sequence of scalars or one scalar. We will use the notation  $X_k$  for the quantity to be estimated in both cases,  $k$  is the sample index and, in most cases, it will be

also the discrete-time index. When  $X_k$  is a scalar constant, we have  $X_k = x$ . The simplified problem, which can also be called scalar remote sensing problem, is depicted in Fig. 2.

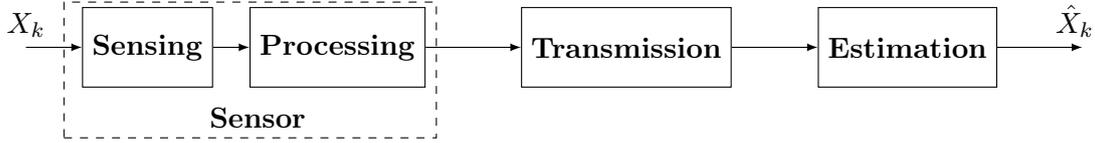


Figure 2: Scalar remote sensing problem. A scalar single terminal simplification of the problem depicted in Fig. 1.

The parameter  $X_k$  is measured with continuous amplitude additive noise  $V_k$ . The continuous measurement will be denoted  $Y_k = X_k + V_k$ .

The estimation problem we mainly deal here is location estimation, as  $X_k$  in this case is a location parameter characterizing the measurement distribution. Other technical considerations about the noise sequence will be presented later. In some points of the thesis we will not constrain  $X_k$  to be a location parameter and we will let it be a general parameter.

According to the previous discussion on the design constraints, the processing block is replaced by a scalar quantizer. Thus, each noisy continuous measurement  $Y_k$  will generate a quantized measurement  $i_k$  according to the quantizer function  $Q(\cdot)$ . Each quantized measurement is defined in a finite set of values so the rate (number of possible values per measurement of the alphabet) is fixed and known. We suppose that the rate in bits per unit of time is chosen such that the transmission channel capacity is not exceeded, thus, by adding proper channel coding in the transmission block, we can consider that the channel is perfect.

For each time  $k$  we are interested in estimating  $X_k$  based on the set of past measurements  $i_1, i_2, \dots, i_k$ . The problem is then depicted in Fig. 3.

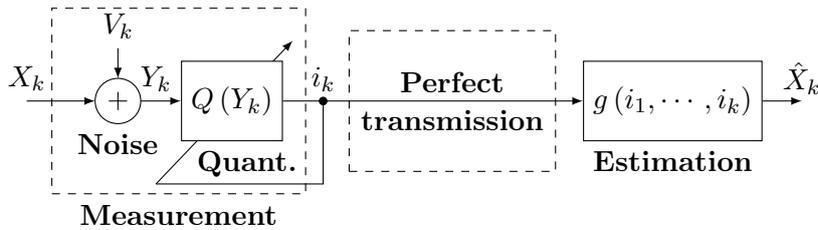


Figure 3: Estimation based on quantized measurements. A parameter is measured with additive noise, the measurements are then quantized and transmitted through a perfect channel. Based on the past quantized measurements, the objective is to estimate  $X_k$  for each time  $k$  with the sequence of mappings  $g(\cdot)$ .

As it is shown in Fig. 3, we also consider that the quantizer structure can depend on the past quantized measurements.

**What we want to study.** We want to propose algorithms for estimating  $X_k$  based on  $i_k$ . The parameter  $X_k$ , which will be detailed later, can be either a deterministic constant or a slowly varying random process.

After proposing the algorithms, we want to evaluate their performance. Given the algorithm performance, we want to study the effects of the quantizer function parameters, the quantization thresholds, and of the quantizer resolution, the number of quantization intervals or bits.

For assessing how quantization impacts on estimation, we will also compare the estimation performance of the proposed algorithms with the estimation performance of their corresponding continuous measurement versions.

The objective here is to estimate  $X_k$  based on the interval information (we know only in which interval the measurement is) of a noisy version of it.

**What we do not want to study (and we will not study).** We do not want to reconstruct the measurement  $Y_k$  from the quantized measurements and then estimate  $X_k$  based on the reconstructed measurements, as if they were continuous. By doing this, we would simply join their optimal separated solutions, which are well known.

We do not want to consider quantization as additive noise either. We want to consider the problem in its true form, that is to study how to exploit the information contained in intervals and not in continuous values.

**What we want to study but we will not study.** To specify in a precise way the scope of the thesis, we also have to state the problems we may have consciously overlooked. Consciously overlooked in this case means that, differently from the class of problems above, we wanted to study them, but to keep the subject simple, they will be neglected. These subjects are: vector parameters and vector quantization, presence of noisy channels (fading or additive) and channel coding, fast varying signals, estimation of continuous time signals and Bayesian estimation of a random constant.

## Structure of the thesis and outline

This thesis is formed by a general introduction, two parts and a general conclusion. Each part is divided in introduction, chapters and conclusion. In the first part, there are three chapters and in the second a single chapter. Each chapter is subdivided in three parts: introduction with the main contributions of the chapter, the main development and a summary/conclusion with some directions for future work. The conclusions in the order thesis–part–chapter increase in level of details. The thesis conclusion is a general overview, the part conclusion presents the points that we think we must retain without explaining the technical details and the chapter summary is a detailed account of the points observed in the chapter.

The thesis outline is the following:

- Part I: a study of algorithms/performance for estimation based on quantized measurements.
  - Chapter 1: the main details on the quantizer structure and noise are given. The fundamental algorithms and performance for the estimation of a deterministic scalar constant parameter are presented. Algorithms both for static quantization and adaptive quantization are studied.
  - Chapter 2: the time-varying parameter counterpart of Ch. 1 is presented. We consider the parameter to be a slowly varying scalar Wiener process and we present Bayesian algorithms for tracking the parameter.
  - Chapter 3: Low complexity algorithms are proposed as alternatives to those presented in Ch. 1 and 2. We also study some extensions of the scalar location problem: an extension that considers that the noise scale parameter is unknown and an extension that considers multiple sensors.
- Part II: a high resolution (high-rate) approximate analytical expression for the estimation performance.
  - Chapter 4: an open problem from Part I is how to set completely the quantizer key parameters so that estimation performance is maximized. In this chapter, we study how to solve this problem approximately by considering high resolution approximations (small quantization intervals approximation). We give a practical solution to obtain the optimal quantizer and the corresponding asymptotic estimation performance.

Each part will begin with an example, which can be seen as a background for the presentation of the problem. The examples serve only for presentation purposes and their specific subjects (water management and deep-sea water mining) are not the main subject of this work.

The appendices of this thesis are divided in three parts, one part for presenting proofs that are considered not important to develop in the main text **Why? - App. A.1**, another for giving more details about a subject **More? - App. A.2** and one part for explaining some implementation issues **How? - App. A.3**.

For defining a new abbreviation or acronym we write the expression in **boldface with the abbreviation in parenthesis (.)**. For citing a reference that was already cited similarly elsewhere, we write the reference and the work where it was cited with **(cited in ...)**.

## Publication

During this thesis three papers were presented in international conferences

- Rodrigo C. Farias and Jean-Marc Brossier, *Adaptive Estimation Based on Quantized Measurements*, IEEE International Conference on Communications (ICC), 2013, Budapest, Hungary.
- Rodrigo C. Farias and Jean-Marc Brossier, *Adjustable Quantizers for Joint Estimation of Location and Scale Parameters*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, Vancouver, Canada.
- Rodrigo C. Farias and Jean-Marc Brossier, *Asymptotic Approximation of Optimal Quantizers for Estimation*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, Vancouver, Canada.

One paper was accepted for presentation in a French conference

- Rodrigo C. Farias and Jean-Marc Brossier, "*Quantification asymétrique optimale pour l'estimation d'un paramètre de centrage dans un bruit de loi symétrique*", "Colloque GRETSI", 2013, Brest, France

and one article was published

- Rodrigo C. Farias and Jean-Marc Brossier, *Adaptive quantizers for estimation*, Signal Processing, Elsevier, vol. 93, november 2013.

Part I

Estimation based on quantized  
measurements:

algorithms

and

performance



*"A word to the wise is enough" - popular saying.*

## Motivation

As a background to introduce this part, we start with an application example. A recent trend in placing water as a key element in government strategic decisions (including possible future military interventions) lead to the choice of the motivational example.

Agriculture is responsible for 70% of freshwater withdrawals. Food production for satisfying the daily caloric needs of a person consumes 3000 liters of water, a very large quantity when compared with the 2-5 liters used for drinking. Add to these ingredients the fact that world population is growing and that a large part of the population is changing its diet, consuming more meat and vegetables and therefore even more water [Molden 2007] and we have a possible recipe for future water scarcity.

One possible policy for preventing future water scarcity is to develop or improve irrigation systems in sub-developed countries, where water use efficiency is very low [Molden 2007]. For doing so, measuring accurately the soil moisture of crop fields is a main issue. Thus, as a background scenario for introducing this chapter, we will consider the problem of estimating the moisture level of crop fields.

Consider that multiple crop field areas will have each a set of sensors noisy measurements of some quantities related to soil moisture. All the data will be transmitted to a central processor, which after estimating the moisture levels, will decide which crops must be irrigated. As the number of sensed areas can be large, for example, when the irrigation system is integrated for an entire geographic region, quantization will be applied to respect communication constraints.

The solution to this problem can be simplified by assuming that the decision (control) part of the problem can be decoupled from the estimation part. We will focus here only on the estimation part. In a first approach, we can assume that the moisture levels are unknown deterministic scalars, unrelated from one region to the other and that they are approximately constant for a block of  $N$  independent measurements. If humidity sensors are used, from the symmetry of the problem and the assumption that the moisture levels are not related, the joint estimation problem of all levels decouples into many scalar estimation problems with identical general form. This general form is the following:

**(a) Estimate a constant scalar location parameter  $x$ , based on  $N$  independent noisy measurements**

$$Y_{1:N} = \{Y_1 = x + V_1, \dots, Y_N = x + V_N\},$$

**which are scalarly quantized with a quantizer function  $Q$  (to be defined later)**

$$i_{1:N} = \{i_1 = Q(Y_1), \dots, i_N = Q(Y_N)\}.$$

A more detailed meaning for "estimate" is

- (1) Give an analytical form or a procedure describing the parameter estimator  $\hat{X}$ .
- (2) Give the estimation performance or an approximation of the estimation performance as a function of
  - number of measurements;
  - noise characteristics;
  - the quantizer function.

After giving a solution for this problem, we may be interested in considering a more complex model for  $x$ , for instance, instead of considering it as a constant, we can assume that it varies randomly with time.

A simple dynamical model is

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

where  $k$  is the discrete-time index,  $W_k$  is an **independent and identically distributed (i.i.d.)** Gaussian process, with zero mean and variance  $\sigma_w^2$ . Thus, if  $X_0$  is Gaussian,  $X_k$  is a Gaussian process known as a **discrete-time Wiener process** or as a discrete-time random walk process. This type of process is commonly used to describe slowly varying parameters when their evolution is random but with unknown form. A reason to use this model is that by constraining the increments to be Gaussian distributed, minimal quantity of information is imposed for a given increment variance (in terms of information theory quantity of information).

Now, suppose we have statistics about precipitation on the crop field region, for example its average, we also know the last quantities of water irrigated on the field and how to relate both precipitation and irrigated water to average increase in moisture level, denoted  $u_k$ . This will allow us to use a more precise dynamical model for  $X_k$ , using as increments Gaussian **random variables (r.v.)** with mean  $u_k$ . Consequently, our model will become a **discrete-time Wiener process with a deterministic drift**.

The objective is the same as before, estimate  $X_k$  based on scalarly quantized  $Y_k$ . However, the relation between measurements can be exploited now. Instead of considering the static estimation problem for separate blocks of measurements, we can now use all past measurements in the estimation of the varying parameter, under the constraint that the parameter evolution must follow the dynamical model.

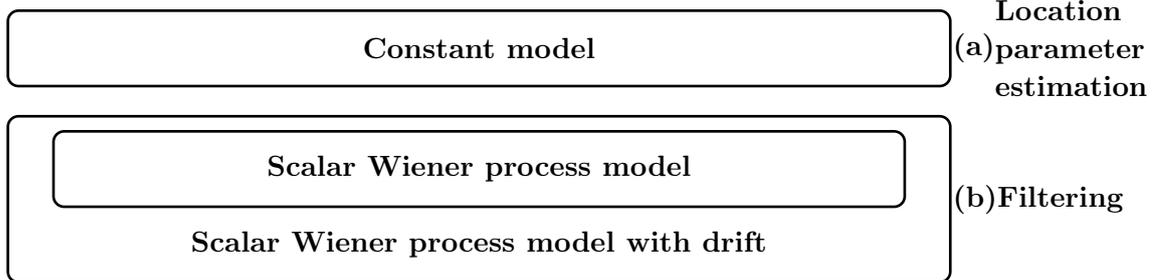
Therefore, we are also interested in solving the following problem:

**(b) Estimate a varying random parameter  $X_k$  at time  $k$  based on the last and present scalarly quantized measurements  $i_{1:k}$ .**

This is a filtering problem, as estimates depend not only on past measurements but also on the present measurement [Jazwinski 1970]. The problems of estimation based on past measurements, *i.e.* prediction, and of estimation based on additional future measurements, *i.e.* smoothing, will not be treated in this thesis.

## Outline for this part

For the problem at hand, 3 types of model with increasing complexity can be considered, these 3 models are related to the two estimation problems (a) and (b) as it is shown below:



Many other practical estimation problems rely on the models presented above and consequently can be cast as (a) or (b). We will look now for their solutions.

First, we will present algorithms and performance for the estimation of a constant location parameter. We will study maximum likelihood estimators and their asymptotic performance through the Cramér–Rao bound. We will see that estimation performance is sensible to the distance between the quantizer dynamic range and the parameter. For commonly used noise models, we will see that estimation performance actually degrades when the dynamic range is far from the parameter. As a solution, we will search for adaptive schemes that place the dynamic range close to the parameter. We will show that in the binary case, the asymptotically optimal adaptive algorithm is given in a simple recursive form.

After that, we will focus on filtering. A general solution using recursive integral expressions will be given. As this solution is analytically intractable, an approximate solution based on sequential Monte Carlo methods (particle filtering) will be considered. Its performance will be assessed through a lower bound, the Bayesian Cramér–Rao bound. Then, by analyzing the bound, we will see that a good estimation scheme can be obtained by quantizing the measurement prediction error, usually called the innovation. We will show that the asymptotically optimal filter based on the quantized innovation is also given in a simple recursive form when

the parameter varies slowly.

Motivated by the recursive forms that are obtained asymptotically, both in the constant and varying parameter cases, we will present a low complexity adaptive algorithm for estimation using quantized measurements. The estimation performance and the optimal algorithm parameters will be obtained for constant and Wiener process models. Extensions of the algorithm for the cases when multiple sensors with a fusion center are used and when the noise scale factor (a measure of its amplitude) is unknown will also be obtained.

At the end of this part some conclusions will be drawn on the overall aspects of estimation based on quantized measurements.

# Estimation of a constant parameter: what is done and a little more

---

In this chapter we study the problem of estimation of a constant location parameter based on quantized measurements. We start the chapter with the measurement model, which is mainly the noise model and the definition of the quantizer. The first sections of the chapter deal with a fixed quantizer structure (fixed quantization thresholds), while in the last sections, we present estimation schemes with an adaptive quantizer structure.

In the part concerning a fixed quantizer structure, we start by giving a general estimation algorithm based on the maximum likelihood method. Its performance is given in terms of the Cramér–Rao bound. Then, we study the general effects of quantization on estimation performance. This is done through the analysis of the Cramér–Rao bound, a quantity that is directly related to the Fisher information. We also analyze the performance of binary and multibit quantization as a function of the quantizer tuning parameter. We give a detailed implementation of the maximum likelihood estimator for general noise distributions in the binary case, while in the multibit case, the maximum likelihood estimator is detailed for a more restricted class of noise distributions, more precisely, log-concave distributions.

As a main result of the performance analysis for the fixed threshold scheme, we will see that, for commonly used noise models, the estimation performance degrades as the quantization dynamic range is distant from the true parameter. This is used as a motivation to study estimation schemes that adaptively place the quantizer dynamic range close to the true parameter. We study two adaptive schemes. One based on a simple update of the quantizer main parameter, but with the final estimate given by maximum likelihood estimation and the other based on the use of the maximum likelihood last estimate as the quantizer main parameter. Their performances are given also in terms of the Cramér–Rao bound. We will also see that the estimator based on the maximum likelihood threshold update is asymptotically equivalent to a low complexity recursive algorithm.

We finish this chapter with a summary of the main points that were studied and with the directions for further research. The directions will point for further work that is presented in other chapters or that will be studied in the future.

**Contributions presented in this chapter:**

- *Global and local analysis for binary quantization.* By reading carefully the literature on the subject, we have the impression that setting the quantization threshold on the true parameter value is optimal for symmetric distributions [Wang 2010, p. 265]. But this affirmation is actually false. We present here global and local conditions on the noise distribution that guarantees that this threshold value (equal to the parameter value) is indeed optimal.
- *Asymmetric threshold case.* Differently from the literature where only the symmetric cases are shown, we show some cases where the noise distribution is symmetric and the optimal quantization threshold is not the median.
- *Laplacian noise.* In the literature, most of the analysis is focused on the Gaussian noise case, where, as it is expected, quantization strictly decreases estimation performance. Here, we study also the Laplacian case. The Laplacian case is easier to analyze and it is a nice counterexample to the intuition that quantization strictly decreases estimation performance (see p. 55).
- *Adaptive binary quantization scheme in a finite grid.* We present a method to obtain the asymptotic threshold probabilities in the adaptive binary threshold scheme (see (More? - App. A.2.4)). Differently from the method presented in [Fang 2008], where a truncation approximation is used, in the method presented here, we define boundaries on the possible threshold values so that the number of threshold values is finite and the asymptotic probabilities can be evaluated analytically.
- *Multibit adaptive scheme based on the maximum likelihood estimator and its convergence.* We extend the binary adaptive scheme presented in [Fang 2008] to the multibit case and we also extend its proof of convergence to the general multibit non Gaussian case.
- *Asymptotic binary adaptive scheme based on the MLE.* We give a less heuristic proof that the adaptive quantization scheme based on the maximum likelihood estimator is given asymptotically in a simple recursive form.

---

**Contents**


---

<b>1.1</b>	<b>Measurement model . . . . .</b>	<b>34</b>
1.1.1	Noise model . . . . .	34
1.1.2	Quantization model . . . . .	35
<b>1.2</b>	<b>Maximum likelihood, Cramér–Rao bound and Fisher information . .</b>	<b>37</b>
1.2.1	Maximum likelihood estimator . . . . .	38
1.2.2	Cramér–Rao bound and the Fisher information . . . . .	39
1.2.3	Quantization loss . . . . .	41
<b>1.3</b>	<b>Binary quantization . . . . .</b>	<b>44</b>
1.3.1	The Gaussian case . . . . .	44
1.3.2	The Laplacian case . . . . .	45
1.3.3	The general case . . . . .	46
1.3.4	Asymmetric threshold: surprising cases . . . . .	48
1.3.5	Conclusions on binary quantization performance . . . . .	52
1.3.6	MLE for binary quantization . . . . .	53
<b>1.4</b>	<b>Multibit quantization . . . . .</b>	<b>54</b>
1.4.1	The Laplacian case . . . . .	55
1.4.2	The Gaussian and Cauchy cases under uniform quantization . . . . .	56
1.4.3	Summary of the main points . . . . .	58
1.4.4	MLE for multibit quantization with fixed thresholds . . . . .	58
<b>1.5</b>	<b>Adaptive quantizers: the high complexity fusion center approach . .</b>	<b>60</b>
1.5.1	MLE for the adaptive binary scheme . . . . .	61
1.5.2	Performance for the adaptive binary scheme . . . . .	62
1.5.3	Adaptive scheme based on the MLE . . . . .	66
1.5.4	Performance for the adaptive multibit scheme based on the MLE . . . . .	66
1.5.5	Equivalent low complexity asymptotic scheme . . . . .	70
<b>1.6</b>	<b>Chapter summary and directions . . . . .</b>	<b>73</b>

---

## 1.1 Measurement model

We start by explaining the measurement model. The unknown deterministic scalar constant parameter to be estimated is

$$x \in \mathbb{R}$$

and it is measured  $N$  times,  $N \in \mathbb{N}^*$ , with i.i.d. additive noise  $V_k$ . For  $k \in \{1, \dots, N\}$  the continuous measurements are

$$Y_k = x + V_k. \quad (1.1)$$

### 1.1.1 Noise model

The continuous sequences of r.v.  $Y_k$  and  $V_k$  are defined on the probability space  $\mathcal{P} = (\Omega, \mathcal{F}, \mathbb{P})$  with values on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . For simplification purposes the following hypotheses on the noise distribution will be considered:

**Assumptions (on the noise distribution):**

**AN1** The marginal **cumulative distribution function (CDF)** of the noise, denoted  $F$ , admits a **probability density function (PDF)**  $f$  **with respect to (w.r.t.)** the standard Lebesgue measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

**AN2** The PDF  $f(v)$  is a strictly positive even function and it strictly decreases w.r.t.  $|v|$ .

Assumption AN1 is a commonly used assumption that in practice will be used when the derivative of  $F$  w.r.t. its arguments is needed. AN2 means that the noise distributions are unimodal and symmetric around zero and it will be used for the following reasons:

1. The unimodal behavior of the noise will allow to have a general qualitative characterization of estimation performance as a function of quantization parameters. More precisely, it will be observed that for unimodal densities very poor estimation performance occurs for quantizers having the dynamical range far away from  $x$ .
2. It will be used as a condition for the convergence of some new adaptive estimation algorithms presented in this thesis.
3. In the lack of physical constraints (*e.g.* positivity), there should be no reason for the components of the noise to be asymmetric. Thus, if we consider that the noise is a normalized sum of an infinite number of symmetric i.i.d. r.v. (an infinite sum of small perturbations), then it is known that the resulting noise r.v. distribution is a symmetric stable distribution [Samorodnitsky 1994], which is unimodal.

Even if not all unimodal symmetric distributions are stable the generalized central limit theorem above serves as an additional motivation.

## 1.1.2 Quantization model

From the reasons presented in the Introduction and in the motivational example given above, the measurements are quantized. We will consider that they are scalarly quantized, which means that each measurement is quantized separately from the others. The quantizer output can be written as

$$i_k = Q(Y_k), \quad (1.2)$$

where  $i_k$  is a value from a finite set  $\mathcal{I}$  of  $\mathbb{R}$  with  $N_I$  elements. Due to notation issues, we denote both the quantized measurement random variable and its realization with lowercase  $i$ .  $N_I$  is the number of quantization intervals. A simple example of quantizer  $Q$  with uniform threshold spacing is given in Fig. 1.1.

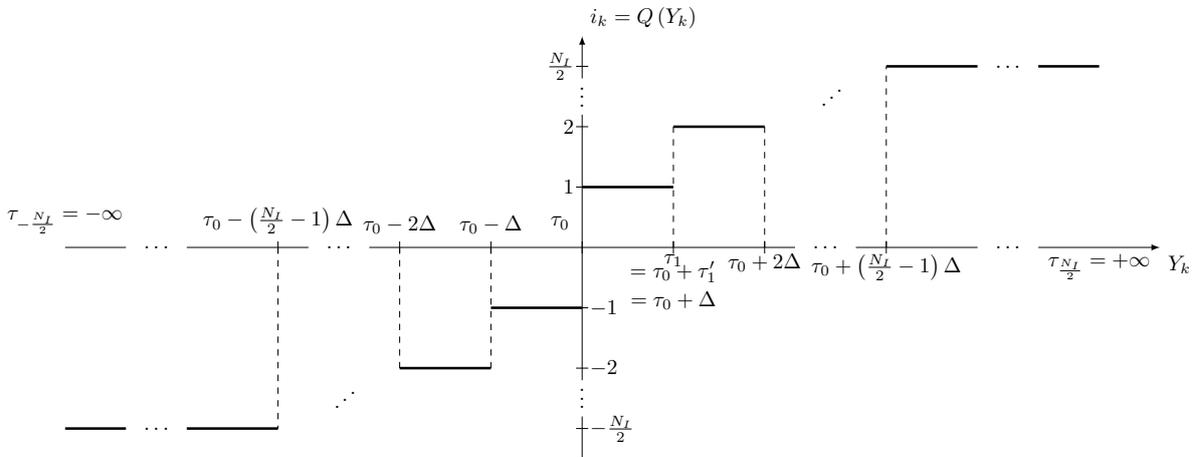


Figure 1.1: Quantizer function  $Q(Y_k)$  with  $N_I$  quantization intervals and uniform threshold spacing with length  $\Delta$ . The number of quantization intervals  $N_I$  is even, the quantizer is symmetric around the central threshold  $\tau_0$  and the output indexes are integers without the zero.

Except for the uniform thresholds, Fig. 1.1 shows the main elements of the general quantization model that will be used:

- the number of quantization intervals  $N_I$  will be an even number, this will lead to a clearer presentation, as in each analysis we will not need to deal with the additional central interval.
- The outputs of the quantizer will be defined on a set of integers from  $-\frac{N_I}{2}$  to  $\frac{N_I}{2}$ , without zero. This will simplify the notation of the algorithms that will be presented later. Note that as we will consider that the output of the quantizer is obtained without additional noise (it passes through a noiseless channel), the assignment of the output values  $i_k$  is not important, as long as the assigned values are different. For estimation purposes, only a label is needed at quantization output. The estimator or parts of the estimation procedure will carry out the role of the output quantization levels, as they are used in standard quantization, by generating estimates (values) based on the information from the intervals (indicated by the labels) where the continuous measurements lie.

Observe that if we introduce in the model a noisy communication channel and constraints on transmission power, the assignment of the output values becomes important. As it was stated in the Introduction, we are not going to consider this model in this thesis, but we can keep this extended problem as a possibility for future work.

- The quantizer is defined by  $N_I + 1$  thresholds  $\tau_i$ , which can be separated in three types: one central threshold  $\tau_0$ ,  $\frac{N_I}{2} - 1$  thresholds that are larger than  $\tau_0$  with an additional threshold at  $+\infty$  and  $\frac{N_I}{2} - 1$  that are smaller with an additional threshold at  $-\infty$ . We will consider that the non central thresholds are symmetric w.r.t.  $\tau_0$ , thus, for example, the threshold  $\tau_i$  is given by  $\tau_0$  plus a variation  $\tau'_i$  and the threshold  $\tau_{-i}$  is given by  $\tau_0$  minus the same variation. In the figure, the variations are integer multiples of  $\Delta$ , which corresponds to uniform quantization. In general, we will not impose uniform quantization.

The assumption on the symmetry of the quantizer is difficult to justify at this point, but the main idea is that, as it will be shown further for commonly used noise models, the best central threshold for estimation purposes is exactly  $x$ , thus if we set  $\tau_0 = x$ , from the assumption of noise symmetry, it seems reasonable to assume that the quantizer (a good one) is symmetric. In Part II, it will be shown that for large  $N_I$  the optimal quantizer is indeed symmetric around  $\tau_0$  for symmetric noise distributions.

The infinite thresholds for the extreme positive and negative thresholds are used to have the same notation for the probabilities of the granular (region inside the quantizer input dynamic range) and overload regions (region outside the quantizer input dynamic range).

From Fig. 1.1 and the explanations above, the quantizer function can be described as follows: if we have a measurement  $Y_k \geq \tau_0$  that falls in the quantization interval  $q_i = [\tau_{i-1}, \tau_i)$ , then its output will be  $i$ . Otherwise, if  $Y_k \leq \tau_0$  and it falls in  $q_{-i} = [\tau_{-i}, \tau_{-i+1})$ , then the quantizer output will be  $-i$ .

As an example, consider that we have a uniform quantizer with 16 quantization levels,  $\tau_0 = 0$  and uniform quantization step-length  $\Delta = 1$ , then for the input

$$y_{1:10} = \{-20, -8.5, -3.4, -5.6, -0.1, 0.7, 3.2, 10.7, 7.1, -2.3\},$$

we obtain

$$\begin{array}{ll} y_1 = -20 & \rightarrow i_1 = -8, & y_6 = 0.7 & \rightarrow i_6 = 1, \\ y_2 = -8.5 & \rightarrow i_2 = -8, & y_7 = 3.2 & \rightarrow i_7 = 4, \\ y_3 = -3.4 & \rightarrow i_3 = -4, & y_8 = 10.7 & \rightarrow i_8 = 8, \\ y_4 = -5.6 & \rightarrow i_4 = -6, & y_9 = 7.1 & \rightarrow i_9 = 8, \\ y_5 = -0.1 & \rightarrow i_5 = -1, & y_{10} = -2.3 & \rightarrow i_{10} = -3. \end{array}$$

Observe that by using the threshold variations  $\tau'_i$ , we can write the input–output relation in a more compact way:

$$i_k = i \operatorname{sign}(Y_k - \tau_0), \quad \text{for } |Y_k - \tau_0| \in [\tau'_{i-1}, \tau'_i]. \quad (1.3)$$

Note that the index  $k$  here is the time or sample index and it is not the particular value of  $i$ . Before proceeding, we will state explicitly the assumptions on the quantizer.

**Assumptions (on the quantizer):**

**AQ1**  $N_I$  is considered to be an even natural number and the set  $\mathcal{I}$  where  $i_k$  is defined is

$$\mathcal{I} = \left\{ -\frac{N_I}{2}, \dots, -1, 1, \dots, \frac{N_I}{2} \right\}.$$

**AQ2** The quantizer is symmetric around the central threshold. This means that the vector of thresholds  $\boldsymbol{\tau}$  is given by ( $\top$  is the transpose operator)

$$\boldsymbol{\tau} = \left[ \tau_{-\frac{N_I}{2}} = \tau_0 - \tau'_{\frac{N_I}{2}} \quad \dots \quad \tau_{-1} = \tau_0 - \tau'_1 \quad \tau_0 \quad \tau_1 = \tau_0 + \tau'_1 \quad \dots \quad \tau_{\frac{N_I}{2}} = \tau_0 + \tau'_{\frac{N_I}{2}} \right]^\top$$

with the threshold vector elements forming a strictly increasing sequence and the non-negative vector of threshold variations w.r.t. the central threshold given by

$$\boldsymbol{\tau}' = \left[ \tau'_0 = 0 \quad \tau'_1 \quad \dots \quad \tau'_{\frac{N_I}{2}} = +\infty \right]^\top.$$

## 1.2 Maximum likelihood, Cramér–Rao bound and Fisher information

We want to estimate  $x$  based on  $i_{1:N} = \{i_1, \dots, i_N\}$  (problem (a)). For doing so, we will look for an estimator

$$\hat{X}(i_{1:N}) - \text{which is a r.v. as it is a function of r.v.,}$$

that must be as close as possible to  $x$ . In our case, we are going to choose the quantitative meaning of "as close as possible" to be with minimum (or small) **mean squared error (MSE)**:

$$\text{MSE} = \mathbb{E} \left[ \left( \hat{X} - x \right)^2 \right], \quad (1.4)$$

$\mathbb{E}$  is the expectation w.r.t. the joint distribution of the noise. The MSE is a commonly used performance criterion for estimation problems. Although it is widely used, it has the inconvenient that it is impossible to find in a general form the  $\hat{X}$  minimizing it by direct analytical minimization [Van Trees 1968, p. 64].

### 1.2.1 Maximum likelihood estimator

A common solution for this problem is to suppose that  $N$  is large, in theory  $N$  must tend to infinity, and that  $\hat{X}$  is constrained to be unbiased, which means

$$\mathbb{E}[\hat{X}] = x,$$

in this case, the optimal  $\hat{X}$  minimizing the MSE is known to be the **maximum likelihood estimator (MLE)** [Kay 1993, p. 160]. The MLE consists of maximizing the likelihood function which is the joint distribution of the measurements considering that the measurements are fixed parameters and that the parameter  $x$  is variable<sup>1</sup>. For the estimation problem considered here, the likelihood for an independent block of measurements  $i_{1:N}$  is

$$L(x; i_{1:N}) = \prod_{k=1}^N \mathbb{P}(i_k; x), \quad (1.5)$$

where  $\mathbb{P}(i_k; x)$  is the probability of having a quantizer output  $i_k$  at time  $k$  for a parameter  $x$ . This probability can be rewritten using the noise CDF and the thresholds:

$$\mathbb{P}(i_k; x) = \begin{cases} \mathbb{P}(\tau_{i_k-1} \leq Y_k < \tau_{i_k}), & \text{if } i_k > 0, \\ \mathbb{P}(\tau_{i_k} \leq Y_k < \tau_{i_k+1}), & \text{if } i_k < 0, \end{cases}$$

using the definition of  $Y_k = x + V_k$  given by (1.1)

$$\begin{aligned} \mathbb{P}(i_k; x) &= \begin{cases} \mathbb{P}(\tau_{i_k-1} \leq x + V_k < \tau_{i_k}), & \text{if } i_k > 0, \\ \mathbb{P}(\tau_{i_k} \leq x + V_k < \tau_{i_k+1}), & \text{if } i_k < 0, \end{cases} \\ &= \begin{cases} F(\tau_{i_k} - x) - F(\tau_{i_k-1} - x), & \text{if } i_k > 0, \\ F(\tau_{i_k+1} - x) - F(\tau_{i_k} - x), & \text{if } i_k < 0. \end{cases} \end{aligned} \quad (1.6)$$

The MLE is the value of  $x$  maximizing  $L(x; i_{1:N})$  for a given  $i_{1:N}$ :

$$\hat{X}_{ML,q} = \hat{X}_{ML}(i_{1:N}) = \underset{x}{\operatorname{argmax}} L(x; i_{1:N}). \quad (1.7)$$

The subscript  $q$  is used to make explicit that the estimation is done with quantized measurements. As the logarithm is a strictly increasing function on  $\mathbb{R}_+^*$  and most used likelihood functions are given in exponential form, it is common to solve an equivalent maximization problem:

$$\hat{X}_{ML,q} = \underset{x}{\operatorname{argmax}} \log L(x; i_{1:N}).$$

---

<sup>1</sup>Clearly, this is an inversion of roles from the modeling point of view and this is the main reason why we do not call the likelihood function simply by joint PDF.

### 1.2.2 Cramér–Rao bound and the Fisher information

The MLE is the procedure to find the estimate. We still need its performance. Unfortunately, no finite sample (finite  $N$ ) performance results are available for the MLE. We will focus then on asymptotic results for which in some sense, as stated before, the MLE is optimal.

The MSE for the MLE can be written as

$$\mathbb{E} \left[ \left( \hat{X}_{ML,q} - x \right)^2 \right] = \left[ \mathbb{E} \left( \hat{X}_{ML,q} - x \right) \right]^2 + \text{Var} \left( \hat{X}_{ML,q} \right) = \text{bias}^2 + \text{variance}.$$

As it was stated, the MLE is asymptotically unbiased:

$$\mathbb{E} \left[ \hat{X}_{ML,q} \right] \underset{N \rightarrow \infty}{=} x. \quad (1.8)$$

Therefore, it is characterized asymptotically only by its variance.

The **Cramér–Rao bound (CRB)** is a lower bound on the variance of any unbiased estimator [Kay 1993, p. 30] and the bound is valid even for finite  $N$ . Under some regularity conditions, the asymptotic variance of the MLE is known to be minimum and it attains the CRB [Kay 1993, p. 160]:

$$\text{Var} \left( \hat{X}_{ML,q} \right) \underset{N \rightarrow \infty}{\sim} \text{CRB}_q, \quad (1.9)$$

later, we will compare this  $\text{CRB}_q$  with its corresponding version for continuous measurements that we will denote  $\text{CRB}_c$ . The symbol  $\underset{N \rightarrow \infty}{\sim}$  used here means that both quantities are equivalent

$$\lim_{N \rightarrow \infty} \frac{\text{Var} \left( \hat{X}_{ML,q} \right)}{\text{CRB}_q} = 1.$$

As the MLE is asymptotically unbiased and with asymptotically minimum variance it is usually called an asymptotically efficient estimator in classical estimation terms.

Note that the optimality in asymptotic variance does not imply optimality in MSE sense, as a biased estimator can attain a lower asymptotic MSE when compared with the MLE. Also, it is important to stress that the variance of the MLE will tend to the CRB only if the maximum of the likelihood can be achieved. This can be an issue when we need to evaluate the maximum of the likelihood through a numerical method, in this case we have to ensure that the numerical method will converge to the global maximum. In what follows, we will assume that the MLE, either evaluated analytically or numerically, is always the global maximum of the likelihood. For further discussion on the issues of finding the MLE see (More? - App. A.2.1).

The CRB is the inverse of the **Fisher information (FI)** [Kay 1993, p. 30]. The FI is given by the variance of the score function  $S_q$ . As the expected value of the score function is zero [Kay 1993, p. 67], the FI is given by the second order moment of the score function. Starting from the definition of the score function for  $N$  quantized measurements and going in

the direction of the asymptotic variance of the MLE, we have the following expressions:

$$\begin{aligned}
S_{q,1:N} &= \frac{\partial \log L(x; i_{1:N})}{\partial x} && \text{- score function,} \\
I_{q,1:N} &= \mathbb{E} [S_{q,1:N}^2] = \mathbb{E} \left\{ \left[ \frac{\partial \log L(x; i_{1:N})}{\partial x} \right]^2 \right\} && \text{- FI,} \\
\text{Var} \left( \hat{X}_{ML,q} \right) &\underset{N \rightarrow \infty}{\sim} \text{CRB}_q = \frac{1}{I_{q,1:N}} = \frac{1}{\mathbb{E} \left\{ \left[ \frac{\partial \log L(x; i_{1:N})}{\partial x} \right]^2 \right\}} && \text{- variance and CRB.}
\end{aligned}$$

the subscript is used to indicate that these quantities are related to the quantized measurements  $i_{1:N}$ . Due to the fact that the measurements are i.i.d., whenever we want to refer to the score function and FI for one measurement  $i_k$ , we can drop the sample indexes, thus writing  $S_q$  and  $I_q$ . Under the assumption of independent measurements (independent noise), we have the following:

- the joint probability in the FI expression decomposes in a product of marginal probabilities.
- The logarithm of the product of marginal probabilities becomes the sum of the logarithm of each probability.
- After differentiating the sum of logarithms w.r.t.  $x$ , the square of the differentiated sum can be decomposed in a sum of squared terms and a sum of products between different terms.
- The expectation of the products between different terms is zero because the factors in the products are independent and with zero mean (they are score functions thus having zero mean [Kay 1993, p. 67]).
- The expectation of each squared term is the FI for the corresponding individual measurement.

Therefore, as the measurements are also identically distributed, the FI for  $N$  quantized measurements is  $N$  times the FI for one measurement  $I_q$ :

$$\text{Var} \left( \hat{X}_{ML,q} \right) \underset{N \rightarrow \infty}{\sim} \text{CRB}_q = \frac{1}{NI_q}. \quad (1.10)$$

The score function for one measurement  $S_q$  is

$$S_q = \frac{\partial \log L(x; i_k)}{\partial x} = \frac{\frac{\partial \mathbb{P}(i_k; x)}{\partial x}}{\mathbb{P}(i_k; x)} \quad (1.11)$$

and the corresponding FI is

$$\begin{aligned}
I_q &= \mathbb{E} \left\{ \left[ \frac{\partial \log L(x; i_k)}{\partial x} \right]^2 \right\} = \sum_{i_k \in \mathcal{I}} \left[ \frac{\frac{\partial \mathbb{P}(i_k; x)}{\partial x}}{\mathbb{P}(i_k; x)} \right]^2 \mathbb{P}(i_k; x), \\
&= \sum_{i_k \in \mathcal{I}} \frac{\left[ \frac{\partial \mathbb{P}(i_k; x)}{\partial x} \right]^2}{\mathbb{P}(i_k; x)}. \quad (1.12)
\end{aligned}$$

Defining the difference between the central threshold and the parameter as  $\varepsilon = \tau_0 - x$ , using the CDF, PDF notations and the symmetry of the quantization thresholds we have

$$I_q = \sum_{i_k=1}^{i_k=\frac{N_I}{2}} \left\{ \frac{[f(\varepsilon + \tau'_{i_k-1}) - f(\varepsilon + \tau'_{i_k})]^2}{F(\varepsilon + \tau'_{i_k}) - F(\varepsilon + \tau'_{i_k-1})} + \frac{[f(\varepsilon - \tau'_{i_k}) - f(\varepsilon - \tau'_{i_k-1})]^2}{F(\varepsilon - \tau'_{i_k-1}) - F(\varepsilon - \tau'_{i_k})} \right\}. \quad (1.13)$$

The solution to problem (a) (p. 27) given by the MLE is the following:

**Solution to (a) - MLE for a fixed thresholds set  $\tau$**

**(a1) 1) Estimator**

$$\hat{X}_{ML,q} = \underset{x}{\operatorname{argmax}} L(x; i_{1:N})$$

**or**

$$\hat{X}_{ML,q}(i_{1:N}) = \underset{x}{\operatorname{argmax}} \log L(x; i_{1:N}),$$

**with  $L(x; i_{1:N})$  given by (1.5)**

$$L(x; i_{1:N}) = \prod_{k=1}^N \mathbb{P}(i_k; x).$$

---

**2) Performance (asymptotic)**

**$\hat{X}_{ML,q}$  is asymptotically unbiased**

$$\mathbb{E} \left[ \hat{X}_{ML,q} \right]_{N \rightarrow \infty} = x$$

**and its asymptotic MSE or variance is given by**

$$\operatorname{Var} \left( \hat{X}_{ML,q} \right)_{N \rightarrow \infty} \sim \mathbf{CRB}_q = \frac{1}{NI_q},$$

**with  $I_q$  given by (1.13).**

The CRB given above is not only related to the MLE, but can be used to approximately assess the performance of any good (close to optimal) estimator. In our case, it can be used to characterize the performance of the measurement/estimation system (Fig. 1.2) independently of the estimator.

### 1.2.3 Quantization loss

The solution given above does not contain any direct characterization of the estimation performance as a function of  $N_I$  and/or  $\tau$ . We are going to look into these details now.

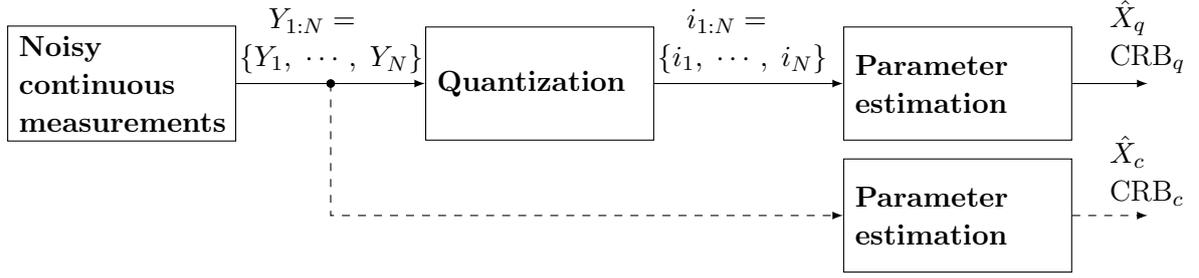


Figure 1.2: Scheme representing the general measurement/estimation system. The continuous measurements sequence  $Y_{1:N}$  is scalarly quantized and the quantized sequence  $i_{1:N}$  is used for estimation.  $\hat{X}_q$  and  $\hat{X}_c$  are the estimators based on quantized or continuous measurements and  $\text{CRB}_q$  and  $\text{CRB}_c$  are their respective CRB.

### Loss with respect to the continuous measurement

We will start analyzing the general effect of quantization on estimation. An approximate way of doing this (exact for  $N \rightarrow \infty$ ) is to study the quantized FI for one measurement  $I_q$  and its difference with respect to the continuous measurement FI  $I_c$ .  $I_q$  was given in (1.13), while  $I_c$  is given by

$$I_c = \mathbb{E} [S_c^2], \quad (1.14)$$

where  $S_c$  is the score function for continuous measurements given by

$$S_c(y) = \frac{\partial \log f(y-x)}{\partial x}. \quad (1.15)$$

The difference between  $I_c$  and  $I_q$  can be obtained by evaluating the quantity  $\mathbb{E} [(S_c - S_q)^2]$  [Marano 2007]. Indeed,

$$\mathbb{E} [(S_c - S_q)^2] = \mathbb{E} [S_c^2] + \mathbb{E} [S_q^2] - 2\mathbb{E} [S_c S_q] = I_c + I_q - 2\mathbb{E} [S_c S_q]$$

and it can be shown that  $\mathbb{E} [S_c S_q] = \mathbb{E} [S_q^2]$  (Why? - App. A.1.1). Thus from above, we have<sup>2</sup>

$$I_c - I_q = \mathbb{E} [(S_c - S_q)^2] \geq 0, \quad (1.16)$$

as the **right-hand side (RHS)** is the expectation of a squared function, the FI difference is nonnegative, meaning that the FI for quantized measurements is always less or equal to its continuous measurement equivalent. Therefore, as the corresponding CRB will have larger or equal values, it is clear, as it was already expected, that quantization of measurements reduces estimation performance (see Fig. 1.2 for the two estimation settings).

<sup>2</sup>Special attention must be given to the fact that to obtain (1.16), the measurement PDF form  $f(y-x)$  is not used, in the proof in App. A.1.1 a general form  $f(y;x)$  is used, thus the conclusion above is also valid for general parameter estimation problems, not only location parameter estimation.

### Loss with respect to the number of quantization intervals

Even if performance loss is positive or zero, nothing guarantees, until now, that estimation performance increases with increasing  $N_I$ , as it is intuitively expected. We will suppose that we have a threshold set  $\boldsymbol{\tau}$  for  $N_I$  quantization intervals. We will suppose  $\varepsilon = 0$  for simplification. We will add one threshold  $\tau'$  between two thresholds  $\tau_{i-1}$  and  $\tau_i$  ( $\tau_i > \tau' > \tau_{i-1}$ ),  $i > 0$  is assumed only to simplify notation. The sum elements defining  $I_q$  does not change, except for the term corresponding to interval  $q_i$ . The old and new FI only for this region are respectively

$$I_{q,i}^{\boldsymbol{\tau}} = \frac{[f(\tau_i) - f(\tau_{i-1})]^2}{F(\tau_i) - F(\tau_{i-1})} = \left[ \frac{f(\tau_i) - f(\tau_{i-1})}{F(\tau_i) - F(\tau_{i-1})} \right]^2 [F(\tau_i) - F(\tau_{i-1})], \quad (1.17)$$

$$I_{q,i}^{\{\boldsymbol{\tau}\} \cup \{\tau'\}} = \frac{[f(\tau_i) - f(\tau')]^2}{F(\tau_i) - F(\tau')} + \frac{[f(\tau') - f(\tau_{i-1})]^2}{F(\tau') - F(\tau_{i-1})}. \quad (1.18)$$

We can expand (1.17) adding and subtracting a term  $f(\tau')$  in the numerator of the first factor, adding and subtracting  $F(\tau')$  in the denominator of the first factor and multiplying and dividing the results numerator terms by  $F(\tau_i) - F(\tau')$  and  $F(\tau') - F(\tau_{i-1})$ . This gives

$$I_{q,i}^{\boldsymbol{\tau}} = \left\{ \frac{\frac{[f(\tau_i) - f(\tau')]}{[F(\tau_i) - F(\tau')]} [F(\tau_i) - F(\tau')] + \frac{[f(\tau') - f(\tau_{i-1})]}{[F(\tau') - F(\tau_{i-1})]} [F(\tau') - F(\tau_{i-1})]}{[F(\tau_i) - F(\tau')] + [F(\tau') - F(\tau_{i-1})]} \right\}^2 \times \{ [F(\tau_i) - F(\tau')] + [F(\tau') - F(\tau_{i-1})] \}. \quad (1.19)$$

The Jensen's inequality tells us the following [Hardy 1988, p. 74]: for a sequence of values  $a_i$ , positive weights  $b_i$  and a convex function  $\phi$  we have

$$\phi \left( \frac{\sum_i a_i b_i}{\sum_i b_i} \right) \leq \frac{\sum_i b_i \phi(a_i)}{\sum_i b_i}. \quad (1.20)$$

Multiplying both sides of (1.20) by  $\sum_i b_i$  and identifying in (1.19)  $b_i$  with  $F(\tau_i) - F(\tau')$  and  $F(\tau') - F(\tau_{i-1})$ ,  $a_i$  with  $\frac{[f(\tau_i) - f(\tau')]}{[F(\tau_i) - F(\tau')]}$  and  $\frac{[f(\tau') - f(\tau_{i-1})]}{[F(\tau') - F(\tau_{i-1})]}$  and  $\phi(x)$  with  $x^2$ , we have the following:

$$I_{q,i}^{\boldsymbol{\tau}} \leq I_{q,i}^{\{\boldsymbol{\tau}\} \cup \{\tau'\}}. \quad (1.21)$$

As it was expected, adding a threshold, or equivalently a quantizer interval, increases the FI and, as consequence, it decreases the CRB, enhancing estimation performance. Note that this is also true if we start with an optimal partition (a partition that maximizes the FI) and we add a threshold arbitrarily, however, in this case, the final interval partition may not be optimal within the class of quantizers with  $N_I + 1$  intervals, even if we try to optimize the new threshold position.

As adding thresholds increases the FI and as  $I_q$  is bounded above by  $I_c$ , the FI tends to a limit value when  $N_I$  tends to infinity. An interesting point to be studied is to know if we can make it converge to  $I_c$ . This will be done in Part II, where we will see that, under some regularity assumptions on the quantizer intervals,  $I_q$  converges to  $I_c$ .

Now, to have a more precise characterization of the estimation performance as a function of  $N_I$ , we must first describe how it is influenced by  $\boldsymbol{\tau}$ . For the optimal  $\boldsymbol{\tau}$ , we will be able to obtain the dependence of the estimation performance only on  $N_I$  and the noise characteristics.

### 1.3 Binary quantization

We begin the analysis by the binary case,  $N_I = 2$ . For binary observations ( $\tau'_{-1} = -\infty$  and  $\tau'_1 = \infty$ ), the CRB for  $N$  measurements can be written by using (1.13) in the CRB. As  $f(\varepsilon + \tau'_1) = 0$ ,  $f(\varepsilon + \tau'_{-1}) = 0$ ,  $1 - F(\varepsilon + \tau'_1) = 0$  and  $F(\varepsilon + \tau'_{-1}) = 0$  by assumption AN2, we obtain

$$\text{CRB}_q^B = \frac{F(\varepsilon)[1 - F(\varepsilon)]}{Nf^2(\varepsilon)}. \quad (1.22)$$

The analysis of performance in this case reduces to the analysis of the function

$$B(\varepsilon) = N\text{CRB}_q^B = \frac{F(\varepsilon)[1 - F(\varepsilon)]}{f^2(\varepsilon)}. \quad (1.23)$$

#### 1.3.1 The Gaussian case

This function was studied in the Gaussian noise case in [Papadopoulos 2001] and revisited in [Ribeiro 2006a]. In this case,

$$f(\varepsilon) = \frac{1}{\sqrt{\pi}\delta} \exp\left[-\left(\frac{\varepsilon}{\delta}\right)^2\right], \quad (1.24)$$

where  $\delta$  is the noise scale factor, which can be linearly related to the standard deviation  $\sigma$  ( $\delta = \sqrt{2}\sigma$ ). By plotting  $B$  as a function of  $\varepsilon$  (see Fig. 1.3), it was noted in [Papadopoulos 2001] that the minimum value  $B^*$  is attained for  $\varepsilon = 0$  and that  $B(\varepsilon)$  increases when  $|\varepsilon|$  increases. Thus, the optimal threshold  $\tau_0^*$  must be equal to  $x$  and the minimum value of  $B(\varepsilon)$  is  $B^* = \frac{1}{4f^2(0)} = \frac{\pi\delta^2}{4}$ . We can compare the CRB for one continuous measurement  $B_c = \frac{1}{I_c}$  with  $B^*$ , to have an idea about the loss of performance. Using (1.14), (1.15) and the expression for the PDF of the Gaussian distribution (1.24), we have:

$$B_c = \frac{1}{I_c} = \frac{1}{\mathbb{E}\left\{\left[\frac{\partial \log f(y-x)}{\partial x}\right]^2\right\}} = \frac{\delta^2}{2} = \frac{2}{\pi}B^*,$$

or equivalently

$$B^* = \frac{\pi}{2}B_c \approx 1.57B_c.$$

The performance loss due to binary quantization is surprisingly small. However, note that this requires  $\tau_0 = x$ , which is impossible to do in practice as  $x$  is the unknown parameter to be estimated. For increasing  $|\varepsilon| > 0$ , we can observe that  $B$  increases in a rather sensitive way.

An upper bound on  $B$  was given in [Ribeiro 2006a] by noting that the product in the numerator can be bounded by the following exponential (Why? - App. A.1.2):

$$F(\varepsilon)[1 - F(\varepsilon)] \leq \frac{1}{4} \exp\left[-\left(\frac{\varepsilon}{\delta}\right)^2\right]. \quad (1.25)$$

This bound can be used in (1.23) with (1.24) to obtain

$$B(\varepsilon) \leq \bar{B}(\varepsilon) = \frac{\pi}{4} \exp\left[+\left(\frac{\varepsilon}{\delta}\right)^2\right], \quad (1.26)$$

which is a function that increases exponentially with  $\varepsilon$ . To confirm that the bound is tight, at least for moderate  $\varepsilon$ , we plot the function  $\bar{B}$  also in Fig. 1.3.

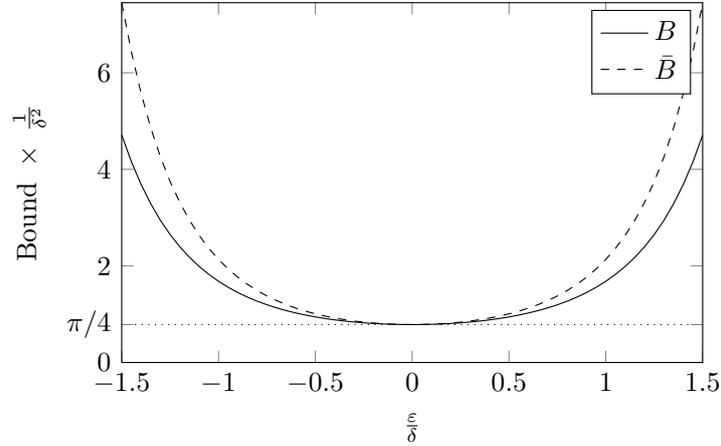


Figure 1.3: Quantity related to the CRB for quantized measurements,  $B$ , as a function of the normalized difference  $\frac{\varepsilon}{\delta}$  between threshold and parameter.  $\bar{B}$  is its upper bound, which has an exponential form. The noise distribution is the Gaussian distribution and the normalizing factor  $\delta$  is the Gaussian noise scale parameter. The normalizations in both axis are done to be able to have a plot independent of  $\delta$ .

Therefore, for the Gaussian case, we can conclude that the estimation performance loss for the binary case is relatively small if we set the threshold at the true parameter value, but it increases rapidly when we quantize far from it.

### 1.3.2 The Laplacian case

We can try to look to another symmetric unimodal distribution to see if the same happens. For example, we can consider the Laplacian distribution, whose PDF and CDF are

$$f(\varepsilon) = \frac{1}{2\delta} \exp\left(-\left|\frac{\varepsilon}{\delta}\right|\right), \quad (1.27) \quad F(\varepsilon) = \frac{1}{2} + \frac{\text{sign}(\varepsilon)}{2} \left[1 - \exp\left(-\left|\frac{\varepsilon}{\delta}\right|\right)\right], \quad (1.28)$$

where  $\text{sign}(\varepsilon)$  is the sign function

$$\text{sign}(\varepsilon) = \begin{cases} 1 & , \text{if } \varepsilon > 0, \\ 0 & , \text{if } \varepsilon = 0, \\ -1 & , \text{if } \varepsilon < 0. \end{cases}$$

Applying (1.27) and (1.28) to (1.23), we get

$$\begin{aligned} B &= \frac{\left\{ \frac{1}{2} - \frac{\text{sign}(\varepsilon)}{2} \left[1 - \exp\left(-\left|\frac{\varepsilon}{\delta}\right|\right)\right] \right\} \left[ \frac{1}{2} + \frac{\text{sign}(\varepsilon)}{2} \left[1 - \exp\left(-\left|\frac{\varepsilon}{\delta}\right|\right)\right] \right]}{\frac{1}{4\delta^2} \exp\left(-2\left|\frac{\varepsilon}{\delta}\right|\right)} \\ &= \frac{\frac{1}{4} - \frac{1}{4} \left[1 - \exp\left(-\left|\frac{\varepsilon}{\delta}\right|\right)\right]^2}{\frac{1}{4\delta^2} \exp\left(-2\left|\frac{\varepsilon}{\delta}\right|\right)} = \frac{\frac{1}{2} \exp\left(-\left|\frac{\varepsilon}{\delta}\right|\right) - \frac{1}{4} \exp\left(-2\left|\frac{\varepsilon}{\delta}\right|\right)}{\frac{1}{4\delta^2} \exp\left(-2\left|\frac{\varepsilon}{\delta}\right|\right)} \\ &= \delta^2 \left[ 2 \exp\left(\left|\frac{\varepsilon}{\delta}\right|\right) - 1 \right] \end{aligned} \quad (1.29)$$

and we can see that  $B$  and consequently the CRB is minimized for  $\tau_0 = x$  and that it is sensible to  $\varepsilon$ , growing exponentially when we increase  $|\varepsilon| = |\tau_0 - x|$ .

### 1.3.3 The general case

We can try to verify if the increasing behavior of  $B(\varepsilon)$  w.r.t.  $|\varepsilon|$  will be observed in the general case, when the noise PDF is unimodal and symmetric.

#### Attempt of global analysis: dead end $\bullet$

For unimodal symmetric distributions we have that  $f(\varepsilon) = f(-\varepsilon)$  and  $F(\varepsilon) = 1 - F(-\varepsilon)$ . Therefore, as it was observed for the specific Gaussian and Laplacian cases,  $B(\varepsilon)$  is a symmetric function. For analyzing if the increasing behavior is true in general, we can concentrate the analysis on the first derivative of  $B$  w.r.t.  $\varepsilon$ , for  $\varepsilon > 0$ . The derivative is

$$\frac{dB}{d\varepsilon} = \frac{f^2(\varepsilon)[1 - 2F(\varepsilon)] - 2F(\varepsilon)[1 - F(\varepsilon)]f^{(1)}(\varepsilon)}{f^3(\varepsilon)}, \quad (1.30)$$

where  $f^{(1)}(\varepsilon)$  is the first derivative of the PDF w.r.t.  $\varepsilon$ , supposed to exist<sup>3</sup>. Observe that if the distribution is symmetric, we have  $1 - 2F(0) = 0$  and only the second term in the numerator can be nonzero for  $\varepsilon = 0$ . Adding the condition that  $f^{(1)}(0) = 0$  makes  $\varepsilon = 0$  to be a local extremum of  $B$ , being a candidate point to be a local minimum.

In a first attempt to verify if  $\varepsilon = 0$  is a global minimum, we can calculate the second derivative and look if its sign is negative for all  $\varepsilon$ . If we calculate the second derivative we get

$$\frac{d^2B}{d\varepsilon^2} = \frac{-3f^2(\varepsilon)f^{(1)}(\varepsilon)[1 - 2F(\varepsilon)] + F(\varepsilon)[1 - F(\varepsilon)][6f^{(1)2}(\varepsilon) - 2f(\varepsilon)f^{(2)}(\varepsilon)]}{f^4(\varepsilon)} - 2, \quad (1.31)$$

with  $f^{(2)}(\varepsilon)$  the second derivative supposed to exist<sup>3</sup>. Even using the assumptions on the noise distribution, we cannot get any conclusion on the sign of the second derivative. Thus, we can try to go back to the first derivative and analyze its sign. Using the symmetry of  $B$ , a sufficient condition for  $\varepsilon = 0$  to be a global maximum is that  $\frac{dB}{d\varepsilon} \geq 0$  for  $\varepsilon > 0$ . The derivative  $\frac{dB}{d\varepsilon}$  has the same sign of the numerator in the RHS of (1.30), therefore, we can obtain the condition

$$-f^{(1)}(\varepsilon) \geq \frac{f^2(\varepsilon)[2F(\varepsilon) - 1]}{2F(\varepsilon)[1 - F(\varepsilon)]},$$

using the fact that the density is monotonically decreasing ( $f^{(1)}(\varepsilon) < 0$  for  $\varepsilon < 0$ ) and symmetric ( $[2F(\varepsilon) - 1] > 0$  for  $\varepsilon > 0$ ), we can write

$$|f^{(1)}(\varepsilon)| \geq \frac{f^2(\varepsilon)[2F(\varepsilon) - 1]}{2F(\varepsilon)[1 - F(\varepsilon)]}. \quad (1.32)$$

Unfortunately, by using the assumptions on the distribution we cannot go further. But, at least, we can use the condition above (1.32) to verify empirically, for commonly used noise

<sup>3</sup>This rules out the evaluation of this quantity for  $\varepsilon = 0$  in the Laplacian case, which is not a problematic case, as we know analytically that  $B$  is strictly increasing with  $|\varepsilon|$  in this case.

models, the increasing behavior of  $B(\varepsilon)$  with  $|\varepsilon|$ . For doing so, we (re)tested the Gaussian and Laplacian distributions with (1.32), we also added a heavy-tailed distribution<sup>4</sup> to see if in this case the conclusions change. The heavy-tailed distribution is the Cauchy distribution with PDF and CDF given respectively by

$$f(\varepsilon) = \frac{1}{\pi\delta} \frac{1}{\left[1 + \left(\frac{\varepsilon}{\delta}\right)^2\right]}, \quad (1.33) \quad F(\varepsilon) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{\varepsilon}{\delta}\right). \quad (1.34)$$

For the three distributions (Gaussian, Laplacian and Cauchy), we calculated the quantity  $M = |f^{(1)}(\varepsilon)| - \frac{f^2(\varepsilon)[2F(\varepsilon)-1]}{2F(\varepsilon)[1-F(\varepsilon)]}$ , which must be positive to have the monotonic increasing behavior of  $B$  w.r.t.  $|\varepsilon|$ . The result is displayed in Fig. 1.4, where we observe that this is indeed true.

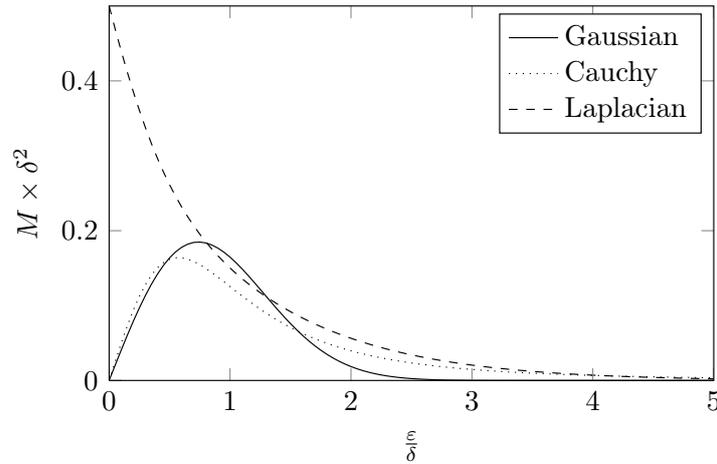


Figure 1.4:  $M \times \delta^2$  as a function of  $\frac{\varepsilon}{\delta} > 0$ . The plot is given for three types of noise distribution: Gaussian, Cauchy and Laplacian. All distributions have a noise scale parameter denoted  $\delta$ . The function must be positive for the optimal threshold in binary quantization to be exactly placed at the true parameter  $\tau^* = x$ . The normalizations in both axis are done to be able to have a plot independent of  $\delta$ .

### Local analysis

As condition (1.32) is difficult to verify in general, we can try to analyze the local behavior of  $B(\varepsilon)$  around  $\varepsilon = 0$ . Even if the results will be weaker, as they will only be local results, we can expect that the conditions for  $\varepsilon = 0$  to be a local minimum of  $B(\varepsilon)$  will be easy to verify.

We saw above that if  $f^{(1)}(0) = 0$ , then we have an extremum of  $B(\varepsilon)$  at  $\varepsilon = 0$ . If we use one more time the assumption  $f^{(1)}(0) = 0$  on the second derivative at zero and the symmetry ( $F(0)[1-F(0)] = \frac{1}{4}$ ), we get

$$\left. \frac{d^2 B}{d\varepsilon^2} \right|_{\varepsilon=0} = -\frac{1}{2} \frac{f^{(2)}(0)}{f^3(0)} - 2.$$

<sup>4</sup>A heavy-tailed distribution is a distribution whose ratio between  $1 - F(x+y)$  and  $1 - F(x)$  is equal to one when  $x$  tends to infinity [Sigman 1999]. A subclass of this family is the class of all sub-exponential distributions, where the Student-t distributions (for which the Cauchy distribution is a special case) and Paretian distributions are included.

For  $\varepsilon = 0$  to be a local minimum of  $B(\varepsilon)$ , we have the condition  $\left. \frac{d^2 B}{d\varepsilon^2} \right|_{\varepsilon=0} > 0$ . When we apply this condition to the expression above, we can obtain the following condition on the noise PDF and its second derivative:

$$-f^{(2)}(0) > 4f^3(0). \quad (1.35)$$

For the Gaussian distribution this condition is satisfied as we have

$$-f^{(2)}(0) = \frac{1}{\delta^3} \frac{2}{\sqrt{\pi}} > 4f^3(0) = \frac{1}{\delta^3} \frac{4}{\pi^{\frac{3}{2}}}$$

and also for the Cauchy distribution

$$-f^{(2)}(0) = \frac{1}{\delta^3} \frac{2}{\pi} > 4f^3(0) = \frac{1}{\delta^3} \frac{4}{\pi^3}.$$

### 1.3.4 Asymmetric threshold: surprising cases $\triangle!$

Surprisingly, we can find symmetric distributions and even a class of unimodal symmetric distributions for which the condition (1.35) is not satisfied, as a consequence, for these distributions,  $\varepsilon = 0$  can be a local maximum instead of a local minimum.

#### The uniform/Gaussian case

A simple way to define a symmetric distribution that does not satisfy (1.35) is to set the values of the PDF around zero to a nonzero constant, in this way  $f(0) > 0$  and  $f^{(2)}(0) = 0$ . This makes the second derivative at  $\varepsilon = 0$  to be negative, leading to a local maximum of  $B(\varepsilon)$  at that point.

As an example we can consider a noise PDF that is uniform in the interval  $[-\frac{\alpha}{2}, \frac{\alpha}{2}]$ , where  $\alpha \in \mathbb{R}_+$ , and that decreases as a Gaussian distribution with a standard deviation parameter  $\sigma$  outside this interval. We call this noise distribution the uniform/Gaussian distribution and the analytic expression for its PDF is

$$f(\varepsilon) = \begin{cases} f_{GL}(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon+\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{for } \varepsilon < -\frac{\alpha}{2}, \\ f_U(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma}, & \text{for } -\frac{\alpha}{2} \leq \varepsilon \leq \frac{\alpha}{2}, \\ f_{GR}(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon-\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{for } \varepsilon > \frac{\alpha}{2}, \end{cases} \quad (1.36)$$

where  $C = 1 + \frac{\alpha}{\sqrt{2\pi}\sigma}$  is a normalization constant that makes the integral of the PDF to be equal to one. This PDF is depicted in Fig. 1.5.

To obtain the function  $B(\varepsilon)$ , we have to describe the CDF of the uniform/Gaussian r.v.. If we denote  $\Phi(\varepsilon)$  the CDF of a standard Gaussian distribution (the CDF for a Gaussian with  $\sigma = 1$ ), we obtain the following:

$$F(\varepsilon) = \begin{cases} \frac{1}{C} \Phi\left(\frac{\varepsilon+\frac{\alpha}{2}}{\sigma}\right), & \text{for } \varepsilon < -\frac{\alpha}{2}, \\ \frac{1}{C} \left[ \frac{1}{2} + \frac{1}{\sqrt{2\pi}\sigma} \left(\varepsilon + \frac{\alpha}{2}\right) \right], & \text{for } -\frac{\alpha}{2} \leq \varepsilon \leq \frac{\alpha}{2}, \\ \frac{1}{C} \left[ \frac{\alpha}{\sqrt{2\pi}\sigma} + \Phi\left(\frac{\varepsilon+\frac{\alpha}{2}}{\sigma}\right) \right], & \text{for } \varepsilon > \frac{\alpha}{2}. \end{cases} \quad (1.37)$$

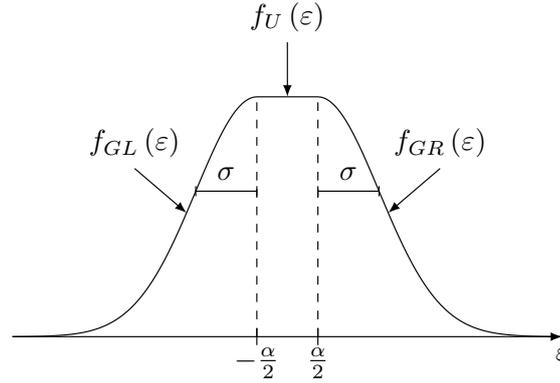


Figure 1.5: PDF for the uniform/Gaussian distribution. The center region is uniform with width  $\alpha$ , while the left and right sides are Gaussian with standard deviation parameter  $\sigma$ .

Using (1.36) and (1.37) in the expression for  $B(\varepsilon)$  (1.23), we get

$$\begin{aligned}
 B(\varepsilon) &= \frac{F(\varepsilon)[1 - F(\varepsilon)]}{f^2(\varepsilon)} = \\
 &= \begin{cases} 2\pi\sigma^2 \exp\left[\left(\frac{\varepsilon + \frac{\alpha}{2}}{\sigma}\right)^2\right] \Phi\left(\frac{\varepsilon + \frac{\alpha}{2}}{\sigma}\right) \left[C - \Phi\left(\frac{\varepsilon + \frac{\alpha}{2}}{\sigma}\right)\right], & \text{for } \varepsilon < -\frac{\alpha}{2}, \\ 2\pi\sigma^2 \left[\left(\frac{1}{2} + \frac{\alpha}{2} \frac{1}{\sqrt{2\pi}\sigma}\right)^2 - \frac{\varepsilon^2}{2\pi\sigma^2}\right], & \text{for } -\frac{\alpha}{2} \leq \varepsilon \leq \frac{\alpha}{2}, \\ 2\pi\sigma^2 \exp\left[\left(\frac{\varepsilon - \frac{\alpha}{2}}{\sigma}\right)^2\right] \left[\frac{\alpha}{\sqrt{2\pi}\sigma} + \Phi\left(\frac{\varepsilon - \frac{\alpha}{2}}{\sigma}\right)\right] \left[1 - \Phi\left(\frac{\varepsilon - \frac{\alpha}{2}}{\sigma}\right)\right], & \text{for } \varepsilon > \frac{\alpha}{2}. \end{cases} \quad (1.38)
 \end{aligned}$$

observe that in the interval  $[-\frac{\alpha}{2}, \frac{\alpha}{2}]$ , the function is concave, so we really have a local maximum at zero.

For observing the global behavior of this bound, we plotted  $\text{CRB}_q^B$  for a number of samples  $N = 500$ ,  $\alpha = 1$ ,  $\sigma = 1$  and for values of  $\varepsilon$  in the interval  $[-2, 2]$ . For verifying that the behavior of the bound was close to the true MSE of the MLE, we simulated the MLE  $10^5$  times for  $N = 500$ , the simulation results were used for evaluating a simulated MSE. The details on the implementation of the MLE for binary quantization will be presented further in (a1.1) in Sec. 1.3.6 and for more specific implementation details about the uniform/Gaussian case see (More? - App. A.2.2). The simulation of the noise was done by exploiting the fact that the uniform/Gaussian distribution is a mixture of distributions that are easy to sample (How? - App. A.3.1). The results are shown in Fig. 1.6.

We can observe the concave behavior of the bound around  $\varepsilon = 0$  and the presence of two minima at points different from  $\varepsilon = 0$ . This shows that for this type of noise, binary quantization must be done in an asymmetric way, by shifting the central threshold to a zone where the noise is not uniform. Note also that if we shift too much, the performance starts too degrade again. We suspect that this asymmetric behavior comes from the fact that for the uniform distribution the most informative points in statistical sense are the boundaries of the distribution (where it passes from a positive value to zero). Finally, we can also see that the MSE for the MLE is quite close to the bound, indicating that we can use the bound for analyzing the behavior of the MSE.

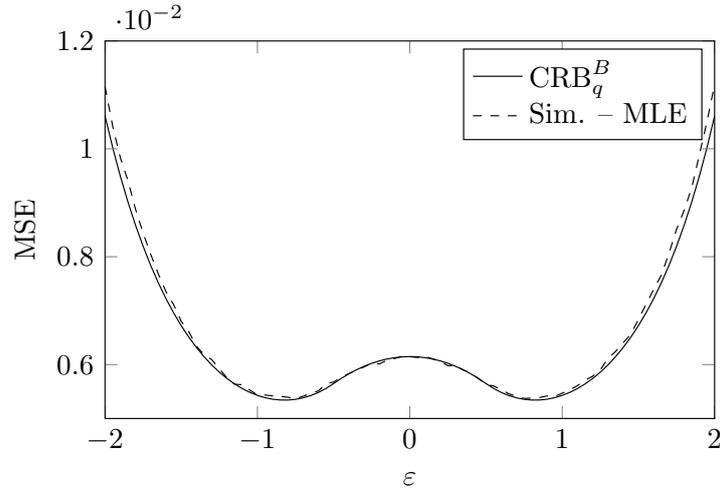


Figure 1.6:  $\text{CRB}_q^B$  and simulated MLE MSE for uniform/Gaussian noise. Both the bound and simulated MSE were evaluated for a number of samples  $N = 500$  and for  $\varepsilon$  in the interval  $[-2, 2]$ . The MSE for the MLE was evaluated through Monte Carlo simulation using  $10^5$  realizations of blocks with 500 samples. We considered the following noise parameters:  $\alpha = 1$  and  $\sigma = 1$ .

### The generalized Gaussian case

We can also look if there are noise distributions without the central uniform behavior for which the condition on the second derivative (1.35) is not respected. All distributions that have zero second derivative at  $\varepsilon = 0$  will not respect the condition. To have zero second derivative at zero, the PDF must be flat around zero. A class of distributions for which we can control the flatness around zero by changing a parameter is the **generalized Gaussian distribution (GGD)**. A more detailed presentation of the GGD will be given in Ch. 3 with the motivation for using it as a noise model. Here, we will present only its PDF and CDF, which are given respectively by

$$f(\varepsilon) = \frac{\beta}{2\delta\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left|\frac{\varepsilon}{\delta}\right|^\beta\right), \quad (1.39)$$

$$F(\varepsilon) = \frac{1}{2} \left[ 1 + \text{sign}(\varepsilon) \frac{\gamma\left(\frac{1}{\beta}, \left|\frac{\varepsilon}{\delta}\right|^\beta\right)}{\Gamma\left(\frac{1}{\beta}\right)} \right], \quad (1.40)$$

where  $\delta$  is the noise scale parameter,  $\beta$  is a shape parameter which allows for controlling the flatness around zero. Both  $\delta$  and  $\beta$  are constrained to be strictly positive  $\delta > 0$ ,  $\beta > 0$ .  $\Gamma(\cdot)$  is the gamma function

$$\Gamma(x) = \int_0^{+\infty} z^{x-1} \exp(-z) dz$$

and  $\gamma(\cdot, \cdot)$  is the incomplete gamma function

$$\gamma(x, w) = \int_0^w z^{x-1} \exp(-z) dz.$$

We need to calculate  $f^{(2)}(\varepsilon)$  at  $\varepsilon = 0$ . For doing so, we will evaluate the derivatives for  $\varepsilon < 0$  and  $\varepsilon > 0$  and then we will evaluate their limits when  $\varepsilon$  tends to zero.

For the first derivative we have

$$f^{(1)}(\varepsilon) = \begin{cases} D \left(\frac{-\varepsilon}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{-\varepsilon}{\delta}\right)^\beta\right], & \text{for } \varepsilon < 0, \\ -D \left(\frac{\varepsilon}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{\varepsilon}{\delta}\right)^\beta\right], & \text{for } \varepsilon > 0, \end{cases}$$

where  $D = \frac{\beta^2}{2\delta^2\Gamma\left(\frac{1}{\beta}\right)}$ . Observe that if  $\beta \leq 1$ , then the first derivative at zero is not defined. For  $\beta > 1$ , the derivative is zero.

For the evaluation of the second derivative we will consider  $\beta > 1$ . We get the following second derivative:

$$f^{(2)}(\varepsilon) = \begin{cases} D \left[-\frac{(\beta-1)}{\delta} \left(\frac{-\varepsilon}{\delta}\right)^{\beta-2} + \frac{\beta}{\delta} \left(\frac{-\varepsilon}{\delta}\right)^{2(\beta-1)}\right] \exp\left[-\left(\frac{-\varepsilon}{\delta}\right)^\beta\right], & \text{for } \varepsilon < 0, \\ D \left[-\frac{(\beta-1)}{\delta} \left(\frac{\varepsilon}{\delta}\right)^{\beta-2} + \frac{\beta}{\delta} \left(\frac{\varepsilon}{\delta}\right)^{2(\beta-1)}\right] \exp\left[-\left(\frac{\varepsilon}{\delta}\right)^\beta\right], & \text{for } \varepsilon > 0. \end{cases}$$

We can see that for  $1 < \beta < 2$ , the derivatives when  $\varepsilon$  approaches zero are both  $-\infty$ . For these cases, the point  $\varepsilon = 0$  is a local minimum of  $B(\varepsilon)$ . In the Gaussian case  $\beta = 2$ , the second derivative has a finite negative value and we saw before that  $\varepsilon = 0$  is a local minimum (empirically we also observed that it is a global minimum). For the cases  $\beta > 2$ , the second derivative is zero, thus corresponding to the special cases of local maximum that we were looking for.

The function  $B(\varepsilon)$ , that we expect to be a "w" shaped function for  $\beta > 2$ , can be evaluated using (1.39) and (1.40) in the expression for  $B(\varepsilon)$  (1.23). This gives

$$B(\varepsilon) = \frac{F(\varepsilon) [1 - F(\varepsilon)]}{f^2(\varepsilon)} = \frac{\delta^2 \Gamma^2\left(\frac{1}{\beta}\right)}{\beta^2} \left[1 - \frac{\gamma^2\left(\frac{1}{\beta}, \left|\frac{\varepsilon}{\delta}\right|^\beta\right)}{\Gamma^2\left(\frac{1}{\beta}\right)}\right] \exp\left(2 \left|\frac{\varepsilon}{\delta}\right|^\beta\right). \quad (1.41)$$

As in the uniform/Gaussian case, we also plotted  $\text{CRB}_q^B$  and the simulated MSE of the MLE. We used  $N = 500$ ,  $\beta = 4$ ,  $\delta = 1$  and values of  $\varepsilon$  in the interval  $[-1, 1]$ . We simulated the MLE  $10^5$  times and the results were used to obtain an estimate of the true MSE for this estimator. For more specific implementation details about the MLE in the GGD case see (More? - App. A.2.3). The GGD noise was generated using transformations of gamma variates (How? - App. A.3.2). The results are shown in Fig. 1.7.

We can notice again that the optimal threshold must be placed in an asymmetric way and also that the simulated estimation performance is close to the bound. Contrary to the

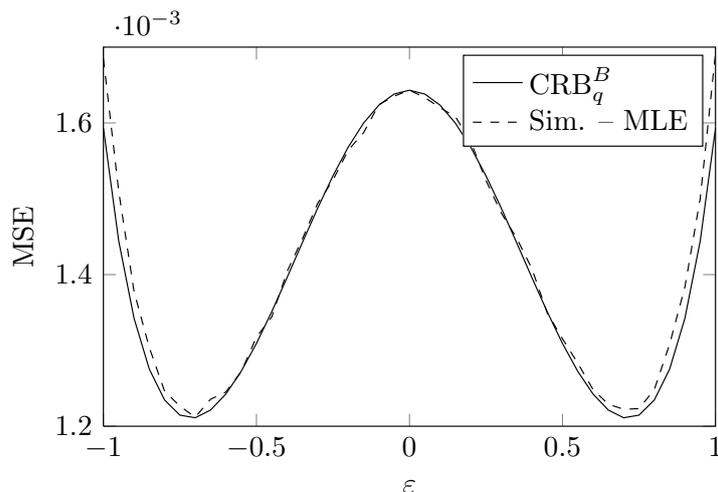


Figure 1.7:  $\text{CRB}_q^B$  and simulated MLE MSE for GGD noise. Both the bound and simulated MSE were evaluated for a number of samples  $N = 500$  and for  $\varepsilon$  in the interval  $[-1, 1]$ . The MSE for the MLE was evaluated through Monte Carlo simulation using  $10^5$  realizations of blocks with 500 samples. We considered the following noise parameters:  $\beta = 4$  and  $\delta = 1$ .

uniform/Gaussian case, we cannot have a clear interpretation on the position of the minimum point. The minimum point was observed to be sensible to changes in  $\beta$  and  $\delta$ . It was also observed that as we set  $\beta$  closer to 2 (the Gaussian case), we obtain a difference of performance between the point  $\varepsilon = 0$  and the minimum point that gets smaller. On the other hand, as we increase  $\beta$  (getting closer to the uniform distribution), the difference seems to increase.

### 1.3.5 Conclusions on binary quantization performance

To conclude, we can say that the best estimation performance in the binary case for commonly used noise models ( $\text{CRB}_q^B$ ) is obtained for  $\varepsilon = 0$  or  $\tau_0 = x$ :

$$\text{CRB}_q^{B,*} = \frac{F(0)[1 - F(0)]}{Nf^2(0)} = \frac{1}{4Nf^2(0)}, \quad (1.42)$$

which is also a lower bound on the asymptotically achievable performance. However, even under the unimodal symmetric assumption, this rather intuitive conclusion is not always true. From the local condition on the second derivative, we can see that if the noise PDF is slightly flat around zero, then a "w" shaped performance function will appear, leading to an optimal threshold that might be placed asymmetrically w.r.t. to its input r.v. distribution.

The variation between the performance for the point  $\varepsilon = 0$  and the minimum  $\text{CRB}_q^{B,*}$  in the asymmetric cases seems to depend on the flatness of the distribution. An increased flatness around zero, seems to be related to an increased performance variation. This strong dependence between the shape of the CRB and the noise distribution seems to be a good subject for future work.

Another interesting direction for future work on this issue about asymmetry is to analyze how it can appear on the detection problem using binary quantized measurements. It

appears that such behavior will be present for the same noise distributions considered above (uniform/Gaussian and GGD) in the problem of local optimum detection of signals based on binary quantized measurements. For this problem, it can be shown that the asymptotic performance depends also on the FI for quantized measurements [Kassam 1977].

### 1.3.6 MLE for binary quantization

The specific implementation of (a1) in the binary case with a fixed threshold can be done in a simple way [Papadopoulos 2001] (and revisited in [Ribeiro 2006a]). The sequence of  $N$  quantized measurements can be observed as a sequence of  $N$  i.i.d. samples from a Bernoulli distribution with probability  $p = \mathbb{P}(i_k = 1) = 1 - F(\tau_0 - x)$ . Thus, hiding the functional dependency on  $x$  and  $\tau_0$ , we can calculate the likelihood of  $p$  with the sequence  $i_{1:N}$ .

The likelihood of  $p$  for  $i_{1:N}$  can be written in a simple form by observing the following:

- for a measurement  $i_k$ ,  $\mathbb{P}(i_k = 1; p) = p$  and  $\mathbb{P}(i_k = -1; p) = 1 - p$ . We can write  $\mathbb{P}(i_k; p)$  in a form  $p^{f_1(i_k)} (1 - p)^{f_{-1}(i_k)}$ , where the functions  $f_1$  and  $f_{-1}$  are respectively 1 and 0 when  $i_k = 1$ , and 0 and 1 when  $i_k = -1$ . A Simple choice for these functions is  $f_1(i_k) = \frac{i_k+1}{2}$  and  $f_{-1}(i_k) = \frac{1-i_k}{2}$ .
- As the measurements are independent, the likelihood for the sequence  $i_{1:N}$  will be the product of the marginal likelihoods  $\mathbb{P}(i_k = 1; p)$ .

This leads to

$$L(p, i_{1:N}) = \prod_{k=1}^N p^{\frac{i_k+1}{2}} (1-p)^{\frac{1-i_k}{2}}.$$

Calculating its logarithm and then evaluating the MLE for  $p$  denoted  $\hat{P}_{ML}$ , we get the following [Wasserman 2003, p. 123]:

$$\hat{P}_{ML} = \frac{1}{N} \sum_{k=1}^N \frac{1+i_k}{2}. \quad (1.43)$$

The MLE (in general) has the property that if we want to estimate a parameter  $x$  which is an invertible function of  $z$ ,  $x = g(z)$  and we know the MLE for  $z$ ,  $\hat{Z}_{ML}$ , then the MLE for  $x$  is  $\hat{X}_{ML} = g(\hat{Z}_{ML})$  [Kay 1993, p. 176]. This property is known as functional invariance. For our problem we can write

$$x = g(p) = \tau_0 - F^{-1}(1 - p), \quad (1.44)$$

$F^{-1}$  is the inverse of the noise CDF. By definition  $F^{-1}$  is invertible, as  $F$  is strictly increasing due to the monotonicity assumption on  $F$ , so the function  $g$  in this case is invertible. Thus, by the functional invariance of the MLE, we can obtain  $\hat{X}_{ML,q}$ , after replacing  $p$  in (1.44) by  $\hat{P}_{ML}$  given by (1.43). This leads to an analytical expression for the MLE:

$$\hat{X}_{ML,q} = g(\hat{P}_{ML}) = \tau_0 - F^{-1}\left(1 - \hat{P}_{ML}\right) = \tau_0 - F^{-1}\left[\frac{1}{2}\left(1 - \frac{1}{N} \sum_{k=1}^N i_k\right)\right]. \quad (1.45)$$

Therefore, the solution to problem (a) (p. 27) in the binary case can be detailed as follows

**Solution to (a) - MLE for binary quantized measurements and fixed threshold  $\tau_0$**

**(a1.1) 1) Estimator**

$$\begin{aligned}\hat{X}_{ML,q} = g(\hat{P}_{ML}) &= \tau_0 - F^{-1}\left(1 - \hat{P}_{ML}\right) \\ &= \tau_0 - F^{-1}\left[\frac{1}{2}\left(1 - \frac{1}{N}\sum_{k=1}^N i_k\right)\right].\end{aligned}$$

**2) Performance (asymptotic)**

$\hat{X}_{ML,q}$  is asymptotically unbiased

$$\mathbb{E}\left[\hat{X}_{ML,q}\right]_{N \rightarrow \infty} = x$$

and its asymptotic MSE or variance is given by

$$\text{Var}\left(\hat{X}_{ML,q}\right)_{N \rightarrow \infty} \sim \text{CRB}_q^B = \frac{F(\tau_0 - x)[1 - F(\tau_0 - x)]}{Nf^2(\tau_0 - x)},$$

which is minimal for commonly used noise models (Gaussian, Laplacian and Cauchy distributions) if  $\tau_0 = x$ , attaining  $\frac{1}{4Nf^2(0)}$  and increases with  $|\tau_0 - x|$ .

Notice that this algorithm can be used for any noise distribution, not only for symmetric unimodal distributions.

## 1.4 Multibit quantization

Now, we study the multiple interval (multibit) case,  $N_I > 2$ . The expression characterizing estimation performance for this case is given by (1.13):

$$I_q(\varepsilon) = \sum_{i_k=1}^{i_k=\frac{N_I}{2}} \left\{ \frac{[f(\varepsilon + \tau'_{i_k}) - f(\varepsilon + \tau'_{i_k-1})]^2}{F(\varepsilon + \tau'_{i_k}) - F(\varepsilon + \tau'_{i_k-1})} + \frac{[f(\varepsilon - \tau'_{i_k-1}) - f(\varepsilon - \tau'_{i_k})]^2}{F(\varepsilon - \tau'_{i_k-1}) - F(\varepsilon - \tau'_{i_k})} \right\}.$$

We remind that a larger  $I_q(\varepsilon)$  gives a better asymptotic estimation performance. We will start by analyzing the influence of the central threshold.

For verifying symmetry, we replace  $\varepsilon$  by  $-\varepsilon$ .

$$I_q(-\varepsilon) = \sum_{i_k=1}^{i_k=\frac{N_I}{2}} \left\{ \frac{[f(-\varepsilon + \tau'_{i_k}) - f(-\varepsilon + \tau'_{i_k-1})]^2}{F(-\varepsilon + \tau'_{i_k}) - F(-\varepsilon + \tau'_{i_k-1})} + \frac{[f(-\varepsilon - \tau'_{i_k-1}) - f(-\varepsilon - \tau'_{i_k})]^2}{F(-\varepsilon - \tau'_{i_k-1}) - F(-\varepsilon - \tau'_{i_k})} \right\}.$$

The following equalities come from the symmetry assumptions:

$$\begin{aligned} f(-\varepsilon + \tau'_{i_k}) &= f(\varepsilon - \tau'_{i_k}), & F(-\varepsilon + \tau'_{i_k}) &= 1 - F(\varepsilon - \tau'_{i_k}), \\ f(-\varepsilon + \tau'_{i_k-1}) &= f(\varepsilon - \tau'_{i_k-1}), & F(-\varepsilon + \tau'_{i_k-1}) &= 1 - F(\varepsilon - \tau'_{i_k-1}), \\ f(-\varepsilon - \tau'_{i_k-1}) &= f(\varepsilon + \tau'_{i_k-1}), & F(-\varepsilon - \tau'_{i_k-1}) &= 1 - F(\varepsilon + \tau'_{i_k-1}), \\ f(-\varepsilon - \tau'_{i_k}) &= f(\varepsilon + \tau'_{i_k}), & F(-\varepsilon - \tau'_{i_k}) &= 1 - F(\varepsilon + \tau'_{i_k}). \end{aligned}$$

Applying these expressions to  $I_q(-\varepsilon)$  and multiplying by  $-1$  inside the squared terms we get that  $I_q(\varepsilon) = I_q(-\varepsilon)$ , thus, the even symmetry observed in the binary case can be extended to this case.

#### 1.4.1 The Laplacian case

Now, we start with the Laplacian case which is easy to be treated analytically. If we set  $\varepsilon = 0$ ,

$$I_q(0) = \sum_{i_k=1}^{i_k=\frac{N_I}{2}} \left\{ \frac{[f(\tau'_{i_k}) - f(\tau'_{i_k-1})]^2}{F(\tau'_{i_k}) - F(\tau'_{i_k-1})} + \frac{[f(-\tau'_{i_k-1}) - f(-\tau'_{i_k})]^2}{F(-\tau'_{i_k-1}) - F(-\tau'_{i_k})} \right\},$$

using also the symmetry assumption (similar development as above), one can easily observe that the second term inside the sum terms is equal to the first, which means that we can rewrite the sum as

$$I_q(0) = 2 \sum_{i_k=1}^{i_k=\frac{N_I}{2}} \left\{ \frac{[f(\tau'_{i_k}) - f(\tau'_{i_k-1})]^2}{F(\tau'_{i_k}) - F(\tau'_{i_k-1})} \right\}. \quad (1.46)$$

Using the PDF and CDF for the Laplacian distribution (1.27) and (1.28), separating the last term of the sum and simplifying the notation for the absolute value and sign functions ( $\tau'_{i_k} \geq 0$ ), we obtain

$$\begin{aligned} I_q(0) &= 2 \left\{ \left\{ \sum_{i_k=1}^{i_k=\frac{N_I}{2}-1} \frac{\frac{1}{4\delta^2} \left[ \exp\left(-\frac{\tau'_{i_k}}{\delta}\right) - \exp\left(-\frac{\tau'_{i_k-1}}{\delta}\right) \right]^2}{\frac{1}{2} \left[ \exp\left(-\frac{\tau'_{i_k-1}}{\delta}\right) - \exp\left(-\frac{\tau'_{i_k}}{\delta}\right) \right]} \right\} + \frac{\frac{1}{4\delta^2} \left[ \exp\left(-\frac{\tau'_{\frac{N_I}{2}-1}}{\delta}\right) \right]^2}{\frac{1}{2} \exp\left(-\frac{\tau'_{\frac{N_I}{2}-1}}{\delta}\right)} \right\} \\ &= \frac{1}{\delta^2} \left\{ \left\{ \sum_{i_k=1}^{i_k=\frac{N_I}{2}-1} \left[ \exp\left(-\frac{\tau'_{i_k-1}}{\delta}\right) - \exp\left(-\frac{\tau'_{i_k}}{\delta}\right) \right] \right\} + \exp\left(-\frac{\tau'_{\frac{N_I}{2}-1}}{\delta}\right) \right\}. \end{aligned}$$

The terms inside the sum (in the  $\Sigma$  operator) cancel each other except for the first and last term, the last term and the term outside the sum also cancel each other.  $I_q(0)$  is then given

by only one term, which is  $I_q(0) = \frac{1}{\delta^2} \exp\left(-\frac{\tau'_0}{\delta}\right)$ , as  $\tau'_0 = 0$ , we have

$$I_q(0) = \frac{1}{\delta^2}.$$

Surprisingly, this is exactly the same as the FI for continuous measurements (Why? - App. A.1.3). Thus, this means that not only  $\tau_0 = x$  is optimal for the Laplacian distribution but also that no loss of performance is observed. As the quantized measurement FI can only increase by adding quantization intervals and as it is upper bounded by the continuous FI, we see that once we have placed the threshold at  $x$ , the quantized measurement FI will be the same for all  $N_I \geq 2$ . This means that in practice, as we want to minimize the rate, the optimal choice of number of quantization intervals will be  $N_I = 2$ .

### 1.4.2 The Gaussian and Cauchy cases under uniform quantization

Instead of diving into calculus for trying to obtain some characterization of  $I_q$  as a function of  $\varepsilon$ , we preferred to directly plot its influence for a given set of thresholds. We evaluated  $I_q$  given by (1.13) as a function of  $\frac{\varepsilon}{\delta}$  with  $\frac{\varepsilon}{\delta} \in [-10, 10]$ . The evaluation was done for the Gaussian and Cauchy distributions. The quantizer was assumed to have  $N_I = 8$  and a uniform step  $\Delta$  between thresholds, which means that  $\boldsymbol{\tau}' = [0 \quad \Delta \quad 2\Delta \quad 3\Delta \quad +\infty]$ . Here, uniform quantization was assumed only to simplify the presentation.<sup>5</sup> Three different  $\Delta$  were chosen for the evaluation,  $\Delta = 0.1\delta$ ,  $\Delta^*$  and  $2\delta$ .  $\Delta^*$  was chosen as the maximizer of  $I_q$  when  $\varepsilon = 0$  and it was obtained by exhaustive search. The results are given in Fig. 1.8 where the continuous FI  $I_c$  is also plotted for comparison. Remember that for the Gaussian distribution  $I_c = \frac{2}{\delta^2}$ . For the Cauchy distribution we have  $I_c = \frac{1}{2\delta^2}$  (Why? - App. A.1.4).

Observe that in all cases the point  $\varepsilon = 0$  gives maximum  $I_q$ . Note that differently from the binary case, the FI does not strictly decrease when  $|\varepsilon|$  increases, this only happens when  $|\varepsilon|$  is outside the quantizer range. We can also see that the optimal  $\Delta$  gives  $I_q$  values very close to  $I_c$ .

It is also interesting to observe that when we choose  $\Delta$  very large compared with  $\Delta^*$ , we obtain a maximum  $I_q$  smaller than for  $\Delta^*$ , but this  $I_q$  does not decrease to zero inside the quantizer range. This indicates that when we have a prior information on the interval of values where  $x$  is located, then a more robust solution can be found by using a large quantization step (for example by using a  $\Delta$  that is equal to the prior interval length divided by  $N_I$ ). Clearly in this case, the price to pay is that even if we have  $\varepsilon = 0$  the performance is lower than the optimal, being very close to the performance for a binary quantizer.

Differently from the binary case, after evaluating  $I_q(\varepsilon)$  for the GGD with  $\beta > 2$  and  $N_I > 2$ , it was observed that when we use  $\Delta^*$  as quantization step, the symmetric quantizer assumption seems to force the performance to be optimal for  $\varepsilon = 0$ . Less surprisingly now, when the quantization step is chosen too large, the asymmetric behavior appears, this is due to the fact that the performance around  $\varepsilon = \tau'_i$  is very close to the binary quantizer performance. In the same way as for the other noise distributions considered above, it was also observed that when the parameter is outside the quantizer range the performance is degraded.

<sup>5</sup>It will be shown in Part II, that for large  $N_I$ , the optimal quantization intervals may not be uniform.

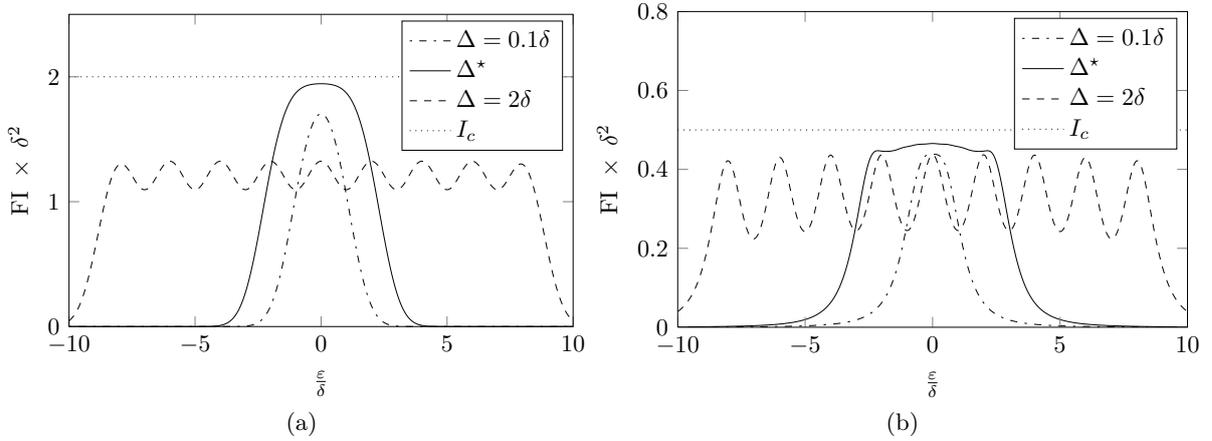


Figure 1.8: FI for a range  $[-10, 10]$  of normalized difference  $\frac{\varepsilon}{\delta}$  between the central threshold  $\tau_0$  and the true parameter  $x$ . The quantizer with  $N_I = 8$  is uniform with quantization interval length (in the granular region)  $\Delta$ . In (a), the noise distribution is Gaussian and  $\Delta = 0.1\delta, \Delta^* = 0.399\delta, 2\delta$ .  $\Delta^*$  is the optimal quantization step for  $\frac{\varepsilon}{\delta} = 0$ . In (b), the Cauchy noise distribution is used and  $\Delta = 0.1\delta, \Delta^* = 0.5878\delta, 2\delta$ .  $\Delta^*$  is also the optimal quantization step for  $\frac{\varepsilon}{\delta} = 0$ . For both cases,  $I_c$  is the FI for the continuous measurement. In the Gaussian case  $I_c = 2\delta^2$ , while in the Cauchy case  $I_c = \frac{1}{2}\delta^2$ . The normalizations on the difference range and also on FI were done to be able to have a plot independent of  $\delta$ .

In all tested cases<sup>6</sup> (under the symmetry assumptions), it was observed that when  $\Delta^*$  is used,  $\varepsilon = 0$  is the optimal solution. Thus, we can say that for commonly used noise models, if the quantization thresholds are well chosen,  $\tau_0 = x$  is optimal. The "commonly used" term here seems to be larger than in the binary case, as all the GGD with  $\beta > 2$  do not have anymore the asymmetric behavior for the optimal central threshold.

After setting  $\tau_0 = x$ , we still need to characterize the other thresholds to have a full performance characterization depending only on  $N_I$ . This can be equivalently stated as finding the variations from the central thresholds  $\tau'$  maximizing  $I_q(0)$  given by (1.46):

$$I_q^* = \operatorname{argmax}_{\tau'} I_q(0). \quad (1.47)$$

Unfortunately, an analytical solution cannot be found in general. An efficient solution for this problem could be obtained if this problem was convex or convexifiable [Boyd 2004], but this is not the case, so this is a very complicated multidimensional maximization problem. A possible solution to it is to fix the quantizer to be uniform, then in this case the problem is still one dimensional and it can be solved by exhaustive search (searching for the maximum on a fine grid of possible values). Existence of a non-degenerate solution ( $0 < \Delta^* < \infty$ ) is guaranteed by the following argument: for  $\Delta^* \rightarrow +\infty$ , all the distribution is concentrated on the first quantizer interval (remember that  $\varepsilon = 0$ ), thus  $I_q$  will be equal to the binary case  $I_q$  and for  $\Delta^* = 0$ , we get directly the binary quantization performance. As it was explained above,  $I_q$

<sup>6</sup>Two families of distributions were tested, the GGD and the Student-t distribution which will be presented later in Ch. 3. They were tested with uniform symmetric quantizers.

increases when we add thresholds, so at least one non-degenerate solution must exist. For a non-uniform solution, we can try to use local approximations by using Taylor series, this subject will be left to Part II.

### 1.4.3 Summary of the main points

Thus, up to this point we have:

- estimation performance based on quantized measurements is bounded above by the estimation performance based on continuous measurements.
- Adding quantization levels does not decrease estimation performance (it might increase in most of the cases).
- The optimal central threshold  $\tau_0$  must be placed at the true parameter  $x$  for commonly used noise models (Gaussian, Laplacian, Cauchy distributions). If we consider  $N_I > 2$ , symmetric thresholds w.r.t. the central threshold and well chosen quantization intervals, then it seems that  $\tau_0 = x$  may be optimal for a large class of symmetric unimodal distributions (for all the distributions above plus other members of the GGD).
- Maximizing the estimation performance w.r.t. the other thresholds (1.47) is in general a complicated problem.

### 1.4.4 MLE for multibit quantization with fixed thresholds

As it was done in the binary case, we still need to precise how to implement the MLE. Note that in this case the likelihood is given by (1.5)

$$L(x; i_{1:N}) = \prod_{k=1}^N \mathbb{P}(i_k; x).$$

Now, the MLE cannot be written in simple form and we must resort to numerical maximization. In general, we could use a steepest ascent algorithm, to iteratively climb the likelihood function. As it was developed in [Ribeiro 2006a], an efficient solution can be found when the noise distribution is log-concave. A log-concave distribution is a distribution for which its logarithm is concave, a simple example is the Gaussian distribution. If  $f$  is log-concave, it is known that  $\mathbb{P}(i_k; x)$  is log-concave [Boyd 2004, p. 107] and also that their product (expression above) is log-concave [Boyd 2004, p.105]. Thus, under this assumption the log of  $L$  is concave, an efficient solution for finding the MLE is the Newton's algorithm [Boyd 2004, p. 496] given by [Ribeiro 2006a]:

$$\hat{X}_{ML,j} = \hat{X}_{ML,j-1} - \frac{\frac{\partial \log \mathbb{P}(i_k; x)}{\partial x}}{\frac{\partial^2 \log \mathbb{P}(i_k; x)}{\partial x^2}} \Bigg|_{x=\hat{X}_{ML,j-1}}, \quad (1.48)$$

the subscript  $j$  is used to represent the iteration index and  $\Big|_{x=\hat{X}_{ML,j-1}}$  means that the function on its left is evaluated at the point  $x = \hat{X}_{ML,j-1}$ . After starting the algorithm with an

arbitrary  $\hat{X}_{ML,0}$ , the iterations are done until a pre-specified small minimum value  $\varepsilon_{\min}$  for the variations  $|\hat{X}_{ML,j} - \hat{X}_{ML,j-1}|$  is crossed. All the interest in obtaining a concave problem formulation comes from the fact that the Newton's algorithm not only guarantees convergence to a global maximum but also does it with quadratic convergence, i.e. when the iterates gets close to the optimal value, at each iteration  $\hat{X}_{ML,j}$  gets 2 digits closer to  $x$  [Boyd 2004, p. 489].

Therefore, for  $N_I > 2$ , with a fixed set of thresholds and considering that the distribution is log-concave we have the following solution for problem (a) (p. 27):

**Solution to (a) - MLE for quantized measurements with log-concave noise distribution,  $N_I > 2$  and fixed  $\tau$**

**(a1.2) 1) Estimator**

Define an initial guess on the estimate  $\hat{X}_{ML,0}$ . Until  $|\hat{X}_{ML,j} - \hat{X}_{ML,j-1}| < \varepsilon_{\min}$ , do

$$\hat{X}_{ML,j} = \hat{X}_{ML,j-1} - \frac{\frac{\partial \log \mathbb{P}(i_k; x)}{\partial x}}{\frac{\partial^2 \log \mathbb{P}(i_k; x)}{\partial x^2}} \Bigg|_{x=\hat{X}_{ML,j-1}}$$

and set  $j = j + 1$ . Then,  $\hat{X}_{ML,q}$  is set to the last  $\hat{X}_{ML,j}$ .

**2) Performance (asymptotic)**

$\hat{X}_{ML,q}$  is asymptotically unbiased

$$\mathbb{E} \left[ \hat{X}_{ML,q} \right]_{N \rightarrow \infty} = x$$

and its asymptotic MSE or variance is given by

$$\text{Var} \left( \hat{X}_{ML,q} \right)_{N \rightarrow \infty} \sim \text{CRB}_q = \frac{1}{NI_q},$$

with  $I_q$  given by (1.13).

## 1.5 Adaptive quantizers: the high complexity fusion center approach

The analysis and results above indicate that to get optimal estimation performance from quantized measurements we must, in general, place the central threshold close to the true parameter<sup>7</sup>. This can be done by using the information given by the measurements to move adaptively the central threshold. Main work that has been already done on this subject will be presented in this section.

An adaptive scheme to estimate  $x$  based on a sensor network of binary quantizers is presented in [Li 2007]. The main idea is that enhanced estimation performance can be obtained if the sensors can place dynamically their thresholds around  $x$ . Here, we present an equivalent sequential version using only one sensor. The following is proposed:

1. a sensor can communicate binary measurements to a fusion center. The sensor measurement noise sequence is supposed to be i.i.d..
2. The sensor starts with a known binary threshold  $\tau_{0,0}$ , where the second subscript is for the discrete-time index. Note that now the threshold will be considered to be varying.
3. At each instant  $k$ , the sensor obtains a binary quantized measurement  $i_k$  ( $i_k \in \{-1, 1\}$ ).
4. The sensor then updates the threshold by the following simple cumulative rule:

$$\tau_{0,k} = \tau_{0,k-1} + \gamma i_k, \quad (1.49)$$

where  $\gamma$  is a constant positive adaptation step (see the remarks after the MLE definition).

5. The sensor sends its measurement  $i_k$  to the fusion center.
6. The fusion center updates its  $\tau_{0,k}$ , and stocks in a memory both  $i_k$  and  $\tau_{0,k}$ . Note that the fusion center threshold is exactly the same as the one obtained in the sensor threshold update.
7. After a predefined number of iterations, for example  $N$ , or at each iteration  $k$ , the fusion center can get a more precise estimate of  $x$  (more precise than  $\tau_{0,k}$ ) by using a MLE based on all past  $i_k$ .

---

<sup>7</sup>The literature on the subject also points in the same direction.

The case when  $x$  is constrained to lie in a bounded interval  $\mathcal{X}$  of  $\mathbb{R}$  was extensively studied in [Papadopoulos 2001]. Main attention was given to the effects of different schemes for setting  $\tau_0$ . The schemes considered were: fixed, varying but random and i.i.d, varying deterministically and based on feedback. For each scheme, the worst case  $\text{CRB}_q$  ( $x$  was chosen to maximize the CRB) was evaluated and divided by the continuous measurement CRB to give a measure on the performance loss induced by quantization. The loss was shown to be more sensible w.r.t. an equivalent signal-to-noise ratio (the interval  $\mathcal{X}$  length divided by the noise scale factor) in the fixed case and insensible in the feedback case. Some solutions based on iterative maximum likelihood techniques, which puts the new threshold on the last ML estimate, were presented but no theoretical proofs that they reach the minimum  $\text{CRB}_q$  were given.

In [Ribeiro 2006a], where the binary quantization Gaussian noise case was mainly studied, it was pointed out that the sensibility of the estimation performance to  $\varepsilon$  and its optimality for  $\varepsilon = 0$  indicates that, to enhance performance, we could move adaptively the binary threshold, placing it on the last available estimate  $\hat{X}$  to get closer and closer to the true  $x$ .

### 1.5.1 MLE for the adaptive binary scheme

As the threshold is dependent on the measurements, the measurements are not independent anymore. However, as a measurement  $i_k$  is dependent only on past measurements and this dependence is done through  $\tau_{0,k-1}$ , conditioned on the threshold that was used, the measurements are independent. This leads to the following likelihood and log-likelihood for the measurements until time  $N$ :

$$\begin{aligned} L(x; i_{1:N}) &= \mathbb{P}(i_{1:N}; x) = \prod_{k=1}^N \mathbb{P}(i_k | i_{k-1}, \dots, i_1; x) \\ &= \prod_{k=1}^N \mathbb{P}(i_k | \tau_{0,k-1}; x) \\ &= \prod_{k=1}^N [1 - F(\tau_{0,k-1} - x)]^{\frac{1+i_k}{2}} F(\tau_{0,k-1} - x)^{\frac{1-i_k}{2}}, \end{aligned} \quad (1.50)$$

$$\log L(x; i_{1:N}) = \sum_{k=1}^N \left\{ \frac{1+i_k}{2} \log [1 - F(\tau_{0,k-1} - x)] + \frac{1-i_k}{2} \log F(\tau_{0,k-1} - x) \right\}, \quad (1.51)$$

where the vertical bar inside the probability symbol means that the probability measure is evaluated for the r.v. on the left side of the bar, conditioned on the r.v. on the right side of the bar. The conditioning makes the output  $i_k$  depend on  $\tau_{0,k-1}$  as if it was a deterministic parameter, that is why we can use the same notation with CDF  $F$  parametrized by a fixed nonrandom threshold.

At the fusion center at time  $N$ , all the thresholds and binary measurements are known, the maximum likelihood estimator can then be calculated by maximizing (1.50) or (1.51):

$$\hat{X}_{ML,q} = \underset{x}{\operatorname{argmax}} \prod_{k=1}^N [1 - F(\tau_{0,k-1} - x)]^{\frac{1+i_k}{2}} F(\tau_{0,k-1} - x)^{\frac{1-i_k}{2}} \quad (1.52)$$

or

$$\hat{X}_{ML,q} = \underset{x}{\operatorname{argmax}} \sum_{k=1}^N \left\{ \frac{1+i_k}{2} \log [1 - F(\tau_{0,k-1} - x)] + \frac{1-i_k}{2} \log F(\tau_{0,k-1} - x) \right\}.$$

Note that the threshold moves with each measurement, while the estimate is obtained only at the end of the measurement block. Observe also that when the noise distributions are log-concave, the MLE can also be obtained by using the Newton's algorithm, as it was discussed in the previous section.

**Remarks:** it is intuitive to expect that the mean  $\tau_{0,k}$  will reach an equilibrium after some time. If the threshold is above the parameter, iteration (1.49) will reduce its value, in the other case, if the threshold is below the parameter, iteration (1.49) will increase its value. In the mean equilibrium we have  $\mathbb{E} \left[ \frac{\tau_{0,k} - \tau_{0,k-1}}{\gamma} \right] = \mathbb{E}[i_k] = 0$ , thus as  $i_k = 1$  or  $i_k = -1$  the only possibility for this to happen is when  $\mathbb{P}(i_k = 1; x) = \mathbb{P}(i_k = -1; x) = \frac{1}{2}$ , which in the case of symmetric noise distributions means to say that  $\mathbb{E}[\tau_{0,k}] = x$ .

The variance of the thresholds will depend on the noise distribution but also on the parameter  $\gamma$ , if we choose  $\gamma$  to be relatively small, once the threshold is close to the parameter, it will fluctuate around it with a small variance. The fact that the threshold updates are easy to implement (it is just a cumulative sum) and that the estimator is a complex one goes well with real implementation constraints, where complexity on the sensor side of the problem is strongly constrained and on the fusion center side, it is less constrained.

### 1.5.2 Performance for the adaptive binary scheme

We must look now to the performance of this scheme. The performance analysis that is presented here was proposed in [Fang 2008].

Even if the measurements are dependent, it is known that, under some conditions that are satisfied here, the MLE will still attain the CRB [Crowder 1976]. Thus, the main problem here will be the evaluation of the FI. As the measurements are dependent, the FI for  $N$  measurements is not  $N$  times the FI for one measurement and we need to evaluate it using the score function for the entire block of measurements. The FI for  $N$  measurements is

$$I_{q,1:N} = \mathbb{E}[S_q] = \mathbb{E} \left\{ \left[ \sum_{k=1}^N \frac{\partial \log \mathbb{P}(i_k | \tau_{0,k-1}; x)}{\partial x} \right]^2 \right\}.$$

It was shown in [Li 2007] that this quantity is equal to

$$I_{q,1:N} = \sum_{k=1}^N \mathbb{E} \left[ \frac{f^2(\tau_{0,k-1} - x)}{F(\tau_{0,k-1} - x)[1 - F(\tau_{0,k-1} - x)]} \right], \quad (1.53)$$

where the expectation is evaluated under the only r.v. that still appears on the expression,  $\tau_{0,k-1}$ . If we assume that  $\tau_{0,0} = 0$  then the  $\tau_{0,k-1}$  is a random walk in an infinite grid (more specifically finite for finite  $k$ ) with values  $\{-\infty, \dots, -2\gamma, -\gamma, 0, \gamma, 2\gamma, \dots, +\infty\}$ .

For understanding how (1.53) was obtained, one can decompose the squared sum of score functions into a sum of squared scores and a sum of score cross products. As the measurements are independent, the expectation of each cross product will be the product of expectations. The product of expectations will be zero because the expectation of a score function is zero [Kay 1993, p. 67]. Therefore,  $I_{q,1:N}$  will be the expectation of the sum of squared scores. Decomposing the expectation into an expectation on  $i_k$  conditioned on the thresholds and an expectation on the thresholds, one gets (1.53).

Denoting the probability of having  $\tau_{0,k-1} = j\gamma$  by  $\mathbb{P}(\tau_{0,k-1} = j\gamma) = p_{j,k-1}$  we have:

$$I_{q,1:N} = \sum_{k=1}^N \sum_{j=-\infty}^{+\infty} \frac{f^2(j\gamma - x)}{F(j\gamma - x)[1 - F(j\gamma - x)]} p_{j,k-1}. \quad (1.54)$$

Note that this is equivalent to obtaining  $N$  measurements from a binary quantizer with a random thresholding scheme that changes its prior threshold distribution  $p_{k-1}$  in time. The prior distribution changes in a way that when  $k \rightarrow \infty$ , it is expected that most of its probability will be concentrated around the parameter. This is in contrast to the methods presented in



To solve this infinite dimensional system [Fang 2008] considered that only a part of the thresholds around the true parameter will have a non negligible probability, for practical purposes it was considered that non negligible thresholds are those in the interval

$$\mathcal{I}_\tau = [-5\sigma_v - |x|, 5\sigma_v + |x|],$$

where  $\sigma_v$  is the standard deviation of the noise<sup>9</sup>. The non negligible probability vector, denoted  $\tilde{\mathbf{p}}_\infty$  will have size  $2 \left\lceil \frac{5\sigma_v + |x|}{\gamma} \right\rceil + 1 = 2j_{\max} + 1$ , where  $\lceil y \rceil$  is the closest integer that is larger than  $y$  and the "+1" comes from the zero threshold. The approximate threshold distribution can then be obtained by solving

$$\tilde{\mathbf{p}}_\infty = \begin{bmatrix} \tilde{p}_{-j_{\max}, \infty} \\ \vdots \\ \tilde{p}_{\infty, 0} \\ \vdots \\ \tilde{p}_{j_{\max}, \infty} \end{bmatrix} = \tilde{\mathbf{T}} \tilde{\mathbf{p}}_\infty, \quad (1.60)$$

where  $\tilde{\mathbf{T}}$  is the truncated transition matrix around the zero threshold (we show only the upper left corner)

$$\tilde{\mathbf{T}} = \begin{bmatrix} 0 & F(-\gamma(j_{\max} - 1) + \varepsilon) & 0 & 0 & 0 \\ 1 - F(-\gamma j_{\max} + \varepsilon) & 0 & F(-\gamma(j_{\max} - 2) + \varepsilon) & 0 & 0 \\ & \ddots & \ddots & \ddots & \ddots \\ 0 & & & & \end{bmatrix}.$$

One can also truncate  $\mathbf{I}'_q$  only for the non negligible probability elements

$$\tilde{\mathbf{I}}'_q = \left[ \frac{f^2(-\gamma j_{\max} + \varepsilon)}{F(-\gamma j_{\max} + \varepsilon)[1 - F(-\gamma j_{\max} + \varepsilon)]}, \dots, \frac{f^2(\gamma j_{\max} + \varepsilon)}{F(\gamma j_{\max} + \varepsilon)[1 - F(\gamma j_{\max} + \varepsilon)]} \right]^\top. \quad (1.61)$$

Following the development in [Fang 2008], after a finite time  $N_c$  the probability vector  $\mathbf{p}_k$  will be indistinguishable from  $\mathbf{p}_\infty$ , thus when  $N \rightarrow \infty$ , an infinity number of terms in  $I_{q,1:N}$  will behave approximately as  $\tilde{\mathbf{I}}'^\top_q \tilde{\mathbf{p}}_\infty$ , which leads to the following asymptotic approximation of the FI:

$$I_{q,1:N} = \sum_{k=1}^N \mathbf{I}'^\top_q \mathbf{p}_{k-1} \underset{N \rightarrow \infty}{\sim} N \tilde{\mathbf{I}}'^\top_q \tilde{\mathbf{p}}_\infty. \quad (1.62)$$

<sup>9</sup>For one of the noise distributions considered here, the Cauchy distribution, the standard deviation is undefined. In this case, one can use the scale parameter  $\delta$  instead of the standard deviation  $\sigma_v$ .

This gives the following solution for problem (a) (p. 27):

**Solution to (a) - MLE for binary quantized measurements with adaptive thresholds given by a simple cumulative sum.**

**(a2.1) 1) Estimator**

Define an initial threshold  $\tau_{0,0}$  and a positive  $\gamma$ , then from  $k = 1$  to  $N$ :

- the sensor obtains a binary measurement  $i_k$  using  $\tau_{0,k-1}$ .
- The sensor sends  $i_k$  to the fusion center and updates the threshold (1.49):

$$\tau_{0,k} = \tau_{0,k-1} + \gamma i_k.$$

- The fusion center stores  $i_k$  and also evaluates and stores  $\tau_{0,k}$ .

With  $i_{1,N}$  and  $\tau_{0,1:N}$ , the fusion center evaluates the MLE (1.52):

$$\hat{X}_{ML,q} = \underset{x}{\operatorname{argmax}} \prod_{k=1}^N [1 - F(\tau_{0,k-1} - x)]^{\frac{1+i_k}{2}} F(\tau_{0,k-1} - x)^{\frac{1-i_k}{2}}.$$

**2) Performance (asymptotic and approximate)**

$\hat{X}_{ML,q}$  is asymptotically unbiased

$$\mathbb{E} \left[ \hat{X}_{ML,q} \right] \underset{N \rightarrow \infty}{=} x$$

and its asymptotic MSE or variance can be approximated by

$$\operatorname{Var} \left( \hat{X}_{ML,q} \right) \underset{N \rightarrow \infty}{\sim} \operatorname{CRB}_q \approx \frac{1}{N \tilde{\mathbf{I}}_q^\top \tilde{\mathbf{p}}_\infty},$$

with  $\tilde{\mathbf{I}}_q$  given by (1.61) and  $\tilde{\mathbf{p}}_\infty$  by (1.60).

An alternative to have analytical results on the vector  $\mathbf{p}_\infty$  without using an approximation with truncation can be obtained by considering that  $x$  lies in a symmetric interval  $[-A, A]$ , where  $A$  is a positive real. We can create boundaries on the possible values of the threshold in such a way that the number of possible thresholds is finite. In this way, the threshold sequence can be modeled as a Markov chain defined in a domain with a finite number of values and we can evaluate the asymptotic threshold distribution without using truncation approximations (More? - App. A.2.4).

### 1.5.3 Adaptive scheme based on the MLE

One of the disadvantages of (a2.1) is that the threshold will fluctuate around  $x$  and it will not converge to  $x$ , producing a performance that is still not optimal. A remedy for this problem was proposed also in [Fang 2008] (and previously in [Papadopoulos 2001]). By accepting a feedback from the fusion center and assuming that the fusion center has enough processing power to evaluate the MLE for the past measurements at each time, instead of using the cumulative sum for updating the threshold, we can use the last MLE estimate. Intuitively, with a growing number of measurements for the MLE, the threshold will be placed closer and closer to  $x$ , producing as a result an MLE with performance approaching the optimal one (for  $\tau_{0,k-1} = x$ ).

The new update is given by

$$\tau_{0,k} = \hat{X}_{ML,k}, \quad (1.63)$$

where  $\hat{X}_{ML,k}$  is the MLE for the measurements  $i_{1:k}$ . The asymptotic performance analysis was also presented in [Fang 2008], the authors claim that in the binary quantization and Gaussian case, the performance (variance) is asymptotically given by  $\frac{\pi\delta^2}{4N}$ . Therefore, this update scheme is asymptotically optimal as  $I_q(0) = \frac{4N}{\pi\delta^2}$  is the maximum FI that can be achieved.

We will mimic some parts of their proof, but we will change some arguments to obtain a more general result for  $N_I \geq 2$ .

### 1.5.4 Performance for the adaptive multibit scheme based on the MLE

Under an adaptive  $\tau_0$  with the vector  $\boldsymbol{\tau}'$  fixed, we can rewrite the FI given in (1.53) for a general  $N_I$  using a parametrization on the error  $\varepsilon_k = \tau_{0,k-1} - x$ , which now depends on time and it is given as follows (Why? - App. A.1.5)

$$I_{q,1:N}^{N_I} = \sum_{k=1}^N \mathbb{E} [I_q(\varepsilon_k)], \quad (1.64)$$

where  $\varepsilon_k$  is a sequence of r.v. defined on  $\mathbb{R}$ , contrary to the previous case when the thresholds were defined in a grid. The function  $I_q(\varepsilon_k)$  is given by (1.13).

For proceeding, we will make additional assumptions on  $I_q(\varepsilon)$  (the assumptions on the noise AN1 p. 34 and AN2 p. 34 are also assumed).

**Assumptions on  $I_q$  for the MLE update to have asymptotically optimal performance:**

**A1.MLE**  $I_q(\varepsilon)$  is maximum for  $\varepsilon = 0$ .

**A2.MLE**  $I_q(\varepsilon)$  is locally decreasing around zero.

**A3.MLE** The function  $I_q(\varepsilon)$  has bounded  $I_q(0)$ ,  $\left. \frac{dI_q(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = 0$ , bounded  $\left. \frac{d^2 I_q(\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0}$ , therefore accepting a Taylor approximation around zero (for small  $\varepsilon'$ ):

$$I_q(\varepsilon') = I_q(0) + \frac{\varepsilon'^2}{2} \left. \frac{d^2 I_q(\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0} + o(\varepsilon'^2), \quad (1.65)$$

where the  $o(\varepsilon'^2)$  here is equivalent to say that the quantity  $\frac{o(\varepsilon'^2)}{\varepsilon'^2}$  tends to zero when  $\varepsilon'$  tends to zero.

If we look to Fig. 1.8 (p. 57), we can see that these assumptions seem to be satisfied by Gaussian and Cauchy distributions. Except for the Laplacian-like distributions with a derivative discontinuity at  $\varepsilon = 0$ , a large class of smooth symmetric unimodal distributions satisfy these assumptions for  $N_I > 2$  and well chosen quantizer intervals. Note that in the binary cases, where the threshold must be placed asymmetrically, we can add a fixed bias in the MLE threshold update to obtain a better performance. Also for the asymmetric cases, all the assumptions can be stated around the maximum point for the FI instead of  $\varepsilon = 0$ .

The objective now will be to bound above and below the quantity  $I_{q,1:N}^{N_I}$  in such a way, that when we make  $N \rightarrow \infty$  both bounds will "squeeze"  $I_{q,1:N}^{N_I}$  on an interval that goes asymptotically to  $NI_q(0)$ .

For a large number of measurements  $M < N$ , the MLE studied here is consistent even if the measurements are dependent, for verifying this, one can check the regularity conditions given in [Crowder 1976]. Thus, for  $\varepsilon' > 0$  and  $\xi > 0$ , it is possible to choose a number of measurements  $M$  such

$$\mathbb{P}(|\varepsilon_k| \leq \varepsilon') \geq 1 - \xi, \text{ for } k \geq M. \quad (1.66)$$

Applying this inequality with the monotonicity property of A2.MLE, we can say that we can find a  $M$  such

$$\mathbb{P}(I_q(\varepsilon_k) \geq I_q(\varepsilon')) \geq 1 - \xi, \text{ for } k \geq M. \quad (1.67)$$

Now the sum in (1.64) can be separated in two sums, one for the terms with  $k < M$ ,  $I_{q,1:M-1}$  and the other with  $k \geq M$ ,  $I_{q,M:N}$ :

$$I_{q,1:N}^{N_I} = I_{q,1:M-1} + I_{q,M:N} = \left\{ \sum_{k=1}^{M-1} \mathbb{E}[I_q(\varepsilon_k)] \right\} + \left\{ \sum_{k=M}^N \mathbb{E}[I_q(\varepsilon_k)] \right\}. \quad (1.68)$$

Using A1.MLE and the fact that  $I_q(\varepsilon_k)$  is a nonnegative quantity, we know that  $I_q(\varepsilon_k) \in [0, I_q(0)]$ . Thus, the first term can be written as:

$$I_{q,1:M-1} = \alpha_M (M - 1) I_q(0), \quad (1.69)$$

with  $\alpha_M \in [0, 1]$ . The terms on  $I_{q,M:N}$  can be lower bounded using the Markov's inequality. The Markov's inequality states that for a nonnegative r.v.  $Y$  and value  $y > 0$ , we must have [Wasserman 2003, p. 63]:

$$\mathbb{P}(Y > y) \leq \frac{\mathbb{E}(Y)}{y}.$$

Using this inequality for an arbitrary term of  $I_{q,M:N}$  with the value  $I_q(\varepsilon')$  gives

$$\mathbb{E}[I_q(\varepsilon_k)] \geq I_q(\varepsilon') \mathbb{P}[I_q(\varepsilon_k) \geq I_q(\varepsilon')], \text{ for } k \geq M. \quad (1.70)$$

Then, supposing that the thresholds are updated using the MLE, we can use (1.67) in (1.70) to have

$$\mathbb{E}[I_q(\varepsilon_k)] \geq I_q(\varepsilon') (1 - \xi), \text{ for } k \geq M. \quad (1.71)$$

For sufficiently large  $M$  (and consequently  $N$ ),  $I_{q,1:N}^{N_I}$  can be lower bounded using (1.71) and (1.69)

$$I_{q,1:N}^{N_I} \geq \alpha_M (M - 1) I_q(0) + [N - (M - 1)] I_q(\varepsilon') (1 - \xi). \quad (1.72)$$

From A1.MLE the FI can be upper bounded by the optimal  $I_q$

$$I_{q,1:N}^{N_I} \leq N I_q(0). \quad (1.73)$$

Joining (1.72) and (1.73) gives the following:

$$\alpha_M (M - 1) I_q(0) + [N - (M - 1)] I_q(\varepsilon') (1 - \xi) \leq I_{q,1:N}^{N_I} \leq N I_q(0). \quad (1.74)$$

For small  $\varepsilon'$ , we can use A3.MLE to obtain

$$\alpha_M (M - 1) I_q(0) + [N - (M - 1)] \left[ I_q(0) + \frac{\varepsilon'^2}{2} \left. \frac{d^2 I_q(\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0} + o(\varepsilon'^2) \right] (1 - \xi) \leq I_{q,1:N}^{N_I} \leq N I_q(0). \quad (1.75)$$

The term on the left of the inequality can be rewritten as

$$N I_q(0) \left\{ (1 - \xi) + \frac{(M - 1) [\alpha_M - (1 - \xi)]}{N} \right\} + [N - (M - 1)] \left[ \frac{\varepsilon'^2}{2} \left. \frac{d^2 I_q(\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0} + o(\varepsilon'^2) \right] (1 - \xi).$$

Separating a factor  $N I_q(0)$  we can write the term above as  $N I_q(0) (1 - \xi')$  with

$$\xi' = \left\{ -\xi + \frac{(M - 1) [\alpha_M - (1 - \xi)]}{N} \right\} + \left[ 1 - \frac{(M - 1)}{N} \right] \left[ \frac{\varepsilon'^2}{2} \left. \frac{d^2 I_q(\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0} + o(\varepsilon'^2) \right] \frac{(1 - \xi)}{I_q(0)}. \quad (1.76)$$

Therefore, the inequality (1.74) becomes

$$N I_q(0) (1 - \xi') \leq I_{q,1:N}^{N_I} \leq N I_q(0).$$

By imposing  $N \gg M$  ( $N$  much larger than  $M$ ) so that  $\frac{(M-1)}{N}$  is arbitrary small and by choosing  $M$  sufficiently large so that  $\xi$  is small, we can make the first term in  $\xi'$  to approach zero. Using also  $N \gg M$  and choosing now  $M$  sufficiently large so that  $\varepsilon'$  is arbitrary small, we can make the second term in  $\xi'$  to approach zero. Therefore, we can make the left side of

the inequality above to be close to  $NI_q(0)$  when  $N$  and  $M$  tend to infinity with the condition  $N \gg M$ . As the upper bound on  $I_{q,1:N}^{N_I}$  is also  $NI_q(0)$ , we have that

$$I_{q,1:N}^{N_I} \underset{N \rightarrow \infty}{\sim} NI_q(0), \quad (1.77)$$

or equivalently

$$\text{CRB}_q \underset{N \rightarrow \infty}{\sim} \frac{1}{NI_q(0)}. \quad (1.78)$$

We have now the following solution for problem (a)<sup>10</sup>:

**Solution to (a) - MLE for quantized measurements with  $N_I \geq 2$  and adaptive thresholds given by the MLE.**

**(a2.2) 1) Estimator**

Define an initial threshold  $\tau_{0,0}$ , then from  $k = 1$  to  $N$ :

- the sensor obtains a binary measurement  $i_k$  using  $\tau_{0,k-1}$ .
- The sensor sends  $i_k$  to the fusion center.
- The fusion center stores  $i_k$ , evaluates and stores  $\hat{X}_{ML,k} = \tau_{0,k}$  following (1.63)

$$\tau_{0,k} = \hat{X}_{ML,k},$$

where the estimate  $\hat{X}_{ML,k}$  is given by

$$\hat{X}_{ML,k} = \underset{x}{\operatorname{argmax}} \prod_{j=1}^k \mathbb{P}(i_j; x, \tau_{0,j-1}).$$

- The fusion center sends  $\tau_{0,k} = \hat{X}_{ML,k}$  to the sensor.

**2) Performance (asymptotic)**

$\hat{X}_{ML,q} = \hat{X}_{ML,k=N}$  is asymptotically unbiased

$$\mathbb{E} \left[ \hat{X}_{ML,q} \right] \underset{N \rightarrow \infty}{=} x$$

and its asymptotic MSE or variance attains the optimal value

$$\operatorname{Var} \left( \hat{X}_{ML,q} \right) \underset{N \rightarrow \infty}{\sim} \frac{1}{NI_q(0)}.$$

Now, we have an estimator with adaptive thresholds (mainly the central threshold) that attains the asymptotically optimal performance. The estimator guides the quantizer dynamic

<sup>10</sup>The threshold  $\tau_{0,k-1}$  is added in the notation of the probabilities to make the dependence on time more explicit.

range close to the parameter by setting the central point of the quantizer with a decreasing fluctuation around  $\tau$ .

### 1.5.5 Equivalent low complexity asymptotic scheme

The main disadvantage of (a2.2) is its high complexity, since the MLE must be obtained at each iteration. In [Papadopoulos 2001] a heuristic based on an approximation of the expectation maximization method for applying the MLE update with reduced complexity on the binary quantization and Gaussian noise case was presented. The proposed threshold/estimate update is given by the following recursive expression:

$$\hat{X}_k = \tau_{0,k} = \hat{X}_{k-1} + \frac{\delta\sqrt{\pi}}{2k} i_k. \quad (1.79)$$

Observe that the difference in complexity is large. In general, the MLE must be obtained with a maximization algorithm, *e.g.* Newton's algorithm, which itself has an inner recursive procedure that may need multiple iterations for reaching convergence for each time  $k$ . In (1.79), we have only a recursive procedure in  $k$ , which requires a multiplication of  $i_k$  by a gain and summation with the last estimate.

We can show that (1.79) can be generalized easily to non Gaussian noise cases. We will use a less heuristic method (less than the method used to obtain (1.79)). We will assume, additionally to symmetry, that the noise PDF has  $f^{(1)}(0) = 0$ . If we consider that  $k$  is large, then from the convergence of the CRB discussed above and the asymptotic normality of the MLE [Kay 1993, p. 167] (or [Crowder 1976]), the error between the threshold used to obtain  $i_k$ ,  $\varepsilon = \hat{X}_{ML,k-1} - x = \tau_{0,k} - x$ , is Gaussian distributed with zero mean and variance  $\frac{1}{(k-1)I_q(0)}$ <sup>11</sup>:

$$f_\varepsilon(\varepsilon) = \sqrt{\frac{(k-1)I_q(0)}{2\pi}} \exp\left[-\frac{(k-1)I_q(0)}{2}\varepsilon^2\right], \quad (1.80)$$

where  $f_\varepsilon$  is the PDF of the error. We can try to estimate the random error using the new quantized observation  $i_k$  and the knowledge about its distribution given by the PDF above. After estimating it, we can correct  $\hat{X}_{ML,k-1}$  using the estimate. As  $\varepsilon$  is random, we will use an estimator equivalent to the MLE, but for random parameters. In this case the **maximum a posteriori estimator (MAP)** will be used. The posterior distribution (the one that might be maximized) is the conditional PDF of  $\varepsilon$  given  $i_k$ . Using Bayes theorem, it is given by

$$p(\varepsilon|i_k) = \frac{\mathbb{P}(i_k|\varepsilon)f_\varepsilon(\varepsilon)}{\mathbb{P}(i_k)}, \quad (1.81)$$

where in the binary case the conditional probability  $\mathbb{P}(i_k|\varepsilon)$  is given by

$$\mathbb{P}(i_k|\varepsilon) = [1 - F(\varepsilon)]^{\frac{1+i_k}{2}} F(\varepsilon)^{\frac{1-i_k}{2}}. \quad (1.82)$$

The denominator is the marginal probability of the output  $i_k$  and it does not depend on  $\varepsilon$ . The MAP is then given by [Kay 1993, p. 350]

$$\hat{\varepsilon}_{MAP} = \underset{\varepsilon}{\operatorname{argmax}} p(\varepsilon|i_k). \quad (1.83)$$

<sup>11</sup>Observe that here we are using the parametrization of the Gaussian distribution with its variance and not with its scale parameter

In the same way as for the MLE, we can maximize the logarithm of the posterior, as  $\mathbb{P}(i_k)$  does not depend on  $\varepsilon$ , we can write an equivalent form for (1.83) as

$$\begin{aligned}\hat{\varepsilon}_{MAP} &= \operatorname{argmax}_{\varepsilon} \log p(\varepsilon|i_k) \\ &= \operatorname{argmax}_{\varepsilon} \{\log [\mathbb{P}(i_k|\varepsilon)] + \log [f_{\varepsilon}(\varepsilon)]\}.\end{aligned}\quad (1.84)$$

Using (1.81) and (1.80) in the RHS (1.84), we obtain

$$\log p(\varepsilon|i_k) = \left(\frac{1+i_k}{2}\right) \log [1-F(\varepsilon)] + \left(\frac{1-i_k}{2}\right) \log [F(\varepsilon)] - \frac{(k-1)I_q(0)}{2}\varepsilon^2. \quad (1.85)$$

Under consistency of the MLE, it is expected that for large  $k$ , the probability of  $|\varepsilon|$  being small is close to 1. Thus, we can look for a maximum point of (1.85) around zero. This leads us to expand  $\log [1-F(\varepsilon)]$  and  $\log [F(\varepsilon)]$  around zero. The expansions are given by

$$\begin{aligned}\log [1-F(\varepsilon)] &= \log [1-F(0)] + \varepsilon \left. \frac{d \log [1-F(z)]}{dz} \right|_{z=0} + \frac{\varepsilon^2}{2} \left. \frac{d^2 \log [1-F(z)]}{dz^2} \right|_{z=0} + o(\varepsilon^2), \\ \log [F(\varepsilon)] &= \log [F(0)] + \varepsilon \left. \frac{d \log [F(z)]}{dz} \right|_{z=0} + \frac{\varepsilon^2}{2} \left. \frac{d^2 \log [F(z)]}{dz^2} \right|_{z=0} + o(\varepsilon^2).\end{aligned}$$

Using the symmetry of the distribution ( $1-F(0) = F(0) = \frac{1}{2}$ ) the terms with logarithms are  $\log(\frac{1}{2}) = -\log(2)$ . The derivatives at the zero point are

$$\begin{aligned}\left. \frac{d \log [1-F(z)]}{dz} \right|_{z=0} &= -\frac{f(0)}{1-F(0)} = -2f(0), \\ \left. \frac{d \log [F(z)]}{dz} \right|_{z=0} &= \frac{f(0)}{F(0)} = 2f(0)\end{aligned}$$

and using the assumption  $f^{(1)}(0) = 0$ , the second derivatives are

$$\begin{aligned}\left. \frac{d^2 \log [1-F(z)]}{dz^2} \right|_{z=0} &= \frac{-f^{(1)}(0)}{1-F(0)} - \frac{f^2(0)}{[1-F(0)]^2} = -4f^2(0), \\ \left. \frac{d^2 \log [F(z)]}{dz^2} \right|_{z=0} &= \frac{f^{(1)}(0)}{F(0)} - \frac{f^2(0)}{F^2(0)} = -4f^2(0).\end{aligned}$$

Applying these expressions to the expansions above, we get

$$\begin{aligned}\log [1-F(\varepsilon)] &= -\log(2) - 2\varepsilon f(0) - 4\frac{\varepsilon^2}{2}f^2(0) + o(\varepsilon^2), \\ \log [F(\varepsilon)] &= -\log(2) + 2\varepsilon f(0) - 4\frac{\varepsilon^2}{2}f^2(0) + o(\varepsilon^2).\end{aligned}$$

These expansions can be used in (1.85), this gives the following:

$$\begin{aligned}\log p(\varepsilon|i_k) &= \left(\frac{1+i_k}{2}\right) \left\{ -\log(2) - 2\varepsilon f(0) - 4\frac{\varepsilon^2}{2}f^2(0) + o(\varepsilon^2) \right\} + \\ &+ \left(\frac{1-i_k}{2}\right) \left\{ -\log(2) + 2\varepsilon f(0) - 4\frac{\varepsilon^2}{2}f^2(0) + o(\varepsilon^2) \right\} + \\ &- \frac{(k-1)I_q(0)}{2}\varepsilon^2.\end{aligned}\quad (1.86)$$

To find the maximum, we differentiate  $\log p(\varepsilon|i_k)$  in (1.86) w.r.t.  $\varepsilon$  and we equate it to zero. This gives

$$\begin{aligned} (k-1)I_q(0)\varepsilon &= \left(\frac{1+i_k}{2}\right)\{-2f(0) - 4\varepsilon f^2(0) + o(\varepsilon)\} + \\ &+ \left(\frac{1-i_k}{2}\right)\{2f(0) - 4\varepsilon f^2(0) + o(\varepsilon)\} \\ &= -2f(0)i_k - 4\varepsilon f^2(0) + o(\varepsilon). \end{aligned}$$

For binary measurements, we know that  $I_q(0) = 4f^2(0)$ . Thus adding  $\varepsilon 4f^2(0)$  on both sides gives

$$k4f^2(0)\varepsilon = -2f(0)i_k + o(\varepsilon).$$

Thus, we have

$$\hat{\varepsilon}_{MAP} \underset{k \rightarrow \infty}{\sim} -\frac{i_k}{2f(0)k}.$$

The optimal new threshold/estimate when  $k \rightarrow \infty$  is then given by

$$\hat{X}_k = \hat{X}_{k-1} - \hat{\varepsilon}_{MAP} \approx \hat{X}_{k-1} + \frac{i_k}{2kf(0)}. \quad (1.87)$$

This is exactly the same recursive estimator obtained by [Papadopoulos 2001] when the noise is Gaussian ( $f(0) = \frac{1}{\sqrt{\pi\delta}}$  for the Gaussian distribution). Note that this recursive update/estimation procedure is asymptotically equivalent to the MLE update, as both procedures (MLE and MAP) have equivalent error distribution for  $k \rightarrow \infty$  [Wasserman 2003, p. 181].

Clearly, some questions arise about the low complexity recursive estimator above:

- can (1.87) converge if we use it when the initial distance  $|\varepsilon| = |\tau_0 - x|$  is arbitrary (not necessarily small)?
- Can we extend this low complexity recursive procedure to the  $N_I > 2$  case?

Answers for these questions will be given in Ch. 3.

## 1.6 Chapter summary and directions

We conclude this chapter with the main points observed until now and directions for future work.

- Estimation performance in terms of MSE can be minimized asymptotically under an unbiasedness constraint by the MLE (a1). The asymptotic performance is then mainly characterized by the CRB which is given in terms of the FI.
- The FI for quantized measurements is upper bounded by the FI for continuous measurements and lower bounded by the FI for binary quantization. Moreover, it increases as additional quantization intervals are used.
- The CRB and FI are very sensitive to the central threshold of the quantizer.
  - For commonly used noise models (Gaussian, Laplacian and Cauchy), the threshold must be placed exactly at the parameter.
  - In the binary quantization case, even if we restrict the noise distribution to be symmetric and unimodal this is not always true. We can find cases (GGD) where quantizing the input r.v. asymmetrically can be optimal. In these cases, it was also observed that the gain of performance obtained by using an asymmetric quantizer seems to be dependent on the noise distribution, however, in general, the gain from using the optimal asymmetric quantizer in the place of a symmetric quantizer seems to be small when compared with the gain that can be obtained by using a symmetric quantizer in the place of a poorly chosen asymmetric quantizer.
  - An interesting subject for future research is to study in more detail the effect of the noise distribution on the shape of the performance function  $B(\varepsilon)$  in the asymmetric cases, for example, we can try to characterize the loss incurred by imposing symmetric quantization w.r.t. optimal quantization. Another possible point for future research is to see if such asymmetric behavior also appears in the problem of detection using binary quantized measurements.
  - In all cases, under symmetry assumptions on the noise and on the quantizer, estimation performance degrades when the quantizer dynamic (the quantizer threshold in the binary case) is very distant to the true parameter.
  - For multibit quantization, also under symmetry assumptions, it seems that if we choose the quantizer thresholds (or equivalently the quantizer intervals) well, then for a large class of unimodal distributions it is optimal to place the central threshold at the true parameter. Note that, quantizing "well" in this case means that we choose the quantization intervals to have a good symmetric quantization performance. An interesting point for future analysis is to see if we can get a better performance than in the symmetric case, when we optimize the quantizer intervals for an asymmetric quantizer (one that is not centered at  $x$ ). A partial answer for this will be given in Part II, where we will see that when the number of quantization intervals tends to infinity, the optimal quantizer is symmetric for symmetric noise distributions.

- Selection of optimal quantization intervals, or equivalently, optimal non central thresholds, was observed to be a difficult problem for nonuniform quantization. The asymptotic design of the optimal quantizer that approaches the optimal finite solution will also be studied in Part II.
- The MLE for binary quantized measurements and a fixed threshold can be obtained in closed form (a1.1). While in the general case it might be obtained numerically. When the noise distribution is log-concave, the Newton's algorithm can be used as an efficient numerical solution (a1.2).
- As the performance degrades when the quantizer range is far from the parameter, the quantizer central threshold must be placed adaptively around the parameter. A simple solution in the binary case is to move the threshold up or down with a constant step. Then, asymptotically, the threshold will settle its mean close to the parameter and it will fluctuates around it. The measurements obtained in this case can be used to have a MLE with asymptotic performance less sensitive to uncertainty on the true parameter value (a2.1).
- By accepting an increased complexity, the central threshold (both in binary and non binary cases) can be set closer and closer to the true parameter by updating it at each time with the MLE based on all the past measurements (a2.2). This scheme asymptotically attains a performance equal to the performance obtained when the threshold is placed at the parameter, which is equivalent to say that this scheme is asymptotically optimal for commonly used noise models.
- When the time goes to infinity the threshold update based on MLE is equivalent to a simple recursive update with decreasing correction gain (1.87). Low complexity recursive schemes of this type and their performance will be studied in detail in Ch. 3.

# Estimation of a varying parameter: what is done and a little more

---

In this chapter we study the estimation of a varying parameter based on quantized measurements. First, we will present the parameter evolution model and the measurement model. Then, we will present the optimal estimator in the MSE sense and its performance. Due to the difficulties that arise when we want to have analytical expressions for the optimal estimator and its performance, we will obtain the optimal estimator using a numerical method. We present and discuss a numerical solution known as particle filtering, which is a method based on Monte Carlo simulation. We give then a bound on its performance using the Bayesian Cramér–Rao bound. After the analysis of the bound, we present a particle filtering scheme based on the quantized prediction error, which is commonly known as quantized innovation. At the end of the chapter, we show that the optimal estimator has, asymptotically, a simple recursive form for a slowly varying parameter. After obtaining the performance for the asymptotically optimal estimator and comparing it to the lower bound on the MSE, we conclude the chapter with a summary and directions for work to be presented in other chapters or to be presented in the future.

## Contributions presented in this chapter:

- *Motivation to use the quantized innovation.* By analyzing a simple signal model, we can obtain a detailed characterization of the bound on the mean squared error for estimation based on quantized measurements. From the bound, we can see clearly that a good estimation scheme can be obtained by quantizing the innovation. This differs from [Ribeiro 2006c] and [You 2008], where the motivation for using the quantized innovation does not come from any quantitative analysis and relies only on intuition.
- *Asymptotically optimal estimator for a slowly varying parameter.* We show that the asymptotically optimal estimator for slowly varying Wiener process parameter can be approximated by a low complexity recursive estimator. We also verify its optimality by comparing it to a lower bound on the mean squared error. The Wiener process model that we consider is a special case of the model in [Ribeiro 2006c], but we do not consider that the noise is Gaussian and we do not impose the quantization to be binary.

**Contents**

---

<b>2.1</b>	<b>Parameter and measurement model . . . . .</b>	<b>77</b>
2.1.1	Parameter model . . . . .	77
2.1.2	Measurement model . . . . .	77
<b>2.2</b>	<b>Optimal estimator . . . . .</b>	<b>78</b>
<b>2.3</b>	<b>Particle Filtering . . . . .</b>	<b>81</b>
2.3.1	Monte Carlo integration . . . . .	81
2.3.2	Importance sampling . . . . .	82
2.3.3	Sequential importance sampling . . . . .	83
2.3.4	Sequential importance resampling . . . . .	85
<b>2.4</b>	<b>Evaluation of the estimation performance . . . . .</b>	<b>87</b>
2.4.1	Online empirical evaluation . . . . .	87
2.4.2	BCRB . . . . .	87
<b>2.5</b>	<b>Quantized innovations . . . . .</b>	<b>90</b>
2.5.1	Prediction and innovation . . . . .	91
2.5.2	Bound for the quantized innovations . . . . .	93
2.5.3	Gaussian assumption and asymptotic estimation of a slow parameter . . . . .	94
<b>2.6</b>	<b>Chapter summary and directions . . . . .</b>	<b>102</b>

---

## 2.1 Parameter and measurement model

### 2.1.1 Parameter model

The parameter to be estimated now is a stochastic process  $\mathbf{X}$  defined on the probability space  $\mathcal{P} = (\Omega, \mathcal{F}, \mathbb{P})$  with values on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . At each instant  $k \in \mathbb{N}^*$ , the corresponding scalar r.v.  $X_k$  will be given by the Wiener process model:

$$X_k = X_{k-1} + W_k, \quad k > 0, \quad (2.1)$$

where  $W_k$  is the  $k$ -th element of a sequence of independent Gaussian r.v.. Its mean is given by  $u_k$  and its variance is a known constant  $\sigma_w^2$ . If  $u_k = 0$  then  $X_k$  forms a standard discrete-time Wiener process, otherwise, it is a Wiener process with drift. The initial distribution of  $X_0$  is supposed to be Gaussian with known mean  $x'_0$  and known variance  $\sigma_0^2$ . The PDF of  $X_0$ , denoted,  $p(x_0)$  is also known as the initial prior of the stochastic process. For estimation purposes, the initial mean represents a guess on the value of  $X_0$  and the initial variance represents the degree of uncertainty on this guess.

From (2.1), we can see that conditioned on  $X_{k-1}$ ,  $X_k$  is independent from the past  $X_{0:k-2}$ . Therefore, this process is a homogeneous Markov process. Until instant  $k$ , it can be characterized by its joint PDF  $p(x_{0:k})$ , which factorizes as follows

$$p(x_{0:k}) = p(x_0) \prod_{j=1}^k p(x_j|x_{j-1}), \quad (2.2)$$

where  $p(x_j|x_{j-1})$  is the conditional PDF of  $X_j$  given  $X_{j-1}$ . This conditional PDF can be written using the Gaussian assumption on  $W_k$  as

$$p(x_j|x_{j-1}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp \left[ -\frac{1}{2} \left( \frac{x_j - x_{j-1} - u_j}{\sigma_w} \right)^2 \right]. \quad (2.3)$$

Therefore, from the knowledge of  $p(x_0)$ ,  $u_k$  and  $\sigma_w$ , we can describe probabilistically the process  $\mathbf{X}$  until any arbitrary instant  $k$  using (2.2) and (2.3).

### 2.1.2 Measurement model

#### Continuous measurement

The process  $\mathbf{X}$  is measured with noise

$$Y_k = X_k + V_k. \quad (2.4)$$

The same assumptions on  $V_k$  as for constant  $x$ , AN1 and AN2, are considered in this case.

#### Quantizer

For tracking the varying parameter, the quantizer will be assumed to be dynamic with varying threshold set  $\boldsymbol{\tau}_k$ :

$$\boldsymbol{\tau}_k = \left[ \tau_{-\frac{N_L}{2},k} \cdots \tau_{-1,k} \quad \tau_{0,k} \quad \tau_{1,k} \cdots \tau_{\frac{N_L}{2},k} \right]^\top.$$

The assumptions on the labeling of the outputs and symmetry, AQ1 and AQ2, are still considered to be valid. The quantized measurements are then given as the output of the quantization function  $Q(\cdot)$  defined in (1.2)

$$i_k = Q(Y_k),$$

where as in the adaptive case, the function  $Q$  can change in time.

## 2.2 Optimal estimator

As it was stated at the beginning of this chapter, we are interested in solving problem (b) (p. 29). That is to estimate  $X_k$  based on the past and present quantized measurements  $i_{1:k}$ . In what follows, we consider that  $\tau_{1:k}$  is a fixed sequence. As in the constant case we want the estimator, or filter in this case, to have **minimum MSE (MMSE)**. We want for all  $k$  an estimator

$$\hat{X}(i_{1:k})$$

minimizing the MSE

$$\text{MSE}_k = \mathbb{E} \left[ \left( \hat{X}_k - X_k \right)^2 \right]. \quad (2.5)$$

As the parameter itself is random, the expectation is evaluated w.r.t. the joint distribution of the measurements  $i_{1:k}$  and the parameter  $X_k$ .

Differently from the deterministic case, when the parameter is random, the general form of the MMSE estimator can be obtained directly from the minimization of  $\text{MSE}_k$ . It can be shown that its general form is [Jazwinski 1970, p. 149]

$$\hat{X}_k = \mathbb{E}_{X_k | i_{1:k}}(X_k), \quad (2.6)$$

where the subscript  $X_k | y_{1:k}$  means that the expectation is evaluated w.r.t. the probability measure of  $X_k$  given a realization of  $i_{1:k}$ . The MMSE estimator is then the posterior mean, *i.e.* the conditional mean of the parameter  $X_k$  given a specific realization sequence of quantized measurements  $i_{1:k}$ <sup>1</sup>.

The MMSE estimator is unbiased since

$$\mathbb{E} \left[ \hat{X}_k \right] = \mathbb{E}_{i_{1:k}} \left[ \mathbb{E}_{X_k | i_{1:k}}(X_k) \right] = \mathbb{E}_{X_k, i_{1:k}}(X_k) = \mathbb{E}(X_k),$$

where the first equality comes from the decomposition of the expectation on the joint variables and the second equality comes from marginalization of the  $i_{1:k}$ .

Similarly, we obtain that the MSE is the mean of the posterior variance

$$\text{MSE}_k = \mathbb{E}_{i_{1:k}} \left\{ \mathbb{E}_{X_k | i_{1:k}} \left\{ \left[ X_k - \mathbb{E}_{X_k | i_{1:k}}(X_k) \right]^2 \right\} \right\} = \mathbb{E}_{i_{1:k}} \left[ \text{Var}_{X_k | i_{1:k}}(X_k) \right]. \quad (2.7)$$

<sup>1</sup>Note that this estimator is different from the MAP, which is the maximum value  $x_k$  that maximizes the posterior  $p(x_k | i_{1:k})$ . It can be shown that the MAP is the optimal estimator under the mean absolute error [Van Trees 1968, pp. 56–57].

Note that for a given realization  $i_{1:k}$ ,  $\text{Var}_{X_k|i_{1:k}}(X_k)$  is the conditional MSE and it can be used when online assessment of the MSE is needed. Online here means that the performance is not averaged on the distribution of the measurements, but evaluated for a given realization.

All the information is contained in the posterior distribution. Its mean is the optimal estimator and its averaged variance is the MMSE. Assuming that the posterior distribution accepts a PDF  $p(x_k|i_{1:k})$ , the MMSE estimator and its MSE are given respectively by

$$\hat{X}_k = \mathbb{E}_{X_k|i_{1:k}}(X_k) = \int_{\mathbb{R}} x_k p(x_k|i_{1:k}) dx_k, \quad (2.8)$$

$$\begin{aligned} \text{MSE}_k &= \mathbb{E}_{i_{1:k}} [\text{Var}_{X_k|i_{1:k}}(X_k)] \\ &= \sum_{i_{1:k} \in \mathcal{I}^{\otimes k}} \left\{ \int_{\mathbb{R}} (x_k - \mathbb{E}_{X_k|i_{1:k}}(X_k))^2 p(x_k|i_{1:k}) dx_k \right\} \mathbb{P}(i_{1:k}). \end{aligned} \quad (2.9)$$

where  $\mathcal{I}^{\otimes k}$  is the joint set where the quantized measurements are defined. To simplify the evaluation of the quantities above, a recursive form for  $p(x_k|i_{1:k})$ , and as a byproduct for  $\mathbb{P}(i_{1:k})$ , can be obtained by using the Markovian property of the dynamical model for the process  $\mathbf{X}$ . The main idea is to write the PDF for prediction  $p(x_k|i_{1:k-1})$  as a function of  $p(x_{k-1}|i_{1:k-1})$  using the dynamical model information  $p(x_k|x_{k-1})$  and then pass from the prediction PDF to the posterior  $p(x_k|i_{1:k})$  using the information given by the measurement  $\mathbb{P}(i_k|x_k)$ . These two expressions, one for prediction using the model and the other for update using the measurement are given respectively by (Why? - App. A.1.6):

$$p(x_k|i_{1:k-1}) = \int_{\mathbb{R}} p(x_k|x_{k-1}) p(x_{k-1}|i_{1:k-1}) dx_{k-1}, \quad (2.10)$$

$$p(x_k|i_{1:k}) = \frac{\mathbb{P}(i_k|x_k) p(x_k|i_{1:k-1})}{\int_{\mathbb{R}} \mathbb{P}(i_k|x'_k) p(x'_k|i_{1:k-1}) dx'_k}. \quad (2.11)$$

The denominator in the RHS of (2.11) is equal to  $\mathbb{P}(i_k|i_{1:k-1})$  (Why? - App. A.1.6), thus this integral can be reused for writing  $\mathbb{P}(i_{1:k})$  in recursive form for  $k > 1$

$$\mathbb{P}(i_{1:k}) = \mathbb{P}(i_k|i_{1:k-1}) \mathbb{P}(i_{1:k-1}) = \left[ \int_{\mathbb{R}} \mathbb{P}(i_k|x_k) p(x_k|i_{1:k-1}) dx_k \right] \mathbb{P}(i_{1:k-1}), \quad (2.12)$$

for  $k = 1$  this probability is

$$\mathbb{P}(i_1) = \int_{\mathbb{R}} \mathbb{P}(i_1|x_0) p(x_0) dx_0. \quad (2.13)$$

In these expressions the prior  $p(x_0)$ , as stated above, is a Gaussian function

$$p(x_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[ -\frac{1}{2} \left( \frac{x_0 - x'_0}{\sigma_0} \right)^2 \right],$$

the conditional PDF  $p(x_k|x_{k-1})$  is given by (2.3) and the probability  $\mathbb{P}(i_k|x_k)$  is given by (1.6) with the dynamical threshold set  $\tau_k$  instead of only one fixed set:

$$\mathbb{P}(i_k|x_k) = \begin{cases} F(\tau_{i_k,k} - x_k) - F(\tau_{i_k-1,k} - x_k), & \text{if } i_k > 0, \\ F(\tau_{i_k+1,k} - x_k) - F(\tau_{i_k,k} - x_k), & \text{if } i_k < 0. \end{cases} \quad (2.14)$$

The general solution to (b) (p. 29) given by the optimal filter is the following:

**Solution to (b) - MMSE estimator for a fixed threshold set sequence  $\tau_{1:k}$**

**(b1) 1) Estimator**

For each time  $k$ , the estimator is given by

$$\hat{X}_k = \mathbb{E}_{X_k|i_{1:k}}(X_k) = \int_{\mathbb{R}} x_k p(x_k|i_{1:k}) \mathbf{d}x_k,$$

where the posterior PDF  $p(x_k|i_{1:k})$  can be evaluated recursively using (2.10) and (2.11).

---

**2) Performance (exact)**

$\hat{X}_k$  is unbiased

$$\mathbb{E}[\hat{X}_k] = \mathbb{E}[X_k]$$

and its MSE for each time  $k$  is

$$\begin{aligned} \text{MSE}_k &= \mathbb{E}_{i_{1:k}} [\text{Var}_{X_k|i_{1:k}}(X_k)] \\ &= \sum_{i_{1:k} \in \mathcal{I}^{\otimes k}} \left\{ \int_{\mathbb{R}} (x_k - \mathbb{E}_{X_k|i_{1:k}}(X_k))^2 p(x_k|i_{1:k}) \mathbf{d}x_k \right\} \mathbb{P}(i_{1:k}) \end{aligned}$$

where now not only (2.10) and (2.11) are used, but also (2.12) and (2.13) to obtain  $\mathbb{P}(i_{1:k})$ .

Some attention must be given to the fact that the MMSE estimator given above and the recursive form for the evaluation of the posterior PDF are quite general and can be applied in many other nonlinear filtering problems.

A major drawback with (b1) is that evaluating the integrals in the prediction/update expressions and in the expectation is analytically intractable. Therefore, we must look for a numerical method for solving it approximately. This will be done next.

## 2.3 Particle Filtering

To obtain the posterior mean (2.8), we must evaluate the integral  $\int_{\mathbb{R}} x_k p(x_k | i_{1:k}) dx_k$ . A general solution is to evaluate it numerically, for example, using a Monte Carlo integration method.

### 2.3.1 Monte Carlo integration

The Monte Carlo integration method consists in approximating the expectation of a function  $g(X)$

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) p(x) dx,$$

where  $p(x)$  is the PDF of  $X$ , by the sample mean calculated using multiple i.i.d. samples  $X^{(j)}$  from the distribution of  $X$  [Robert 1999, p. 83]

$$\mathbb{E}[g(X)] \approx \bar{g}_{N_S} = \frac{1}{N_S} \sum_{j=1}^{N_S} g(x^{(j)}),$$

with  $N_S$  the number of samples and  $x^{(j)}$  the  $j$ -th i.i.d. sample realization.

The approximation is clearly unbiased

$$\mathbb{E} \left[ \frac{1}{N_S} \sum_{j=1}^{N_S} g(X^{(j)}) \right] = \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbb{E} [g(X^{(j)})] = \mathbb{E}[g(X)].$$

By the strong law of large numbers, it converges with probability one to the true expectation  $\mathbb{E}[g(X)]$  [Robert 1999, p. 83]

$$\mathbb{P} \left( \lim_{N_S \rightarrow +\infty} \bar{g}_{N_S} = \mathbb{E}[g(X)] \right) = 1.$$

Moreover, by using a central limit theorem, the asymptotic normalized approximation error  $\varepsilon_{\bar{g}}$  tends to a zero mean Gaussian distribution with variance given by

$$\text{Var}(\varepsilon_{\bar{g}}) = \frac{1}{N_S} \text{Var}[g(X)].$$

Thus, if  $g(X)$  has finite variance, the variance of the approximation reduces by increasing the number of samples.

In our case, we want to approximate the posterior mean

$$\hat{X}_k \approx \frac{1}{N_S} \sum_{j=1}^{N_S} X_k^{(j)}, \quad (2.15)$$

with  $X_k^{(j)}$  i.i.d. samples from the posterior distribution.

Observe that we can also rewrite the posterior mean in an equivalent way using the joint posterior PDF  $p(x_{1:k}|i_{1:k})$ :

$$\hat{X}_k = \int_{\mathbb{R}} x_k p(x_{1:k}|i_{1:k}) dx_{1:k}. \quad (2.16)$$

In this case we will sample independent trajectories  $X_{1:k}^{(j)}$  from  $p(x_{1:k}|i_{1:k})$  and the posterior mean is also given by (2.15).

The main problem here is that the posterior distribution and the joint posterior distribution are usually difficult to sample directly. Therefore, to solve this problem we will use a method called importance sampling.

### 2.3.2 Importance sampling

Retaining the second form of the posterior mean (2.16), the main idea of importance sampling [Robert 1999, p. 92] is to multiply and divide the integrand in the expectation by a PDF  $q(x_{1:k}|i_{1:k})^2$  from which we know how to sample the trajectories  $X_{1:k}$ . This gives

$$\hat{X}_k = \int_{\mathbb{R}} x_k \frac{p(x_{1:k}|i_{1:k})}{q(x_{1:k}|i_{1:k})} q(x_{1:k}|i_{1:k}) dx_{1:k}.$$

Note that the support of the PDF  $q(x_{1:k}|i_{1:k})$  might be strictly larger than the support of the posterior. Denoting the ratio between PDF as an importance weight  $w(x_{1:k})$

$$w(x_{1:k}) = \frac{p(x_{1:k}|i_{1:k})}{q(x_{1:k}|i_{1:k})}, \quad (2.17)$$

the expectation can be approximated by

$$\hat{X}_k = \int_{\mathbb{R}} x_k w(x_{1:k}) q(x_{1:k}|i_{1:k}) dx_{1:k} \approx \frac{1}{N_S} \sum_{j=1}^{N_S} X_k^{(j)} w(X_{1:k}^{(j)}), \quad (2.18)$$

where  $X_{1:k}^{(j)}$  are i.i.d. trajectories from  $q(x_{1:k}|i_{1:k})$ . We can divide the expectation by the integral of the posterior as its value is equal to one, this gives

$$\hat{X}_k = \frac{\int_{\mathbb{R}} x_k w(x_{1:k}) q(x_{1:k}|i_{1:k}) dx_k}{\int_{\mathbb{R}} w(x_{1:k}) q(x_{1:k}|i_{1:k}) dx_k} \approx \frac{\sum_{j=1}^{N_S} X_k^{(j)} w(X_{1:k}^{(j)})}{\sum_{j=1}^{N_S} w(X_{1:k}^{(j)})}.$$

Defining the normalized weights  $\tilde{w}(X_{1:k}^{(j)})$  as

$$\tilde{w}(X_{1:k}^{(j)}) = \frac{w(X_{1:k}^{(j)})}{\sum_{j=1}^{N_S} w(X_{1:k}^{(j)})}, \quad (2.19)$$

---

<sup>2</sup>Note that  $q(x_{1:k}|i_{1:k})$  can depend on the measurements, after we will choose a simplified form which does not depend on the measurements.

we have that the posterior mean can be approximated by

$$\hat{X}_k \approx \sum_{j=1}^{N_S} X_k^{(j)} \tilde{w} \left( X_{1:k}^{(j)} \right). \quad (2.20)$$

By comparing the approximation in (2.20) with the integral  $\int_{\mathbb{R}} x_k p(x_k | i_{1:k}) dx_k$ , we realize that this method is equivalent to approximate the posterior by a discrete distribution with support values chosen randomly and with probabilities given by the normalized weights

$$p(x_k | i_{1:k}) \approx \sum_{j=1}^{N_S} \tilde{w} \left( x_{1:k}^{(j)} \right) \delta_D \left( x_k - x_k^{(j)} \right), \quad (2.21)$$

where  $\delta_D(\cdot)$  is a Dirac distribution.

### 2.3.3 Sequential importance sampling

The remaining problems now are the choice of a PDF  $q(x_{1:k} | i_{1:k})$  easy to sample and the evaluation of the weights.

To be able to sample the trajectory  $X_{1:k}^{(j)}$  without modifying the past trajectory  $X_{1:k-1}^{(j)}$  (so that we do not need to resample the past trajectory), we must choose a distribution  $q(x_{1:k} | i_{1:k})$  for which the marginal distribution for  $k-1$  is exactly  $q(x_{1:k-1} | i_{1:k-1})$ . This can be done using the following form for  $q(x_{1:k} | i_{1:k})$  [Doucet 1998]:

$$q(x_{1:k} | i_{1:k}) = q(x_{1:k-1} | i_{1:k-1}) q(x_k | x_{1:k-1}, i_{1:k}). \quad (2.22)$$

In this case to extend a sample trajectory from realization  $x_{1:k-1}^{(j)}$  to  $x_{1:k}^{(j)}$ , we sample  $q(x_k | x_{1:k-1}^{(j)}, i_{1:k})$  to generate the new point of the trajectory  $x_k^{(j)}$ .

To evaluate the weights, we develop  $p(x_{1:k} | i_{1:k})$  using conditioning and the independence assumptions on the model:

- $X_k$  is independent of  $X_{1:k-2}$  and  $i_{1:k-1}$  conditioned on  $X_{k-1}$ ,
- $i_k$  is independent of  $X_{1:k-1}$  and  $i_{1:k-1}$  conditioned on  $X_k$ .

This gives

$$p(x_{1:k} | i_{1:k}) = \frac{\mathbb{P}(i_k | x_k) p(x_k | x_{k-1})}{\mathbb{P}(i_k | i_{1:k-1})} p(x_{1:k-1} | i_{1:k-1}). \quad (2.23)$$

Replacing the simplified form of  $q(x_{1:k} | i_{1:k})$  (2.22) and the joint posterior above (omitting  $\mathbb{P}(i_k | i_{1:k-1})$ , which is constant in  $x_{1:k}$ ) in the expression (2.17), we have the following weight for the trajectory  $j$ :

$$w \left( x_{1:k}^{(j)} \right) \propto \frac{\mathbb{P}(i_k | x_k^{(j)}) p \left( x_k^{(j)} | x_{k-1}^{(j)} \right) p \left( x_{1:k-1}^{(j)} | i_{1:k-1} \right)}{q \left( x_{1:k-1}^{(j)} | i_{1:k-1} \right) q \left( x_{1:k-1}^{(j)} | i_{1:k-1} \right)}, \quad (2.24)$$

where  $\propto$  is the symbol for proportional. The fact that the weights are defined up to a proportional factor is not important because for approximating the posterior mean we use the normalized weights. Note that the factor  $\frac{p(x_{1:k-1}^{(j)}|i_{1:k-1})}{q(x_{1:k-1}^{(j)}|i_{1:k-1})}$  is the weight for the samples at time  $k-1$ . Thus, we can write a recursive expression that relates the normalized weights for time  $k-1$  with the weights for time  $k$

$$w(x_{1:k}^{(j)}) \propto \frac{\mathbb{P}(i_k|x_k^{(j)}) p(x_k^{(j)}|x_{k-1}^{(j)})}{q(x_k^{(j)}|x_{0:k-1}, i_{1:k})} \tilde{w}(x_{1:k-1}^{(j)}). \quad (2.25)$$

We need now to define the PDF  $q(x_k|x_{0:k-1}, i_{1:k})$  which is used to generate the samples. The two most commonly used choices are the following:

- *choice 1:  $p(x_k|x_{k-1}, i_k)$ , minimum weight variance distribution.* The quality of the approximation of the posterior by the discrete distribution (2.21) is dependent on the variance of the weights and the variance depends on the PDF  $q(x_k|x_{0:k-1}, i_{1:k})$ . It can be shown that conditioned on the past trajectory  $x_{1:k-1}^{(j)}$  realization and on the measurements realization  $i_{1:k}$ , the variance of the weights is minimized for [Doucet 1998]

$$q(x_k|x_{0:k-1}, i_{1:k}) = p(x_k|x_{k-1}, i_k). \quad (2.26)$$

Unfortunately this distribution is difficult to sample directly. In our case, we can sample from it by using a rejection method (More? - App. A.2.5).

- *Choice 2:  $p(x_k|x_{k-1}, i_k)$ , prior distribution.* In order to simplify the evaluation of the weights we can choose

$$q(x_k|x_{0:k-1}, i_{1:k}) = p(x_k|x_{k-1}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left[-\frac{1}{2}\left(\frac{x_k - x_{k-1} - u_k}{\sigma_w}\right)^2\right]. \quad (2.27)$$

Thus for each previous  $x_{k-1}^{(j)}$ , we are going to obtain a sample from a r.v.  $X_k^{(j)}$  using the distribution  $p(x_k|x_{k-1}^{(j)})$ <sup>3</sup>. In our case, this choice reduces the problem to sampling from a Gaussian distribution, which is very simple, and updating the weights following (we chose the proportionality factor to be one)

$$w(x_{1:k}^{(j)}) = \mathbb{P}(i_k|x_k^{(j)}) \tilde{w}(x_{1:k-1}^{(j)}). \quad (2.28)$$

Note that in both cases the sampling and evaluation of the weights do not require the past measurements and the samples  $x_{1:k-2}^{(j)}$ . This leads to memory requirements that do not increase over time. If we compare both choices in terms of complexity, the second choice is better because it only requires sampling from a Gaussian distribution and evaluating the weights with the likelihood. Therefore, from now on, we will use the second choice for the sampling distribution.

We have the following procedure:

---

<sup>3</sup>For details on how to sample from it using a standard Gaussian variate see (How? - App. A.3.3).

1. Sample the prior distribution  $p(x_0)$ . This will generate  $N_S$  samples  $x_0^{(1:N_S)}$ . Set uniform normalized weights  $\tilde{w}(x_0^{(j)}) = \frac{1}{N_S}$ .  
For time  $k$ ,

2. Create  $N_S$  samples each from the corresponding r.v.  $X_k^{(j)}$  with PDF given by (2.27):

$$p(x_k | x_{k-1}^{(j)}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp \left[ -\frac{1}{2} \left( \frac{x_k - x_{k-1}^{(j)} - u_k}{\sigma_w} \right)^2 \right].$$

3. Evaluate the sample weights using the measurement and the last weights with (2.28):

$$w(x_{1:k}^{(j)}) = \mathbb{P}(i_k | x_k^{(j)}) \tilde{w}(x_{1:k-1}^{(j)}).$$

4. Normalize the weights using (2.19):

$$\tilde{w}(x_{1:k}^{(j)}) = \frac{w(x_{1:k}^{(j)})}{\sum_{j=1}^{N_S} w(x_{1:k}^{(j)})}.$$

5. Obtain the estimate with the weighted mean

$$\hat{x}_k \approx \sum_{j=1}^{N_S} x_k^{(j)} \tilde{w}(x_{1:k}^{(j)}).$$

This procedure is the sequential extension of importance sampling applied to filtering and this is the reason for its commonly used name - sequential importance sampling filter. As this method is a special case of importance sampling, it has the same general characteristics, namely, it is biased for a fixed number of samples, but it converges with probability one to the optimal estimator when  $N_S \rightarrow \infty$  [Doucet 1998].

### 2.3.4 Sequential importance resampling

We would expect that by increasing the number of samples the filter would get closer and closer to the optimal estimate. However, the convergence result is asymptotic, it works only when  $N_S$  tends to infinity. When  $N_S$  is finite, it can be shown that the variance of the weights increases over the time [Kong 1994]. This problem is known as the degeneracy problem and what happens in practice is that after some time most of the normalized weights are close to zero, which is equivalent to say that most of the samples are useless [Doucet 1998].

In the case of sampling with  $p(x_k | x_{k-1}^{(j)})$ , the cause of this problem is easy to understand. We start with a given prior distribution, then during the procedure we evaluate the posterior for values of  $X_k$  sampled randomly using  $p(x_k | x_{k-1}^{(j)})$ , as there is no feedback from the measurements in the sampling processes, after some time, the samples can lie very far from the values of  $X_k$  where the posterior has larger values. As a consequence, this will produce a very poor discrete approximation of the posterior. A possible remedy for this problem is to drive the sampling process using the measurements  $i_{1:k}$ .

## Resampling

This can be done in a simple way by reproducing the samples  $x_k^{(j)}$  for which the posterior approximation  $\tilde{w}(x_{1:k}^{(j)})$  is large and deleting the samples for which the posterior is small. This procedure, known as resampling, can be carried out in practice by sampling  $N_S$  times<sup>4</sup> the posterior discrete approximation given by

$$\mathbb{P}(x_k) = \begin{cases} \tilde{w}(x_{1:k}^{(j)}), & \text{if } x_k = x_k^{(j)}, \\ 0, & \text{otherwise.} \end{cases} \quad (2.29)$$

After resampling, for retaining the posterior approximation, the weights of the samples are set to

$$\tilde{w}(x_{1:k}^{(j)}) = \frac{1}{N_S}. \quad (2.30)$$

As the posterior approximation is a multinomial distribution, the procedure of resampling using the approximation of the posterior (2.29) is known as multinomial resampling. Multinomial resampling can be easily implemented using  $N_S$  independent uniform samples, for details see (How? - App. A.3.4) (app4) and for other types of resampling techniques see [Hol 2006].

The process of resampling should not be performed every time as it leads to the impoverishment of the sample set [Berzuini 1997]. Sample impoverishment comes as the opposite extremum of the degeneracy problem, as in this case we simply neglect possible trajectories of  $X_k$  with medium and low likelihood, leading to a not sufficiently rich approximation of the posterior. For triggering the resampling process we can monitor the number of effective samples  $N_{\text{eff}}$ , that is to say, the equivalent number of samples if we were using the true posterior for Monte Carlo evaluation. This number can be approximated by [Doucet 1998]

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^{N_S} \tilde{w}^2(x_{1:k}^{(j)})}. \quad (2.31)$$

Therefore, each time  $N_{\text{eff}} < N_{\text{thresh}}$ , where  $N_{\text{thresh}} \in [1, N_S]$  is a minimum acceptable number of effective samples, the resampling process is triggered.

Sequential importance sampling with the resampling step for general Bayesian estimation was first suggested in [Rubin 1988](cited in [Doucet 1998]) under the name sequential importance resampling. Its widespread use in filtering with the specific choice of  $p(x_k|x_{k-1})$  as the sampling distribution was initiated with [Gordon 1993] under the name of bootstrap filter. This method was proposed for solving general nonlinear non Gaussian filtering problems.

The method presented above can be found in the literature under many other names, the most common is **particle filter (PF)**. In this case "particle" is the name given for a sample  $x_k^{(j)}$ . We will use the terms particle filter and particle from now on.

A proof of convergence of the general PF is given in [Berzuini 1997] for the case with resampling at each iterate. It is shown that when  $N_S \rightarrow \infty$  the error between the optimal

---

<sup>4</sup>We could resample more or less than  $N_S$  samples, we chose  $N_S$  because it is the most commonly used choice in the literature.

estimator and the PF estimate multiplied by  $\sqrt{N_S}$  tends to a Gaussian r.v. with fixed finite variance. This means that, for a large number of particles, when the number of particles increases the PF estimate is more and more concentrated around the optimal estimator.

Application of PF for estimation based on quantized measurements with a fixed sequence of threshold sets are reported in [Ruan 2004] and [Karlsson 2005]. In [Ruan 2004] the main focus is on analyzing the main issues related to the fusion of quantized measurements from multiple sensors for tracking in general, the results reported therein are given by simulation. A more restricted model with  $X_k$  given by a vector linear Gaussian evolution and quantized linear Gaussian measurement is used in [Karlsson 2005], where a theoretical lower bound on estimation performance is obtained and compared with simulation results. The bound that is used is the equivalent counterpart of the CRB for random parameters the **Bayesian Cramér–Rao bound (BCRB)**.

## 2.4 Evaluation of the estimation performance

We have already explained how to obtain the estimates for our problem (b) (p. 29). We still need to evaluate its performance.

### 2.4.1 Online empirical evaluation

The variance of the posterior approximation (supposing that  $N_S$  is sufficiently large for the bias to be negligible)

$$\text{MSE}_k \approx \sum_{j=1}^{N_S} \left( x_k^{(j)} - \hat{x}_k \right)^2 \tilde{w} \left( x_{1:k}^{(j)} \right),$$

gives an online estimate of the MSE. The problem with this approach is that the performance is conditioned on the given measurement sequence  $i_{1:k}$ . In this case, approximated performance can be obtained only after having the measurements, thus no design of the system (choice of the number of quantization intervals  $N_I$ , choice of the sensor quality  $\delta$ ) can be done. Even if we push more into the Monte Carlo philosophy and try to evaluate the mean of the approximated MSE above using Monte Carlo integration, we will have to simulate a large number of times the PF procedure by changing the parameters needed for system design ( $N_I$ ,  $\delta$ ,  $\sigma_w^2$ ). Therefore, it is better to turn our attention to analytical results on performance.

### 2.4.2 BCRB

The analytical form of the MSE (2.9) depends on the posterior distribution. Thus, for the same reason, we cannot have an analytical expression for the estimator, we are not going to have an analytical expression for the MSE. We must resort then to a bound on the MSE. As a consequence, we will follow [Karlsson 2005] and we will also analyze the BCRB. As our case is simpler ( $X_k$  is a scalar Wiener process) than the vector linear case studied in [Karlsson 2005], we will be able to analyze the effects of the measurement system parameters in a more clear and simple way.

The BCRB at instant  $k$ ,  $\text{BCRB}_k$ , is a lower bound on  $\text{MSE}_k$ , it is given by the inverse of the **Bayesian information (BI)** [Van Trees 1968, p. 84]

$$\text{MSE}_k \geq \text{BCRB}_k = \frac{1}{J_k}. \quad (2.32)$$

The BI at time  $k$ ,  $J_k$ , is given by

$$J_k = -\mathbb{E} \left[ \frac{\partial^2 \log p(X_k, i_{1:k})}{\partial X_k^2} \right]. \quad (2.33)$$

As  $X_k$  is random, the expectation here is evaluated using the joint probability measure of  $X_k$  and  $i_{1:k}$ . This result is general and it is not linked particularly to the quantization problem, we could replace  $i_{1:k}$  by any measurement related to  $X_k$ .

By assuming that  $X_k$  is a Markov process (also here  $i_{1:k}$  can be any type of measurement), in [Tichavsky 1998] a recursive form for evaluating the BI is obtained

$$J_k = C_k - \frac{B_k^2}{A_k + J_{k-1}}, \quad (2.34)$$

where<sup>5</sup>  $J_0 = -\mathbb{E} \left[ \frac{\partial^2 \log p(X_0)}{\partial X_0^2} \right]$  and  $A_k = -\mathbb{E} \left[ \frac{\partial^2 \log p(X_k | X_{k-1})}{\partial X_{k-1}^2} \right]$ ,  $B_k = -\mathbb{E} \left[ \frac{\partial^2 \log p(X_k | X_{k-1})}{\partial X_k \partial X_{k-1}} \right]$ ,  $C_k = -\mathbb{E} \left[ \frac{\partial^2 \log p(X_k | X_{k-1})}{\partial X_k^2} \right] - \mathbb{E} \left[ \frac{\partial^2 \log \mathbb{P}(i_k | X_k)}{\partial X_k^2} \right]$ .

Using (2.3) for evaluating the terms  $A_k$ ,  $B_k$  and  $C_k$ , we have

$$\begin{aligned} A_k &= -\mathbb{E} \left\{ \frac{\partial^2 \log \left\{ \frac{1}{\sqrt{2\pi}\sigma_w} \exp \left[ -\frac{1}{2} \left( \frac{X_k - X_{k-1} - u_k}{\sigma_w} \right)^2 \right] \right\}}{\partial X_{k-1}^2} \right\} = -\mathbb{E} \left\{ \frac{\partial^2 \left[ -\frac{1}{2} \left( \frac{X_k - X_{k-1} - u_k}{\sigma_w} \right)^2 \right]}{\partial X_{k-1}^2} \right\} \\ &= \frac{1}{\sigma_w^2}. \end{aligned}$$

In the same way

$$\begin{aligned} B_k &= -\frac{1}{\sigma_w^2}, \\ C_k &= \frac{1}{\sigma_w^2} - \mathbb{E} \left[ \frac{\partial^2 \log \mathbb{P}(i_k | X_k)}{\partial X_k^2} \right]. \end{aligned}$$

Decomposing the expectation above, we obtain

$$C_k = \frac{1}{\sigma_w^2} + \mathbb{E}_{X_k} \left\{ -\mathbb{E}_{i_k | X_k} \left[ \frac{\partial^2 \log \mathbb{P}(i_k | X_k)}{\partial X_k^2} \right] \right\}.$$

The inner expectation is another form of expressing the FI for estimating  $X_k$  when  $X_k$  is considered to be a deterministic parameter [Kay 1993, p. 34]. Thus, by using the parametrization of the FI for quantized measurements with the r.v.  $\varepsilon_k = \tau_{0,k} - X_k$ , we can write

$$C_k = \frac{1}{\sigma_w^2} + \mathbb{E} [I_q(\varepsilon_k)].$$

<sup>5</sup>Note that we are using the notation for discrete measurements  $i_{1:k}$  with  $\mathbb{P}(i_k | x_k)$ .

where the expectation is evaluated using the probability measure of  $\varepsilon_k$ . Using these results in (2.34) gives

$$J_k = \frac{1}{\sigma_w^2} + \mathbb{E} [I_q(\varepsilon_k)] - \frac{1}{\sigma_w^4 \left( \frac{1}{\sigma_w^2} + J_{k-1} \right)}, \quad (2.35)$$

with  $J_0$  given by

$$J_0 = -\mathbb{E} \left\{ \frac{\partial^2 \log \left\{ \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[ -\frac{1}{2} \left( \frac{X_0 - x'_0}{\sigma_0} \right)^2 \right] \right\}}{\partial X_0^2} \right\} = \frac{1}{\sigma_0^2}. \quad (2.36)$$

For commonly used noise models (Gaussian, Laplacian and Cauchy), the FI is maximized for  $\varepsilon_k = 0$ . Thus, we can obtain a simple upper bound on the BI by assuming  $\varepsilon_k = 0$  with probability one. This gives

$$J_k \leq J'_k = \frac{1}{\sigma_w^2} + I_q(0) - \frac{1}{\sigma_w^4 \left( \frac{1}{\sigma_w^2} + J'_{k-1} \right)}, \quad (2.37)$$

with  $J'_0 = J_0$ . This will give a simple lower bound on the BCRB and consequently on the MSE, which can be used to assess approximately the performance of the PF.

The solution to problem (b) (p. 29) given by the PF is

**Solution to (b) - Particle filter for a fixed threshold set sequence  $\tau_{1:k}$**

(b1.1) 1) Estimator

- Set uniform normalized weights  $\tilde{w}(x_0^{(j)}) = \frac{1}{N_S}$  and initialize  $N_S$  particles  $\{x_0^{(1)}, \dots, x_0^{(N_S)}\}$  by sampling the prior

$$p(x_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{x_0 - x'_0}{\sigma_0}\right)^2\right].$$

For each time  $k$ ,

- for  $j$  from 1 to  $N_S$ , sample the r.v.  $X_k^{(j)}$  with PDF (How? - App. A.3.3)

$$p(x_k | x_{k-1}^{(j)}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left[-\frac{1}{2}\left(\frac{x_k - x_{k-1}^{(j)} - u_k}{\sigma_w}\right)^2\right],$$

- for  $j$  from 1 to  $N_S$ , evaluate and normalize the weights

$$w(x_{1:k}^{(j)}) = \mathbb{P}(i_k | x_k^{(j)}) \tilde{w}(x_{1:k-1}^{(j)}), \quad \tilde{w}(x_{1:k}^{(j)}) = \frac{w(x_{1:k}^{(j)})}{\sum_{j=1}^{N_S} w(x_{1:k}^{(j)})},$$

where  $\mathbb{P}(i_k | x_k^{(j)})$  is given by (2.14).

- Obtain the estimate with the weighted mean

$$\hat{x}_k \approx \sum_{j=1}^{N_S} x_k^{(j)} \tilde{w}(x_{1:k}^{(j)}).$$

- Evaluate the number of effective particles

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^{N_S} \tilde{w}^2(x_{1:k}^{(j)})},$$

if  $N_{\text{eff}} < N_{\text{thresh}}$ , then resample using multinomial resampling (How? - App. A.3.4) (app4).

2) Performance (lower bound)

The MSE can be lower bounded as follows

$$\text{MSE}_k \geq \frac{1}{J'_k},$$

with  $J'_k$  given recursively by

$$J'_k = \frac{1}{\sigma_w^2} + I_q(0) - \frac{1}{\sigma_w^4} \frac{1}{\left(\frac{1}{\sigma_w^2} + J'_{k-1}\right)}.$$

## 2.5 Quantized innovations

For commonly used symmetrically distributed noise models (Gaussian, Laplacian and Cauchy distributions), we saw in Ch. 1 that  $I_q(\varepsilon)$  around  $\varepsilon = 0$  is a locally decreasing function with  $|\varepsilon|$ , thus from (2.35) we can see that closer  $\tau_{0,k}$  is to the parameter realization  $x_k$ , higher will

be the BI. If we assume that the BCRB is sufficiently tight for accepting its behavior as an approximation of the behavior of the MSE, closer  $\tau_{0,k}$  is to the parameter realization  $x_k$ , lower will be the MSE. This indicates that the dynamical range of the quantizer must vary in time in order to follow the parameter and produce enhanced estimation performance.

### 2.5.1 Prediction and innovation

**Prediction.** The main problem with the approach  $-\tau_{0,k} = x_k$  is that we do not know  $x_k$ , if we knew, we would not need to estimate it. We might then accept a small loss of performance by using the closest value to  $x_k$  that we have in hand, in our case, a prediction of  $x_k$  using the last estimate value  $\hat{x}_{k-1}$  and the drift  $u_k$ . If  $\hat{X}_{k-1}$  is the MMSE estimator based on  $i_{1:k-1}$ , then the MMSE estimator of  $X_k$  based also on  $i_{1:k-1}$ , denoted  $\hat{X}_{k|k-1}$ , is the conditional mean [Jazwinski 1970, p. 149]

$$\hat{X}_{k|k-1} = \mathbb{E}_{X_k|i_{1:k-1}}(X_k).$$

Using the dynamical model for  $X_k$  and the linearity of the conditional expectation

$$\hat{X}_{k|k-1} = \mathbb{E}_{X_k|i_{1:k-1}}(X_{k-1} + W_k) = \mathbb{E}_{X_{k-1}|i_{1:k-1}}(X_{k-1}) + \mathbb{E}_{W_k|i_{1:k-1}}(W_k).$$

The first term is the optimal estimate for  $X_{k-1}$ . As  $W_k$  is independent of all  $W_n$  with  $n \neq k$  and it is also independent of all  $V_k$ , it does not depend on  $i_{1:k-1}$ . Thus, the second term is simply  $\mathbb{E}_{W_k}(W_k)$ , which is  $u_k$ . This gives

$$\hat{X}_{k|k-1} = \tau_{0,k} = \hat{X}_{k-1} + u_k.$$

Considering that the estimator is good enough (at least for large  $k$ ), we expect to have the r.v.  $\varepsilon_k$  with most of its probability concentrated around zero, thus leading to a higher  $\mathbb{E}[I_q(\varepsilon_k)]$  and, consequently, to a lower MSE.

**Quantizing the Innovation.** Quantizing the prediction error is a known subject in standard quantization. It is widely known under the name predictive quantization [Gersho 1992, Ch. 7]. Note that the procedure considered here is different. Instead of quantizing the prediction error of reconstruction, we quantize the error between the prediction in estimation sense and the noisy measurement  $Y_k - \hat{X}_{k|k-1}$ . The prediction error in this case is commonly called the innovation process in continuous measurement linear filtering theory [Kay 1993, p. 433]. The name comes from the fact that it represents the previously unknown information brought by the new measurement. As a consequence, the quantized prediction error for estimation purposes is called quantized innovation.

We can slightly change solution (b1.1) (p. 90) by adding the adaptive replacement of the central threshold with the prediction<sup>6</sup>. This is what was done in [Sukhavasi 2009b] under the assumption of Gaussian noise, linear and Gaussian vector  $\mathbf{X}_k$  ( $\mathbf{X}_k = \mathbf{A}\mathbf{X}_{k-1} + W_k$ ) and binary quantization. Constraining  $\mathbf{X}_k$  to be the scalar Wiener process considered here and generalizing the algorithm for symmetrically distributed noise and  $N_I \geq 2$ , we have

<sup>6</sup>As the measurements are now linked through the use of the prediction for quantizing, we cannot guarantee the convergence of the particle approximation through standard results and more advanced results are needed [Crisan 2000].

Solution to (b) - Particle filter with adaptive threshold sequence  $\tau_{1:k}$  quantizing the innovation.

(b2.1) 1) Estimator

- Set uniform normalized weights  $\tilde{w}(x_0^{(j)}) = \frac{1}{N_S}$  and initialize  $N_S$  particles  $\{x_0^{(1)}, \dots, x_0^{(N_S)}\}$  by sampling with prior

$$p(x_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{x_0 - x'_0}{\sigma_0}\right)^2\right].$$

For each time  $k$ ,

- for  $j$  from 1 to  $N_S$ , sample the r.v.  $X_k^{(j)}$  with PDF (How? - App. A.3.3)

$$p(x_k | x_{k-1}^{(j)}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left[-\frac{1}{2}\left(\frac{x_k - x_{k-1}^{(j)} - u_k}{\sigma_w}\right)^2\right],$$

- for  $j$  from 1 to  $N_S$ , evaluate and normalize the weights

$$w(x_{1:k}^{(j)}) = \mathbb{P}(i_k | x_k^{(j)}) \tilde{w}(x_{1:k-1}^{(j)}), \quad \tilde{w}(x_{1:k}^{(j)}) = \frac{w(x_{1:k}^{(j)})}{\sum_{j=1}^{N_S} w(x_{1:k}^{(j)})},$$

where  $\mathbb{P}(i_k | x_k^{(j)})$  is given by (2.14).

- Obtain the estimate with the weighted mean

$$\hat{x}_k \approx \sum_{j=1}^{N_S} x_k^{(j)} \tilde{w}(x_{1:k}^{(j)}).$$

- Set the central threshold of the quantizer to the new estimate

$$\tau_{0,k} = \hat{x}_{k-1} + u_k.$$

- Evaluate the number of effective particles

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^{N_S} \tilde{w}^2(x_{1:k}^{(j)})},$$

if  $N_{\text{eff}} < N_{\text{thresh}}$ , then resample using multinomial resampling (How? - App. A.3.4) (app4).

2) Performance (lower bound)

The MSE can be lower bounded as follows

$$\text{MSE}_k \geq \frac{1}{J'_k},$$

with  $J'_k$  given recursively by

$$J'_k = \frac{1}{\sigma_w^2} + I_q(0) - \frac{1}{\sigma_w^4} \frac{1}{\left(\frac{1}{\sigma_w^2} + J'_{k-1}\right)}.$$

### 2.5.2 Bound for the quantized innovations

Observe that the lower bound is still valid because we still have  $\mathbb{E}[I_q(\varepsilon_k)] \leq I_q(0)$ . But, we might have a performance closer to the bound, as  $\varepsilon_k$  might be concentrated mostly around zero. It is also important to note that now even if the bound is tight (which might not be true), we can get very close to it, but in general we cannot attain it. This is due to the fact that the MSE for a varying parameter never goes to zero, leading to a residual spread on the PDF of  $\varepsilon_k$ , which makes  $\mathbb{E}[I_q(\varepsilon_k)] \leq I_q(0)$ .

To have an approximation on the evolution of the MSE, we can analyze the lower bound on the BCRB. Therefore, we are interested in analyzing the evolution of  $J'_k$ . We can start by analyzing the evolution of its increments. Subtracting the expressions for  $J'_k$  and  $J'_{k-1}$ , we have

$$J'_k - J'_{k-1} = \frac{1}{\sigma_w^4} \left( \frac{1}{\frac{1}{\sigma_w^2} + J'_{k-2}} - \frac{1}{\frac{1}{\sigma_w^2} + J'_{k-1}} \right) = \frac{1}{\sigma_w^4} \frac{J'_{k-1} - J'_{k-2}}{\left( \frac{1}{\sigma_w^2} + J'_{k-1} \right) \left( \frac{1}{\sigma_w^2} + J'_{k-2} \right)}. \quad (2.38)$$

The BI is positive by definition, as it is an expectation of a squared quantity, and  $\sigma_w^2$  is also positive by definition, thus the denominator of the expression above is always positive. This leads to a sign of the increment at time  $k-1$  that is the same as the sign of the increment at  $k-2$ . As a conclusion, we can say that the BI is monotonic, it always increases or decreases. For determining if the BI increases or decreases, we can see from the recursive expression above that this will be determined by the first increment  $J'_1 - J'_0$ . By subtracting  $J'_0 = \frac{1}{\sigma_0^2}$  from  $J'_1$ , we obtain

$$J'_1 - J'_0 = \frac{1}{\sigma_w^2} + I_q(0) - \frac{1}{\sigma_w^4} \frac{1}{\frac{1}{\sigma_w^2} + \frac{1}{\sigma_0^2}} - \frac{1}{\sigma_0^2}.$$

Regrouping the terms with factor  $\frac{1}{\sigma_w^2}$  gives

$$J'_1 - J'_0 = I_q(0) + \frac{1}{\sigma_w^2} \left[ 1 - \frac{1}{1 + \frac{\sigma_w^2}{\sigma_0^2}} \right] - \frac{1}{\sigma_0^2} = I_q(0) + \frac{1}{\sigma_w^2 + \sigma_0^2} - \frac{1}{\sigma_0^2}.$$

Thus, if

$$I_q(0) > \frac{1}{\sigma_0^2} - \frac{1}{\sigma_w^2 + \sigma_0^2},$$

$J'_1 - J'_0$  is positive and the BI is always increasing, otherwise it always decreases. As a consequence, if the inequality is satisfied the BCRB is always decreasing, otherwise always increasing.

As stated before, the information bound  $J'_k$  is bounded below by zero. By looking to (2.37)

$$J_k \leq J'_k = \frac{1}{\sigma_w^2} + I_q(0) - \frac{1}{\sigma_w^4} \frac{1}{\left( \frac{1}{\sigma_w^2} + J'_{k-1} \right)},$$

we can see that it is bounded above by  $I_q(0) + \frac{1}{\sigma_w^2}$ , as the other term that is subtracted is always positive. Joining the facts that  $J'_k$  is lower and upper bounded with the fact that it

is always increasing or decreasing, we can conclude that  $J'_k$  will converge to a fixed point (a fixed value). Except in the cases when the inequality above is an equality, from (2.38), we see that the increment  $J'_k - J'_{k-1}$  cannot be zero, as it is equal to the last increment (which is positive) multiplied by a positive value. Therefore, the fixed point of  $J'_k$  is expected to be attained only asymptotically.

Denoting this asymptotic fixed point  $J'_\infty$ , by definition it is the value of  $J'_k$  for which  $J'_k = J'_{k-1}$ . Thus, it can be found by solving

$$J'_\infty = \frac{1}{\sigma_w^2} + I_q(0) - \frac{1}{\sigma_w^4} \frac{1}{\left(\frac{1}{\sigma_w^2} + J'_\infty\right)},$$

which is equivalent to solve

$$J'^2_\infty - I_q(0) J'_\infty - \frac{I_q(0)}{\sigma_w^2} = 0.$$

The solutions for the equation above are

$$\frac{I_q(0) \pm \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}}{2}.$$

In order to have  $J'_\infty$  positive (it is positive by definition), we must take the positive solution. As  $\frac{4I_q(0)}{\sigma_w^2}$  is positive, the positive solution is obtained for the positive sign. Therefore,

$$J'_\infty = \frac{I_q(0) + \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}}{2} \quad (2.39)$$

and the asymptotic MSE is then lower bounded by the inverse of  $J'_\infty$

$$\text{MSE}_\infty \geq \frac{2}{I_q(0) + \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}}. \quad (2.40)$$

The following behaviors can then be obtained for the evolution of the bound: if we start with a very small  $\sigma_0^2$  (small compared with  $\frac{1}{I_q(0)}$ ), as we can see from the inequality related to the monotonicity pattern, the lower bound on the MSE will always increase, tending asymptotically to  $\frac{1}{J'_\infty}$ . If we start with a large  $\sigma_0^2$ , the lower bound will always decrease, also tending asymptotically to  $\frac{1}{J'_\infty}$ .

From the analysis we can see that the MSE, as expected, is always strictly positive, it is lower bounded by  $\sigma_0^2$  when this value is very small compared with  $\frac{1}{I_q(0)}$  and it is lower bounded by  $\frac{1}{J'_\infty}$  when  $\sigma_0^2$  is large compared with  $\frac{1}{I_q(0)}$ .

### 2.5.3 Gaussian assumption and asymptotic estimation of a slowly varying parameter

Other filtering methods based on the quantized innovation are proposed in the literature under the Gaussian noise assumption. In [Ribeiro 2006c], binary measurements are obtained

by applying the sign function to the innovation. A similar procedure to the well-known Kalman filter [Kay 1993, Ch. 13] is derived by assuming that the posterior at instant  $k-1$  is Gaussian. In the same line, [You 2008] proposes a Kalman-like procedure for quantized innovations with  $N_I \geq 2$ . A careful reader of the literature on the subject might note that the idea of considering Gaussian approximations of the posterior for filtering based on quantized data with Gaussian noise dates back to [Curry 1970]. Also, the idea of quantizing the innovation seems to be first exploited in [Borkar 1995]<sup>7</sup> (cited in [Sukhavasi 2009a]).

The general algorithm presented in [You 2008] has its approximate performance dependent on  $I_q(0)$ , with  $I_q(0)$  being evaluated for the Gaussian distribution with variance  $\sigma_v^2 = 1$  (noise scale factor  $\delta = \sqrt{2}$ ). The performance of the algorithms is enhanced by maximizing  $I_q(0)$ . This is in accordance with the lower bound on the MSE studied above for the Wiener process model with symmetric noise,  $\text{MSE}_k \geq \frac{1}{J'_k}$ , which decreases with increasing  $I_q(0)$  ((2.37) shows that  $J'_k$  increases with  $I_q(0)$ ). This gives additional motivation for studying how to maximize  $I_q(0)$  w.r.t. the thresholds.

The assumption that the posterior is a Gaussian distribution for all  $k$  and all  $\sigma_w$  stated in [Curry 1970], [Ribeiro 2006c] and [You 2008] is a very rough approximation. For observing this, consider that the assumption that the prediction PDF  $p(x_k|i_{1:k-1})$  is Gaussian is correct. Then, from the update expression (2.11), we know that the posterior is proportional to the function  $\mathbb{P}(i_k|x_k)p(x_k|i_{1:k-1})$ . The probability  $\mathbb{P}(i_k|x_k)$  is a difference of CDF, which is a function that is approximately a rectangular window with slowly decreasing borders centered at the quantization interval for  $i_k$ . If the standard deviation of the prediction is large or has similar value of the equivalent width of  $\mathbb{P}(i_k|x_k)$  and the prediction distribution has a mean that is different of the quantization interval center, then it is easy to see that the resultant  $\mathbb{P}(i_k|x_k)p(x_k|i_{1:k-1})$  will be a skewed function, not similar at all to a Gaussian function. As an additional remark, we can see that differently from the continuous measurement case, where the measurement noise must be Gaussian for having a Gaussian posterior, the assumption of Gaussian noise does not help here, as the function  $\mathbb{P}(i_k|x_k)$  is not close to Gaussian even in the Gaussian case.

We will use the Gaussian assumption when  $\sigma_w$  is small and  $k$  tends to infinity. Under these assumptions and considering that we quantize the innovations, we will obtain an approximation of the asymptotically optimal estimator and its performance. To verify that the approximation is reasonable, we will compare the approximate asymptotic performance with the asymptotic BCRB.

### 2.5.3.1 Asymptotic estimator for a slowly varying parameter

As it was discussed above, it is reasonable to accept that the estimator MSE will converge to a constant  $\sigma_\infty^2$ . When the Wiener process increment standard deviation  $\sigma_w$  is small compared with the noise scale factor, the estimator has sufficient time for reducing the estimation variance before  $X_k$  changes significantly, thus it is also reasonable to state that  $\sigma_\infty^2$  is small. If

<sup>7</sup>In this case, the true innovation is quantized, *i.e.*, the innovation obtained by using the estimator based on the continuous measurements, this is different from the methods in [Ribeiro 2006c] and [You 2008], where the quantized innovation is the innovation obtained using the estimator based on the quantized measurements.

we assume that the previous posterior after some time is approximately Gaussian with mean  $\mathbb{E}(X_{k-1})$  and variance  $\sigma_\infty^2$ , then, as the prediction PDF is the convolution (2.10) of  $p(x_k|x_{k-1})$ , which is Gaussian, with  $p(x_k|i_{1:k-1})$  which is also Gaussian, we obtain that the prediction PDF conditioned on the past observations is Gaussian distributed with mean  $\mathbb{E}(X_{k-1}) + u_k$  and variance  $\sigma_\infty^2 + \sigma_w^2$ . For estimating the optimal  $X_k$  we must evaluate the conditional mean

$$\hat{X}_k = \int_{\mathbb{R}} x_k \frac{\mathbb{P}(i_k|x_k) p(x_k|i_{1:k-1})}{\int_{\mathbb{R}} \mathbb{P}(i_k|x'_k) p(x'_k|i_{1:k-1}) dx'_k} dx_k = \frac{\int_{\mathbb{R}} x_k \mathbb{P}(i_k|x_k) p(x_k|i_{1:k-1}) dx_k}{\int_{\mathbb{R}} \mathbb{P}(i_k|x'_k) p(x'_k|i_{1:k-1}) dx'_k}.$$

The numerator in the last term of the RHS can be seen as the prediction mean of the r.v.  $X_k \mathbb{P}(i_k|X_k)$  (the mean w.r.t.  $p(x_k|i_{1:k-1})$ ), under the assumption that  $\sigma_\infty$  is small, the factor  $\mathbb{P}(i_k|X_k)$ , which is given by (2.14)

$$\mathbb{P}(i_k|X_k) = \begin{cases} F\left(\tau'_{|i_k|} + \hat{X}_{k|k-1} - X_k\right) - F\left(\tau'_{|i_k-1|} + \hat{X}_{k|k-1} - X_k\right), & \text{if } i_k > 0, \\ F\left(-\tau'_{|i_k+1|} + \hat{X}_{k|k-1} - X_k\right) - F\left(-\tau'_{|i_k|} + \hat{X}_{k|k-1} - X_k\right), & \text{if } i_k < 0, \end{cases}$$

can be well approximated by a first order Taylor series expansion around  $\hat{X}_{k|k-1} - X_k = 0$

$$\begin{aligned} \mathbb{P}(i_k|X_k) &= \mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k} + \left(\hat{X}_{k|k-1} - X_k\right) f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} \\ &\quad + o\left(\hat{X}_{k|k-1} - X_k\right), \end{aligned} \quad (2.41)$$

where  $f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)$  is the first derivative of  $\mathbb{P}(i_k|X_k)$  w.r.t. the prediction error  $\hat{X}_{k|k-1} - X_k$ . It can be written as a function of the noise PDF  $f$

$$\begin{aligned} f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right) &= \\ &= \begin{cases} f\left(\tau'_{|i_k|} + \hat{X}_{k|k-1} - X_k\right) - f\left(\tau'_{|i_k-1|} + \hat{X}_{k|k-1} - X_k\right), & \text{if } i_k > 0, \\ f\left(-\tau'_{|i_k+1|} + \hat{X}_{k|k-1} - X_k\right) - f\left(-\tau'_{|i_k|} + \hat{X}_{k|k-1} - X_k\right), & \text{if } i_k < 0. \end{cases} \end{aligned} \quad (2.42)$$

Note that when  $\hat{X}_{k|k-1} = X_k$  the function  $f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)$  depends only on  $i_k$ . Using (2.41), the numerator in the estimator expression is then the prediction mean of

$$\begin{aligned} X_k \mathbb{P}(i_k|X_k) &= X_k \mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k} + \left(X_k \hat{X}_{k|k-1} - X_k^2\right) f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} \\ &\quad + o\left(X_k \hat{X}_{k|k-1} - X_k^2\right). \end{aligned}$$

The prediction mean of  $X_k$  is the prediction  $\hat{X}_{k|k-1}$ , while  $\hat{X}_{k|k-1}$  is simply a constant for the evaluation of this mean. Thus, using linearity and the fact that

$$\mathbb{E}_{X_k|i_{1:k-1}}^2(X_k) - \mathbb{E}_{X_k|i_{1:k-1}}(X_k^2) = -\text{Var}_{X_k|i_{1:k-1}}(X_k) = -(\sigma_\infty^2 + \sigma_w^2),$$

we have

$$\begin{aligned} \int_{\mathbb{R}} x_k \mathbb{P}(i_k|x_k) p(x_k|i_{1:k-1}) dx_k &= \hat{X}_{k|k-1} \mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k} + \\ &\quad - (\sigma_\infty^2 + \sigma_w^2) f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} + o(\sigma_\infty^2 + \sigma_w^2). \end{aligned} \quad (2.43)$$

To obtain the denominator in the estimation expression, we can use a similar procedure. Note that now the prediction expectation is evaluated for the r.v.  $\mathbb{P}(i_k|X_k)$  instead of  $X_k\mathbb{P}(i_k|X_k)$ . We will use the second order Taylor expansion of  $\mathbb{P}(i_k|X_k)$  in this case:

$$\begin{aligned} \mathbb{P}(i_k|X_k) &= \mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k} + \left(\hat{X}_{k|k-1} - X_k\right) f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} + \\ &+ \frac{\left(\hat{X}_{k|k-1} - X_k\right)^2}{2} f'_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} + \circ\left[\left(\hat{X}_{k|k-1} - X_k\right)^2\right], \end{aligned} \quad (2.44)$$

where  $f'_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)$  is the second derivative of  $\mathbb{P}(i_k|X_k)$  w.r.t. the prediction error. By differentiating  $f_d$  in (2.42), we can observe that, for  $\hat{X}_{k|k-1} = X_k$ , this function also depends only on  $i_k$ . The mean of the first term above is the constant  $\mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k}$ . For the second term  $f_d$  is a constant and the mean of the prediction is zero, as the optimal predictor is unbiased [Jazwinski 1970, p. 150]. The third and last terms depend on the prediction mean of  $\left(\hat{X}_{k|k-1} - X_k\right)^2$ , which is equal to the prediction variance  $\sigma_\infty^2 + \sigma_w^2$ , also due to the unbiasedness of the optimal predictor. This gives the following

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{P}(i_k|x'_k) p(x'_k|i_{1:k-1}) dx'_k &= \mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k} + \\ &+ \frac{(\sigma_\infty^2 + \sigma_w^2)}{2} f'_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} + \circ(\sigma_\infty^2 + \sigma_w^2). \end{aligned} \quad (2.45)$$

Dividing the RHS of the expressions (2.43) and (2.45), we have an expression for the estimator

$$\hat{X}_k = \frac{\hat{X}_{k|k-1} \mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k} - (\sigma_\infty^2 + \sigma_w^2) f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} + \circ(\sigma_\infty^2 + \sigma_w^2)}{\mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k} + \frac{(\sigma_\infty^2 + \sigma_w^2)}{2} f'_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k} + \circ(\sigma_\infty^2 + \sigma_w^2)}.$$

Dividing the numerator and denominator by  $\mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}}$ , we get

$$\hat{X}_k = \frac{\hat{X}_{k|k-1} - (\sigma_\infty^2 + \sigma_w^2) \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k)\Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k}} + \circ(\sigma_\infty^2 + \sigma_w^2)}{1 + \frac{(\sigma_\infty^2 + \sigma_w^2)}{2} \frac{f'_d(i_k, \hat{X}_{k|k-1}, X_k)\Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k}} + \circ(\sigma_\infty^2 + \sigma_w^2)}. \quad (2.46)$$

If  $\left| \frac{f'_d(i_k, \hat{X}_{k|k-1}, X_k)\Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k}} \right|$  is bounded and  $\sigma_\infty^2 + \sigma_w^2 \ll 1$ , the denominator is approximately one and we can approximate the optimal estimator by<sup>8</sup>

$$\hat{X}_k \approx \hat{X}_{k|k-1} - (\sigma_\infty^2 + \sigma_w^2) \frac{f_d\left(i_k, \hat{X}_{k|k-1}, X_k\right)\Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k)|_{\hat{X}_{k|k-1}=X_k}}. \quad (2.47)$$

<sup>8</sup>Note that a first order Taylor series of  $\frac{1}{1+x}$  around  $x = 0$  would produce a more precise approximation, but this would generate a more complicated algorithm for the performance analysis.

### 2.5.3.2 Performance of the asymptotic estimator

Now, we need to calculate the asymptotic MSE of this estimator, which is  $\sigma_\infty^2$ . The idea here to start is to rewrite the asymptotic prediction error as a sum of the estimation error plus a function of the observations. Using (2.47) we have

$$\hat{X}_k - X_k = \hat{X}_{k|k-1} - X_k - (\sigma_\infty^2 + \sigma_w^2) \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}},$$

subtracting the term with  $f_d$  from both sides, we have

$$\hat{X}_k - X_k + (\sigma_\infty^2 + \sigma_w^2) \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} = \hat{X}_{k|k-1} - X_k.$$

Squaring and taking the expectation gives

$$\begin{aligned} \mathbb{E} \left[ (\hat{X}_k - X_k)^2 \right] + 2\mathbb{E} \left[ (\hat{X}_k - X_k) (\sigma_\infty^2 + \sigma_w^2) \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} \right] + \\ + \mathbb{E} \left[ (\sigma_\infty^2 + \sigma_w^2)^2 \frac{f_d^2(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}^2(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} \right] = \mathbb{E} \left[ (\hat{X}_{k|k-1} - X_k)^2 \right]. \end{aligned}$$

The first term is the asymptotic squared error  $\sigma_\infty^2$ . The second term is the expectation of the estimation error multiplied by a function of the measurement  $i_k$ , for small  $\sigma_\infty^2 + \sigma_w^2$  the estimation procedure is optimal (it minimizes the MSE), thus this expectation equals zero [Rhodes 1971]<sup>9</sup>. The constant  $(\sigma_\infty^2 + \sigma_w^2)^2$  can leave the expectation and the term on the RHS is the prediction error  $\sigma_\infty^2 + \sigma_w^2$ . Therefore, we have

$$\sigma_\infty^2 + (\sigma_\infty^2 + \sigma_w^2)^2 \mathbb{E} \left[ \frac{f_d^2(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}^2(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} \right] = \sigma_\infty^2 + \sigma_w^2. \quad (2.48)$$

The expectation that still needs to be evaluated is an expectation under the marginal probability measure of  $i_k$ ,  $\mathbb{P}(i_k)$ . This probability measure can be evaluated by marginalizing on the prediction error  $\varepsilon_k = \hat{X}_k - X_k$

$$\mathbb{P}(i_k) = \int_{\mathbb{R}} \mathbb{P}(i_k|x_k) p(\varepsilon_k) d\varepsilon_k.$$

Remember that in the quantized innovation scheme  $\mathbb{P}(i_k|x_k)$  is a function of  $i_k$  and  $\varepsilon_k$ . The marginal can be also observed as the mean of  $\mathbb{P}(i_k|X_k)$  evaluated w.r.t. the distribution of

<sup>9</sup>This is a more general form of the well-known orthogonal projection theorem.

$\varepsilon_k$ . For evaluating the mean, we can again use a second order Taylor series expansion around  $\varepsilon_k = 0$ . This will lead to the same expression as in (2.45). Then, the remaining expectation is

$$\begin{aligned} \mathbb{E} \left[ \frac{f_d^2(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}^2(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} \right] &= \left\{ \sum_{i_k \in \mathcal{I}} \frac{f_d^2(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} \right\} + \\ &+ (\sigma_\infty^2 + \sigma_w^2) \sum_{i_k \in \mathcal{I}} \frac{f_d^2(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k} f_d'(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}^2(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} + \\ &+ \circ (\sigma_\infty^2 + \sigma_w^2). \end{aligned}$$

The first term on the RHS can be identified as the FI  $I_q(0)$ . Considering that the sum in the second term on the RHS is bounded, then, after multiplying by  $(\sigma_\infty^2 + \sigma_w^2)^2$ , the second term is multiplied by  $(\sigma_\infty^2 + \sigma_w^2)^3$  which is a  $\circ [(\sigma_\infty^2 + \sigma_w^2)^2]$  term. This leads to

$$(\sigma_\infty^2 + \sigma_w^2)^2 \mathbb{E} \left[ \frac{f_d^2(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}^2(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} \right] = (\sigma_\infty^2 + \sigma_w^2)^2 I_q(0) + \circ [(\sigma_\infty^2 + \sigma_w^2)^2].$$

Using the expression above in (2.48), we obtain

$$\sigma_\infty^2 + (\sigma_\infty^2 + \sigma_w^2)^2 I_q(0) + \circ [(\sigma_\infty^2 + \sigma_w^2)^2] = \sigma_\infty^2 + \sigma_w^2,$$

or equivalently

$$(\sigma_\infty^2 + \sigma_w^2)^2 + \circ \left[ \frac{(\sigma_\infty^2 + \sigma_w^2)^2}{I_q(0)} \right] = \frac{\sigma_w^2}{I_q(0)}.$$

If  $\sigma_w^2$  is small enough so that the  $\circ$  term is negligible, we can obtain the following approximation for  $\sigma_\infty^2$ :

$$\sigma_\infty^2 \approx \frac{\sigma_w}{\sqrt{I_q(0)}} - \sigma_w^2. \quad (2.49)$$

Considering that  $\sigma_w$  is small compared with  $\sqrt{I_q(0)}$  and with one, we have a rough approximation for the asymptotic performance

$$\sigma_\infty^2 \approx \frac{\sigma_w}{\sqrt{I_q(0)}}. \quad (2.50)$$

Finally, replacing  $\sigma_\infty^2$  from (2.49) in the approximate expression for  $\hat{X}_k$  (2.47), the following is obtained:

$$\begin{aligned} \hat{X}_k &\approx \hat{X}_{k|k-1} - \frac{\sigma_w}{\sqrt{I_q(0)}} \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}} = \\ &= \hat{X}_{k-1} + u_k - \frac{\sigma_w}{\sqrt{I_q(0)}} \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}. \end{aligned} \quad (2.51)$$

A few remarks are important here. First, in a similar way as it happened for the adaptive estimator of a constant parameter, the asymptotic estimation procedure is very simple, it is a correction on the last estimate which depends on the observation through

$$\frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) |_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k | X_k) |_{\hat{X}_{k|k-1}=X_k}},$$

a function of  $i_k$  only. This means that the corrections can be stored in a table. Second, the correction gain now is even simpler than in the constant parameter case, it is a constant. Third, the rough approximation for  $\sigma_\infty^2$  (2.50) agrees with the intuition on estimation performance, if  $\sigma_w$  increases, the MSE increases, as the estimator has less effective samples to estimate before the parameter changes significantly. If  $I_q(0)$  increases, which is equivalent to say that the noise level is reduced and/or that the quantizer resolution is increased, the MSE decreases as the statistical information given by each sample is reduced.

### 2.5.3.3 Asymptotic lower bound on the BCRB for a slowly varying parameter

To check if the asymptotic estimator above is indeed close to optimal, we can compare its estimation performance with the asymptotic MSE lower bound, which is given by (2.40):

$$\text{MSE}_\infty \geq \frac{2}{I_q(0) + \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}}.$$

Comparing with (2.50) must be done for small  $\sigma_w$ . For evaluating the RHS above in this case, we can multiply its numerator and its denominator by  $\frac{\sigma_w}{2\sqrt{I_q(0)}}$ . This will lead to

$$\frac{2}{I_q(0) + \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}} = \frac{\frac{\sigma_w}{\sqrt{I_q(0)}}}{\frac{I_q(0)\sigma_w}{2\sqrt{I_q(0)}} + \sqrt{\frac{I_q(0)\sigma_w^2}{4} + 1}}.$$

Using the expansion around  $x = 0$ ,  $\sqrt{1+x} = 1 + \frac{x}{2} + \circ(x)$ , on the square root above gives

$$\frac{2}{I_q(0) + \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}} = \frac{\frac{\sigma_w}{\sqrt{I_q(0)}}}{\frac{I_q(0)\sigma_w}{2\sqrt{I_q(0)}} + 1 + \frac{I_q(0)\sigma_w^2}{8} + \circ(\sigma_w^2)},$$

where we used the fact that  $\sigma_w^2$  is small compared with  $I_q(0)$  for making the  $I_q(0)$  to disappear from the  $\circ$  term. Note that this was also supposed to get the rough approximation (2.50) above. We can use again an expansion around zero. Now, we use  $\frac{1}{1+x} = 1 - x + \circ(x)$ . Supposing, additionally that  $\sigma_w$  is small compared with  $\frac{I_q(0)}{\sqrt{I_q(0)}}$ , we can use a  $\circ$  term depending only on  $\sigma_w$ . Thus, we obtain

$$\frac{2}{I_q(0) + \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}} = \frac{\sigma_w}{\sqrt{I_q(0)}} \left[ 1 - \frac{I_q(0)\sigma_w}{2\sqrt{I_q(0)}} + \circ(\sigma_w) \right].$$

The squared terms can be assimilated to  $\circ(\sigma_w)$  leading finally to

$$EQM_\infty \geq \frac{2}{I_q(0) + \sqrt{I_q^2(0) + \frac{4I_q(0)}{\sigma_w^2}}} = \frac{\sigma_w}{\sqrt{I_q(0)}} + \circ(\sigma_w),$$

which for small  $\sigma_w$  is exactly the same as the rough approximation of the asymptotic estimator performance. Consequently, we can say that the asymptotic estimator obtained above is optimal, as in this specific case, it attains the lower bound.

As in the previous section, where the adaptive MLE scheme was shown to have a simple recursive form asymptotically, a question arise:

- can the asymptotic estimator (2.47) for slowly varying  $X_k$  converge when we use it with an arbitrary (not necessarily small) initial error?

The answer for this question will be given in Ch. 3.

## 2.6 Chapter summary and directions

We sum up the main points of this chapter:

- instead of considering that the parameter is constant, we assumed that the parameter can vary in time, more specifically, following a Wiener process model. We saw that, in general, the optimal estimator can be obtained by evaluating the mean of the parameter conditioned on the past and present quantized measurements. Thus, the core of the problem was observed to be the evaluation of the posterior PDF (PDF of  $X_k$  conditioned on  $i_{1:k}$ ). For a Markov process  $X_k$ , which is the case for a Wiener process model, the posterior can be evaluated in a recursive way, first by obtaining a prediction PDF using the posterior at time  $k - 1$  and the evolution model, then by updating the prediction PDF to the posterior at time  $k$ , incorporating the new measurement  $i_k$ .
- The integrals involved in the recursive expressions are complicated to be evaluated analytically, so we must resort to numerical algorithms for solving them. One way of doing this is to apply Monte Carlo integration. This leads to a PF solution. The PF solution is a recursive simplified form of Monte Carlo integration applied to the filtering problem with an additional resampling step. The performance of the optimal estimator could also be obtained using Monte Carlo integration, but it would be very difficult and time consuming to study the effects of the system parameters (noise level, Wiener process increments variance and quantizer resolution) using the Monte Carlo results. Therefore, we considered a simpler solution by using a bound on the MSE for which we can have a simple analytical expression.
- In our case, we used the BCRB, which is the inverse of the Bayesian information. The BI for the Wiener process  $X_k$  can be evaluated recursively. From its recursive expression, we could see that the BI and consequently the bound were affected by the quantization through a  $\mathbb{E}[I_q(\varepsilon_k)]$  term, where  $\varepsilon_k$  is the difference between the central threshold and the parameter. If  $\mathbb{E}[I_q(\varepsilon_k)]$  is increased, then the bound decreases, if it decreases, the bound increases. For commonly used noise models,  $I_q(\varepsilon_k)$  is maximum at  $\varepsilon_k = 0$ , therefore, a practical lower bound can be obtained by using  $I_q(0)$  instead of  $\mathbb{E}[I_q(\varepsilon_k)]$ .
- If we accept that the bound is tight enough to mimic the behavior of the MSE, another consequence of the dependence of the bound on  $\mathbb{E}[I_q(\varepsilon_k)]$  and the fact that  $I_q(\varepsilon_k)$  is large close to  $\varepsilon_k$  is that the central threshold must be placed as close as possible to the true parameter. This can be done in an approximate way, by setting the central threshold to the prediction of  $X_k$  based on the past measurements. Thus, a good estimation procedure might be based on the quantized innovation. In this case, it is expected that the estimation performance will be closer to the bound, when compared with a quantizer with arbitrary central threshold.
- When  $\sigma_w$  is small and  $k$  tends to infinity, the optimal estimator can be approximated by a low complexity recursive expression, with its MSE attaining the BCRB and given by  $\frac{\sigma_w}{\sqrt{I_q(0)}}$ . This shows one more time, the importance of studying the maximization of  $I_q(0)$  w.r.t. the quantization threshold variations. As stated before, the asymptotic analysis

of this problem will be done in Part II. The simplicity of the asymptotic estimator when  $X_k$  varies slowly will be used as a motivation to study in more detail recursive algorithms of the type - prediction + correction based on  $i_k$ . This will be done in Ch. 3.

- A generalization of the signal model used here can be obtained by considering that the dynamical parameter is a vector  $\mathbf{X}_k$  with size  $N$  and that it obeys a linear Gaussian model of the type

$$\mathbf{X}_k = \Phi_k \mathbf{X}_{k-1} + \mathbf{W}_k,$$

where  $\Phi_k$  is a  $N \times N$  matrix and  $\mathbf{W}_k$  is a sequence of independent Gaussian vectors. The continuous measurement is a vector  $\mathbf{Y}_k$  with dimension  $M$

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_k,$$

where  $\mathbf{H}_k$  is a  $M \times N$  matrix and  $\mathbf{V}_k$  is a sequence of independent vectors. Quantization can be done also scalarly, but we might consider two possibilities for the quantization of each  $Y_k$ , scalar quantization of each dimension or vector quantization of the entire vector.

A direct application of the estimation problem with this model is the control of linear systems under rate constraints. We will not go further in this direction in this thesis, but we will keep this generalized version of the problem for future work.



# Adaptive quantizers for estimation

---

As we saw in the previous chapters, to obtain good estimation performance, the quantizer dynamic might be adaptively set around the true parameter to be estimated. We also saw that the asymptotically optimal estimator in the constant parameter case or in the slowly varying parameter case has a simple recursive form. We asked at the end of each chapter:

- can the asymptotic estimator based on binary measurements converge when we use its simplified form (the low complexity equivalent) with an arbitrary initial error (not necessarily small)?
- Can we extend this low complexity recursive procedure to the case  $N_I > 2$ ?
- Can the asymptotic estimator for slowly varying  $X_k$  converge when we use its simplified form (the low complexity equivalent) with an arbitrary initial error (not necessarily small)?

In this chapter we will answer these questions. For doing so, we will impose the estimation algorithm to have a general recursive form that includes the asymptotically optimal estimators as special cases.

We will start the chapter with a brief review of the signal models that will be used (constant, Wiener process without and with drift) and with the definition of the quantizer to be used. Then, we will define the estimation algorithm form and we will study its performance for the signal models defined previously in terms of the mean error and of the MSE. Based on the performance analysis, we will obtain the optimal estimator parameters and the corresponding optimal performance. As in related work [Papadopoulos 2001], the optimal performance will be used to obtain a measurement of performance loss due to quantization. This loss will be evaluated for each signal model by using the corresponding optimal performance for estimation based on continuous measurements. The performance results will be verified through simulation.

We will also propose extensions of the adaptive algorithm in the following cases:

- quantized measurements from a sensor are used for estimating a constant parameter, but in this case, the noise scale parameter is considered to be unknown.
- Multiple sensors and a fusion center are used to estimate a constant parameter. The sensors can send only quantized measurements to the fusion center, while the fusion center can broadcast continuous values to the sensors.

In each of these cases we will follow a similar procedure. We will define the problem and the estimation algorithm to be used. Then, we will obtain the optimal estimator parameters and the corresponding optimal performance. Simulation will be used to check the validity of the results.

At the end of the chapter, we will summarize the main results of the chapter and we will give some directions for future work.

#### Contributions presented in this chapter:

- *Design and analysis of an adaptive estimation algorithm based on multibit quantized noisy measurements.* This differs from [Li 2007] and [Fang 2008], where only binary quantization is treated.
- *Explicit performance analysis for tracking of a slowly varying parameter.* Differently from [Papadopoulos 2001, Ribeiro 2006a, Li 2007, Fang 2008], where the parameter is set to be constant and all subsequent analysis is based on this hypothesis. Even if tracking is treated in a more general way in [Ribeiro 2006c] and [You 2008], we do not state assumptions on noise Gaussianity. Note that the assumption that the parameter varies slowly seems more restrictive than the parameter models considered in [Ribeiro 2006c] and [You 2008], actually, the slowly varying assumption is hidden in the performance evaluation for the binary case given in [Ribeiro 2006c], where it is shown that the performance of the proposed filter reaches the equivalent continuous when the sampling time tends to zero.
- *Low complexity algorithms.* The algorithms proposed here are based on simple recursive techniques that have lower complexity than the methods proposed in [Li 2007] and [Fang 2008].
- *Joint location and scale adaptive estimator.* The algorithm that we propose is an extension of the location estimation problem. This extension is discussed in [Ribeiro 2006b] but only for fixed quantization thresholds.
- *Fusion center approach.* This approach can be seen as a multisensor, multibit, low complexity alternative to the adaptive techniques presented in [Li 2007] and [Fang 2008] and also as an adaptive alternative for the optimal threshold distribution approach given in [Ribeiro 2006a], where a prior distribution on the parameter is needed.

---

**Contents**

---

<b>3.1</b>	<b>Parameter model and measurement model</b>	<b>108</b>
3.1.1	Parameter model	108
3.1.2	Noise model	108
3.1.3	Adjustable quantizer model	109
<b>3.2</b>	<b>General estimation algorithm</b>	<b>111</b>
<b>3.3</b>	<b>Estimation performance</b>	<b>113</b>
3.3.1	Mean ordinary differential equation	113
3.3.2	Asymptotic MSE	121
<b>3.4</b>	<b>Optimal algorithm parameters and performance</b>	<b>125</b>
3.4.1	Optimal algorithm parameters	125
3.4.2	Algorithm performance for optimal gain and coefficients	128
<b>3.5</b>	<b>Simulations</b>	<b>135</b>
3.5.1	General considerations	135
3.5.2	Theoretical performance loss due to quantization	137
3.5.3	Simulated loss	138
3.5.4	Comparison with the high complexity algorithms	143
3.5.5	Discussion on the results	147
<b>3.6</b>	<b>Adaptive quantizers for estimation: extensions</b>	<b>149</b>
3.6.1	Joint estimation of location and scale parameters	149
3.6.2	Fusion center approach with multiple sensors	155
<b>3.7</b>	<b>Chapter summary and directions</b>	<b>164</b>

---

### 3.1 Parameter model and measurement model

#### 3.1.1 Parameter model

We will join the constant and varying models by using the dynamic model (2.1)

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

where  $\{W_k, k = 1, 2, \dots\}$  is a sequence of independent Gaussian r.v. (also independent of  $X_n$ , for  $n < k$ ) whose means form a deterministic sequence  $\{u_k, k = 1, 2, \dots\}$  and its standard deviation is  $\sigma_w$ :

$$W_k \sim \mathcal{N}(u_k, \sigma_w^2).$$

Symbol  $\sim$  means "distributed according to" and  $\mathcal{N}$  is the symbol for the Gaussian distribution.

Differently from what was considered previously, the sequence  $u_k$  will be considered to be a known or unknown constant  $u$  and we will assume that it has small value. We will also assume that  $\sigma_w$  is a known, small constant. "Small" in both cases means that these constants are small when compared with the noise scale parameter. In the Gaussian noise case, this is equivalent to say that they are small when compared with the noise standard deviation.

The fact that we use a constant  $u$  instead of the varying  $u_k$  will allow to have asymptotic performance results. In practice, all the results that will be presented will be valid for varying  $u_k$ , as long as the sequence  $u_k$  is small and slowly varying.

The model above is a compact form to describe the three parameter models that are studied in this thesis:

- *constant*: by taking  $u = \sigma_w = 0$  and  $X_0 = x$ , we have the constant parameter model.
- *Wiener process*: if  $u = 0$ , (small) nonzero  $\sigma_w$  and Gaussian  $X_0$  with unknown mean and variance, then  $X_k$  is a (slowly) varying Wiener process.
- *Wiener process with drift*: in this case  $u$  and  $\sigma_w$  are non zero (and with small amplitudes).

#### 3.1.2 Noise model

The continuous amplitude measurement is again given by the additive model

$$Y_k = X_k + V_k,$$

where the noise r.v. sequence  $V_k$  respects the assumptions considered previously:

- the sequence is i.i.d..
- AN1 (p. 34) – The marginal noise CDF denoted  $F(v)$  accepts a PDF denoted  $f(v)$ .
- AN2 (p. 34) –  $f(v)$  is a strictly positive even function that strictly decreases w.r.t.  $|v|$ .

An additional assumption will be considered on the noise CDF.

**Assumption (on the noise distribution):**

**AN3**  $F$  is locally Lipschitz continuous.

A function  $F(v)$  is Lipschitz continuous in an interval  $\mathcal{V}$ , if for every two points  $v_1$  and  $v_2$  in  $\mathcal{V}$  there exists a constant  $L$  such that

$$|F(v_1) - F(v_2)| \leq L |v_1 - v_2|,$$

the function is locally Lipschitz continuous if for every  $v \in \mathbb{R}$ , we can find an interval  $\mathcal{V}'$  containing  $v$  such that the function is Lipschitz continuous.

This assumption is required by the method of analysis that will be used to assess the performance of the proposed algorithm. Most noise CDF considered in practice are Lipschitz continuous, thus this assumption is generally satisfied.

### 3.1.3 Adjustable quantizer model

We saw in Ch. 1 and 2 that the quantizer central threshold must be dynamically updated to obtain a good estimation performance. We will make explicit this feature by imposing the quantizer to have an adjustable offset  $b_k$ . For adjusting the amplitude of the quantizer input, we can also consider that after offsetting the input, we apply an adjustable gain  $\frac{1}{\Delta_k}$ . The quantized measurements at the output of the adjustable quantizer are given by

$$i_k = Q \left( \frac{Y_k - b_k}{\Delta_k} \right). \quad (3.1)$$

By considering dynamic input offset and gain, we can fix the quantizer to have a static structure with a central threshold that now can be set to zero. Thus, the quantizer thresholds are equal to the threshold variations. This modifies assumption AQ2 (p. 37).

**Assumption (on the quantizer):**

**AQ2'** The quantizer is symmetric around the central threshold which is equal to zero. This means that the vector of thresholds  $\boldsymbol{\tau}$  is given by the vector of threshold variations

$$\boldsymbol{\tau} = \boldsymbol{\tau}' = \left[ -\tau'_{\frac{N_I}{2}} \cdots -\tau'_1 \quad 0 \quad +\tau'_1 \cdots +\tau'_{\frac{N_I}{2}} \right]^\top,$$

where the threshold variations  $\tau'_i$  form an increasing sequence.

The adjustable quantizer output is given by

$$i_k = Q \left( \frac{Y_k - b_k}{\Delta_k} \right) = i \operatorname{sign}(Y_k - b_k), \quad \text{for } \frac{|Y_k - b_k|}{\Delta_k} \in [\tau'_{i-1}, \tau'_i). \quad (3.2)$$

A scheme representing the adjustable quantizer is given in Fig. 3.1. Note that even if the quantizer is not uniform (with constant step-length between thresholds), it can be implemented using a uniform quantizer with a compander approach [Gersho 1992].

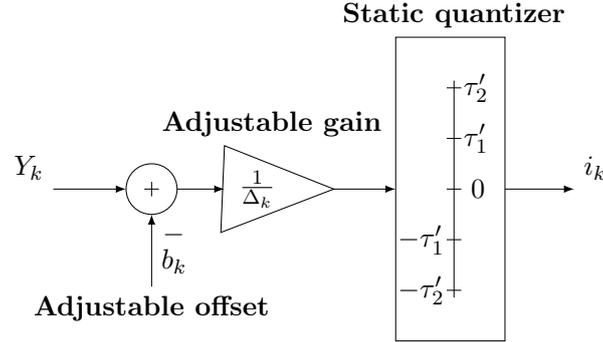


Figure 3.1: Scheme representing the adjustable quantizer. The offset and gain can be adjusted dynamically, while the quantizer thresholds are fixed.

Based on the quantizer outputs, the main objective is to estimate  $X_k$ . A secondary objective is to adjust the parameters  $b_k$  and  $\Delta_k$  to enhance estimation performance. As the estimate  $\hat{X}_k$  of  $X_k$  will be possibly used in real time applications, it might be interesting to estimate it online. Therefore, we are again interested in solving problems (a) and (b), the main difference is that now we want to solve (a) for each time index  $k$ .

It was observed in the previous chapters that

- when estimating a constant, we can place the central threshold in the last estimate to have an asymptotically optimal algorithm.
- When estimating a Wiener process, we can place the central threshold at the prediction. For Wiener process without drift the prediction is exactly  $\hat{X}_{k-1}$  and for Wiener process with drift the prediction is  $\hat{X}_{k-1} + u_k$ .

Based on these observations and for simplification purposes, we will set for all cases  $b_k = \hat{X}_{k-1}$ . Also to simplify, we will consider that the gain is set to be a constant. For the algorithm presented later, the fact that the offset is set to  $\hat{X}_{k-1}$  will have as a consequence asymptotically optimal parameters that do not depend on the mean of  $X_k$ , thus simplifying the analysis.

Some remarks here are important:

- We will see that imposing the use of  $b_k = \hat{X}_{k-1}$ , instead of using the prediction, will make the algorithm parameters and the performance to be different for Wiener process with and without drift.
- If we use the prediction, instead of the last estimate, for setting the quantizer offset and for estimating  $X_k$ , then all the results that we will obtain for a Wiener process without drift will be valid also for the process with drift.
- In the special cases where the optimal central threshold is not the median of the continuous amplitude measurement, we can evaluate the optimal quantizer offset  $\varepsilon^*$  w.r.t. the true parameter (the point of minimum in the "w" shaped CRB curves) and then add this value to the offset of the adaptive quantizer  $b_k = \hat{X}_{k-1} + \varepsilon^*$ .

- As most performance results will be given asymptotically, the simplification brought by using a constant gain  $\frac{1}{\Delta}$  can still be partially achieved if we constrain this gain to be constant after some time or to achieve asymptotically a constant value. In this case, the analysis of error convergence will have to take into account that the measurement system varies in time and we must be able also to evaluate its asymptotic value.
- The gain  $\frac{1}{\Delta}$  will be again considered to be variable further in the chapter, where we will estimate jointly a constant  $X_k$  and the scale parameter of the noise. In this case, the gain will not only be variable, but it will also depend on the measurements.

The general scheme for the estimation of  $X_k$  is depicted in Fig. 3.2 and the main objective will be to find the algorithm that will be placed in the block named **Update**.

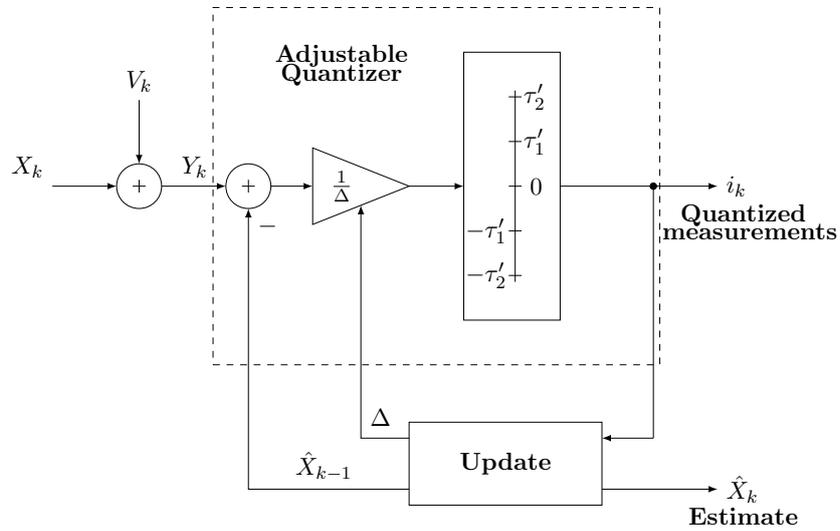


Figure 3.2: Block representation of the estimation scheme. The estimation algorithm and the procedures to set the offset and the gain are represented by the **Update** block.

### 3.2 General estimation algorithm

At the end of Ch. 1, we saw that the estimator in the adaptive binary quantization scheme based on the MLE is asymptotically given by (1.87)

$$\hat{X}_k = \hat{X}_{k-1} + \frac{i_k}{2kf(0)},$$

whereas at the end of Ch. 2, we saw that the asymptotic expression for the optimal estimator of a slowly varying Wiener process is (2.51)

$$\hat{X}_k \approx \hat{X}_{k|k-1} - \frac{\sigma_w}{\sqrt{I_q(0)}} \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}.$$

Both asymptotic estimators have low complexity and both are special cases of the following adaptive algorithm:

$$\hat{X}_k = \hat{X}_{k-1} + \gamma_k \eta \left[ Q \left( \frac{Y_k - \hat{X}_{k-1}}{\Delta} \right) \right], \quad (3.3)$$

where,  $\gamma_k$  is a sequence of positive real gains and  $\eta[\cdot]$  is a mapping from  $\mathcal{I}$  to  $\mathbb{R}$

$$\begin{aligned} \eta : \mathcal{I} &\rightarrow \mathbb{R} \\ j &\rightarrow \eta_j, \end{aligned} \quad (3.4)$$

which is characterized by the sequence of  $N_I$  coefficients  $\left\{ \eta_{-\frac{N_I}{2}}, \dots, \eta_{-1}, \eta_1, \dots, \eta_{\frac{N_I}{2}} \right\}$ . Notice that the coefficients  $\eta[\cdot]$  can be seen as the "estimation equivalent" of the output quantization levels used in standard quantization theory.

Even if nothing guarantees that the algorithm (3.3) is optimal for finite time, the fact that it can be equivalent asymptotically to the optimal estimator and that it has low complexity are strong motivations for using it. Other more intuitive motivations are the following:

- similarly to the binary grid method proposed by [Fang 2008], for a slowly varying or constant parameter, we can choose the coefficients  $\eta[\cdot]$  in a way that the algorithm will tend to be around true parameter at least in the mean.
- 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 (3.3). It will be shown later that for the optimal choice of  $\eta_i$ , algorithm (3.3) is equivalent to a stochastic gradient ascent method to maximize the log-likelihood.
- To estimate a Wiener process, an approximate choice of estimator is a Kalman filter like method based on the quantized innovation, which is also (3.3).

Due to the symmetry of the problem for commonly used noise models, 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  $i_k$ . This symmetry will be useful later for simplification purposes and we will add it to the other assumptions.

**Assumption (on the quantizer output levels):**

**AQ3** The quantizer output levels have odd symmetry w.r.t.  $i$ :

$$\eta_i = -\eta_{-i}, \quad (3.5)$$

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

In the special cases where the threshold must be placed asymmetrically and we put an additional constant value in the quantizer offset ( $\varepsilon^*$ ), the assumption above may lead to an asymptotic estimation bias. For observing this, consider that the quantization offset is already at the parameter. Then, the mean of the correction  $\eta[i_k]$  will be zero, as the distribution of the  $i_k$  is even and  $\eta_i$  is odd. Thus the algorithm is in a mean equilibrium point. As the offset is

placed  $\varepsilon^*$  away from the estimate, the mean of the estimate has an equilibrium point that is different from the true parameter.

The non differentiable non linearity  $Q$  in (3.3) makes it difficult to be analyzed. Fortunately, an analysis based on mean approximations was developed in [Benveniste 1990] for a wide class of adaptive algorithms. Within this framework, the function  $\eta$  can be a general nonlinear non-differentiable function of  $Y_k$  and  $\hat{X}_k$  and it is shown that the gains  $\gamma_k$  that optimize the estimation of  $X_k$  can be chosen as follows:

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

Notice that the gains for the constant and Wiener process models given above have the same form of the asymptotically optimal gains found in Ch. 1 and Ch. 2. The only difference is the gain proportional to  $u^{\frac{2}{3}}$  in the case with drift, which reflects the choice of using  $\hat{X}_{k-1}$  in the place of the prediction.

In the following sections we will consider the gains given above for the algorithm (3.3) and we will apply the general analysis presented in [Benveniste 1990] to obtain its performance.

### 3.3 Estimation performance

To obtain the estimation performance, the analysis is separated in

- *the analysis of the estimator mean.* This gives a rough approximation of the estimator behavior. With this information we can see if the estimator converges in the mean and we can also characterize its bias.
- *The analysis of the estimation variance.* This analysis will give the details on the fluctuation around the mean and it will be obtained, in most cases, asymptotically.

#### 3.3.1 Mean ordinary differential equation

The core of the analysis that we use here and that is presented in a general setting in [Benveniste 1990] is to approximate the mean  $\mathbb{E}(\hat{X}_k)$  by  $\hat{x}(t_k)$ , where  $\hat{x}(t)$  is the solution of the **ordinary differential equation (ODE)**

$$\frac{d\hat{x}}{dt} = h(\hat{x}). \quad (3.6)$$

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

$$h(\hat{x}) = \mathbb{E} \left\{ \eta \left( Q \left( \frac{x - \hat{x} + V}{\Delta} \right) \right) \right\}, \quad (3.7)$$

where the expectation is evaluated w.r.t. the distribution of  $V$ , which is the noise marginal distribution.

A simple heuristic to obtain the approximation is the following: first we rewrite (3.3) as

$$\frac{\hat{X}_k - \hat{X}_{k-1}}{\gamma_k} = \eta \left[ Q \left( \frac{X_k - \hat{X}_{k-1} + V_k}{\Delta} \right) \right],$$

then we consider that the parameter is approximately a constant  $X_k = x$  and that  $\hat{X}_{k-1}$  on the RHS can be approximated by the mean at time  $k$ , *i.e.*  $\hat{X}_{k-1} = \hat{x}$ . Evaluating the expectation on both sides

$$\frac{\mathbb{E}(\hat{X}_k) - \mathbb{E}(\hat{X}_{k-1})}{\gamma_k} = \mathbb{E} \left\{ \eta \left[ Q \left( \frac{x - \hat{x} + V_k}{\Delta} \right) \right] \right\},$$

we see now that the RHS is  $h(\hat{x})$  and if we consider the algorithm gain as a small time step, then  $\frac{\mathbb{E}(\hat{X}_k) - \mathbb{E}(\hat{X}_{k-1})}{\gamma_k}$  is an approximation of the time derivative.

For the approximation given by the ODE (3.6) to be valid as an approximation of  $\mathbb{E}(\hat{X}_k)$  at least after some time  $k$  and for using the results from [Benveniste 1990], some conditions must be satisfied:

- *conditions on the Gains.* The gains must sum to infinity

$$\sum_{k=1}^{\infty} \gamma_k = +\infty,$$

when they are decreasing, the sum of their power must be finite

$$\sum_{k=1}^{\infty} \gamma_k^\alpha < +\infty, \quad \text{for some } \alpha > 1$$

and when they are not decreasing, they must tend to a finite limit

$$\gamma_\infty = \lim_{k \rightarrow \infty} \gamma_k < +\infty.$$

As the cumulative sum of the gains is an equivalent for the time in the ODE approximation, the condition that the sum of the gains goes to infinity is equivalent to say that the time in the ODE can go to infinity, so that the algorithm does not get "stuck" in time. The condition on the sum of the powers of the decreasing gains is used to guarantee that the fluctuations of the estimator will decrease when we want to estimate a constant. The last condition on the limit of the gains is used to have fixed asymptotic performance results. We can see that all these conditions are satisfied for the three types of gain defined previously.

- *Conditions on the continuous measurements.* For a fixed  $X_k = x$ , the continuous measurements  $Y_k$  form a Markov chain with a unique stationary asymptotic distribution.

This condition is also necessary to have fixed asymptotic results. In the problem considered here, the distribution of the continuous measurements given a fixed parameter  $x$  is the distribution of the noise shifted by the parameter  $x$ . As the noise distribution is i.i.d., the distribution of  $Y_k$  is stationary for all  $k$ , thus clearly respecting this condition.

- *Regularity conditions on  $h(\hat{x})$ .* The function  $h(\hat{x})$  is locally Lipschitz continuous.

The main point of using the analysis presented in [Benveniste 1990] is that it is not necessary to have a continuous correction function  $\eta$ . The analysis is mainly based on replacing the mean of the algorithm by the ODE approximations and then evaluating the fluctuations around it. This analysis then reposes mainly on  $h$  and not on  $\eta$ . For the local existence, uniqueness and regularity of the ODE solution, we might impose regularity conditions on  $h$ . Also, for evaluating the fluctuations around the ODE solution we might look to local expansions of  $h$ , which then leads naturally to conditions as the one stated above.

Using the assumptions on the quantizer thresholds and output levels, the expectation in (3.7) can be written as:

$$h(\hat{x}) = \sum_{i=1}^{\frac{N_I}{2}} [\eta_i F_d(i, \hat{x}, x) - \eta_i F_d(-i, \hat{x}, x)], \quad (3.8)$$

where  $F_d$  is a difference of CDFs:

$$F_d = \begin{cases} F(\tau'_i \Delta + \hat{x} - x) - F(\tau'_{i-1} \Delta + \hat{x} - x), & \text{if } i \in \left\{1, \dots, \frac{N_I}{2}\right\}, \\ F(\tau'_{i+1} \Delta + \hat{x} - x) - F(\tau'_i \Delta + \hat{x} - x), & \text{if } i \in \left\{-1, \dots, -\frac{N_I}{2}\right\}. \end{cases} \quad (3.9)$$

From assumption AN3, the function  $h$  is a linear combination of locally Lipschitz continuous functions, this implies that  $h$  is also locally Lipschitz continuous, and the condition is satisfied.

All the conditions are satisfied in our case, therefore, we can apply the performance results from [Benveniste 1990].

### Mean of the algorithm for estimating a constant

For estimating a constant, the gain of the algorithm is of the form [Benveniste 1990]

$$\gamma_k = \frac{\gamma}{k}. \quad (3.10)$$

The ODE is given by (3.6)

$$\frac{d\hat{x}}{dt} = h(\hat{x}),$$

with the time given by  $t_k = \gamma \sum_{j=1}^k \frac{1}{j}$ . The ODE approximation is valid for small gains, so in this case, it is valid for large  $k$ .

The estimation bias after a transient time can be approximated using the ODE above. By denoting the bias as  $\varepsilon(t) = \hat{x}(t) - x$ , the bias ODE is

$$\frac{d\varepsilon}{dt} = \tilde{h}(\varepsilon), \quad (3.11)$$

where  $\tilde{h}(\varepsilon) = h(\varepsilon + x)$  is a function that does not depend on the true parameter  $x$  (to verify this, use  $\varepsilon + x$  in the place of  $\hat{x}$  in the expression for  $F_d$ ).

As the function  $\tilde{h}(\varepsilon)$  depends on a sum of CDF which might not even have analytical form, it is difficult to find analytical solutions for (3.11). The solution in general can be obtained using a numerical method, for example a Runge–Kutta method (see [Golub 1991] for details on numerical solvers).

Even if we cannot obtain in general a characterization of the bias for all  $k$  using the ODE, we can at least analyze what happens asymptotically to the mean of the algorithm.

### Asymptotic stability and asymptotic unbiasedness

An interesting point to study is the asymptotic mean convergence of the algorithm. More precisely, if we prove that  $\varepsilon \rightarrow 0$  as  $t \rightarrow \infty$  for every  $\varepsilon(0) \in \mathbb{R}$ , then we prove that the algorithm is asymptotically unbiased, as its true mean can be approximated by the ODE. The convergence in the mean is not only useful for showing that the algorithm indeed works, at least in the mean, but it is also a requirement for the evaluation of the MSE that will be presented later.

The fact that  $\varepsilon \rightarrow 0$  as  $t \rightarrow \infty$  for every  $\varepsilon(0) \in \mathbb{R}$  means that  $\varepsilon = 0$  is a globally asymptotically stable point [Khalil 1992]. Global asymptotic stability of  $\varepsilon = 0$  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:

$$\mathcal{L}(\varepsilon) = \varepsilon^2, \quad (3.12)$$

which is a positive definite function and tends to infinity when  $\varepsilon$  tends to infinity.

If  $\tilde{h}(\varepsilon) = 0$  for  $\varepsilon = 0$  and  $\frac{d\mathcal{L}}{dt} < 0$  for  $\varepsilon \neq 0$ , then by the Barbashin–Krasovskii theorem [Khalil 1992, p. 124]  $\varepsilon = 0$  is a globally asymptotically stable point.

To show that both conditions are met, expression (3.8) can be rewritten as a function of  $\varepsilon$ :

$$\tilde{h}(\varepsilon) = \sum_{i=1}^{\frac{N_I}{2}} \eta_i \left[ \tilde{F}_d(i, \varepsilon) - \tilde{F}_d(-i, \varepsilon) \right], \quad (3.13)$$

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

When  $\varepsilon = 0$ , the differences between  $\tilde{F}_d$  in the sum are differences between probabilities on symmetric intervals. The symmetry of the noise PDF stated in AN2 and the symmetry of the quantizer stated in AQ2' 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:

$$\frac{d\mathcal{L}}{dt} = \frac{d\mathcal{L}}{d\varepsilon} \frac{d\varepsilon}{dt} = 2\varepsilon \tilde{h}(\varepsilon) < 0, \quad \text{for } \varepsilon \neq 0. \quad (3.14)$$

Thus,  $\tilde{h}(\varepsilon)$  has to respect the following constraints:

$$\tilde{h}(\varepsilon) > 0, \quad \text{for } \varepsilon < 0 \quad \text{and} \quad \tilde{h}(\varepsilon) < 0, \quad \text{for } \varepsilon > 0. \quad (3.15)$$

When  $\varepsilon \neq 0$ , the terms in the sum that gives  $\tilde{h}(\varepsilon)$  are the difference between integrals of the noise PDF under the same interval size but with asymmetric interval centers. Using the symmetry assumptions, for  $\varepsilon > 0$ ,  $\tilde{F}_d(i, \varepsilon)$  is the integration of  $f$  over an interval more distant to zero than for  $\tilde{F}_d(-i, \varepsilon)$ , then by the decreasing assumption on  $f$ ,  $\tilde{F}_d(i, \varepsilon) < \tilde{F}_d(-i, \varepsilon)$  and consequently  $\tilde{h}(\varepsilon) < 0$ . Using the same reasoning for  $\varepsilon < 0$  one can show that  $\tilde{h}(\varepsilon) > 0$ . Therefore, the inequalities in (3.15) are satisfied and  $\frac{d\mathcal{L}}{dt} < 0$  for  $\varepsilon \neq 0$ .

Finally, as both conditions are satisfied, one can say that  $\varepsilon = 0$  is globally asymptotically stable, which means that the estimator is asymptotically unbiased for estimating a constant.

### Mean of the algorithm for estimating a Wiener process

When we want to estimate a Wiener process, the gain of the algorithm is considered to be a constant

$$\gamma_k = \gamma.$$

In this case, if we consider  $\gamma$  to be a small constant, we can also write the ODE approximation to the mean with (3.6)

$$\frac{d\hat{x}}{dt} = h(\hat{x}).$$

Now, the constant  $x$  in the expression for  $h$  is the mean of the Wiener process (which is also the mean of the initial condition  $X_0$ ) and the time is  $t_k = k\gamma$ .

Note that in this case, by imposing a  $\gamma$  sufficiently small the ODE will be valid for all  $k$  and there will be no transient time. Actually, this could also be done for the constant parameter, but as we will see later, the optimal  $\gamma$  minimizing the asymptotic MSE may not be small for estimating a constant and it will indeed be small for estimating a Wiener process with small  $\sigma_w$ .

The bias ODE is also given by (3.11), therefore, for small  $\gamma$  the algorithm is also asymptotically unbiased in this case.

To show an example for which the ODE approximates well the estimation bias, we simulated the adaptive algorithm for  $N_I = 2$  and  $N_I = 4$  in the Gaussian noise case. The quantizer gain was  $\frac{1}{\Delta} = 1$ , the threshold variations and the output coefficients were chosen to be uniform,  $\boldsymbol{\tau}' = [\tau'_1 = 1 \ \tau'_2 = 2]^\top$ ,  $\{\eta_1 = 1, \eta_2 = 2\}$  for  $N_I = 2$  and  $\boldsymbol{\tau}' = [\tau'_1 = 1 \ \tau'_2 = 2 \ \tau'_3 = 3 \ \tau'_4 = 4]^\top$ ,  $\{\eta_1 = 1, \eta_2 = 2, \eta_3 = 3, \eta_4 = 4\}$  for  $N_I = 4$ . The noise scale parameter was chosen to be  $\delta = 1$ , the Wiener process increment standard deviation  $\sigma_w = 10^{-3}$  and the adaptive gain  $\gamma = 10^{-3}$ . We considered the mean of the Wiener process to be  $\mathbb{E}(X_k) = 0$  and the initial condition of the algorithm was set to be  $\hat{X}_0 = 1$ . To obtain an estimation of the bias, we simulated the algorithm 10 times for blocks of  $10^4$  samples. For each sample (each index  $k$ ) we averaged the error through the different simulations. The solution of the bias ODE (3.11) was obtained numerically with a Runge–Kutta method with order 4 and 5. The results are displayed in Fig. 3.3.

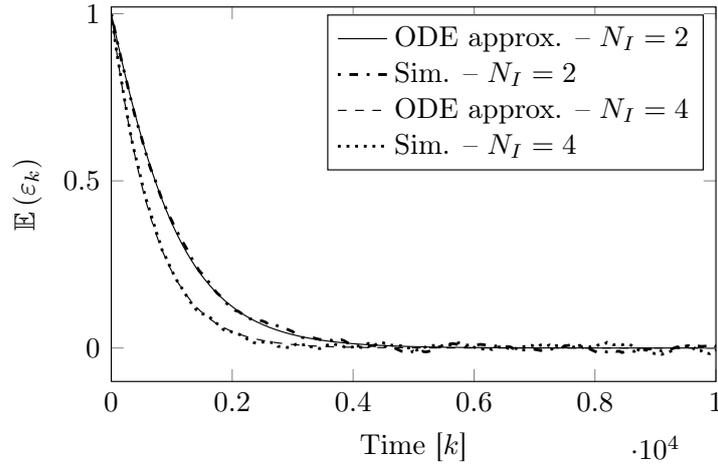


Figure 3.3: ODE bias approximation and simulated bias for the estimation of a Wiener process with the adaptive algorithm. The noise was considered to be Gaussian with  $\delta = 1$ . Both  $N_I = 2$  and  $N_I = 4$  were considered with  $\boldsymbol{\tau}' = [\tau'_1 = 1 \ \tau'_2 = 2]^\top$ ,  $\{\eta_1 = 1, \eta_2 = 2\}$  and  $\boldsymbol{\tau}' = [\tau'_1 = 1 \ \tau'_2 = 2 \ \tau'_3 = 3 \ \tau'_4 = 4]^\top$ ,  $\{\eta_1 = 1, \eta_2 = 2, \eta_3 = 3, \eta_4 = 4\}$ . In both cases, the quantizer input gain was considered to be one. The Wiener process increment standard deviation  $\sigma_w$  and the adaptive gain were set to  $10^{-3}$ . The algorithm was initialized with  $\hat{X}_0 = 1$ , while the true mean of the Wiener process was set to zero. To obtain the simulated bias, we simulated 10 realizations of the estimation procedure for blocks with  $10^4$  samples. The simulated bias was obtained through averaging of the simulations. The ODE approximation of the bias was obtained by solving numerically the ODE (3.11) with a Runge–Kutta method.

We note that the ODE approximation corresponds well to the mean trajectory of the estimation error. For this specific choice of parameters, which corresponds to the binary constant step update presented in [Li 2007] and [Fang 2008] and to a multibit extension of it (when  $N_I = 4$ ), we see that the algorithm can set the mean of the central threshold, which in this case is also the estimator, at the parameter mean even if the parameter is time-varying. We also observe that for the choice of simulation parameters used here, the convergence time of the algorithm for  $N_I = 4$  is smaller than the convergence time for  $N_I = 2$ .

As a final remark on the Wiener process case, when  $\gamma \rightarrow 0$ , the ODE approximation is increasingly accurate as the inherent discretization error (from time discretization) decreases to zero. Also when  $\gamma \rightarrow 0$ , we get the constant  $X_k$  case studied in [Li 2007] and [Fang 2008]. Thus, the proof of asymptotic mean convergence given above is also a proof of convergence of the fixed step algorithms presented in [Li 2007] and [Fang 2008] and multibit extensions of it, when the step of the algorithm is small.

### Mean of the algorithm for estimating a Wiener process with drift

When the Wiener process has a drift, we consider again that the algorithm has a constant gain

$$\gamma_k = \gamma,$$

However, in this case as the mean of the parameter is not stationary, we cannot consider the ODE approximation with a constant  $x$  in the function  $h$ .

To obtain the ODE, we will use again the heuristic presented above, but in this case, we will include the dynamical model of the parameter. We start with the expectation of the increments divided by  $\gamma$

$$\begin{aligned} \frac{\mathbb{E}(X_k) - \mathbb{E}(X_{k-1})}{\gamma} &= \frac{u}{\gamma}, \\ \frac{\mathbb{E}(\hat{X}_k) - \mathbb{E}(\hat{X}_{k-1})}{\gamma} &= \mathbb{E}\left\{\eta\left[Q\left(\frac{x - \hat{x} + V_k}{\Delta}\right)\right]\right\}, \end{aligned}$$

then we approximate it by a pair of coupled ODEs

$$\begin{aligned} \frac{dx}{dt} &= \frac{u}{\gamma}, \\ \frac{d\hat{x}}{dt} &= \tilde{h}(\hat{x} - x), \end{aligned}$$

where the time for both equations is  $t_k = k\gamma$ . Note that the algorithm ODE now depends on the solution of the parameter ODE. By subtracting both expressions, we have an ODE for the bias  $\varepsilon$

$$\frac{d\varepsilon}{dt} = \tilde{h}(\varepsilon) - \frac{u}{\gamma}. \quad (3.16)$$

As the parameter is now moving deterministically with the drift  $u$ , we can assume that most of the algorithm tracking effort will be done to remove the bias  $\varepsilon$ . Therefore, the algorithm must be fast enough to follow the parameter and we must have  $\gamma \gg u$ . This also makes  $\frac{u}{\gamma}$  to be small, thus if all  $\eta_i$  are not too small, we can find an  $\varepsilon_\infty$  such that  $\tilde{h}(\varepsilon_\infty) = \frac{u}{\gamma}$ , which means that  $\varepsilon_\infty$  is an equilibrium point for the bias.

It was shown above that the bias ODE without the forcing term  $\frac{u}{\gamma}$  is globally asymptotically stable, thus for a slowly varying parameter, we can expect that the algorithm will tend to get close to the true parameter. After a time  $t_{k-1}$ , we can assume that the algorithm is sufficiently close to the true parameter, so that we can approximate the function  $\tilde{h}(\varepsilon)$  with a first order Taylor expansion around  $\varepsilon = 0$

$$\tilde{h}(\varepsilon) = \tilde{h}(0) + \tilde{h}^{(1)}(0)\varepsilon + o(\varepsilon),$$

where  $\tilde{h}^{(1)}(0)$  is the derivative of  $\tilde{h}(\varepsilon)$  with respect to  $\varepsilon$  evaluated at  $\varepsilon = 0$ . The ODE can then be rewritten as

$$\frac{d\varepsilon}{dt} = \tilde{h}^{(1)}(0)\varepsilon - \frac{u}{\gamma} + o(\varepsilon), \quad \text{for } t > t_{k-1}. \quad (3.17)$$

For  $t_k$  sufficiently large we can neglect the  $o(\varepsilon)$  term. Thus, the bias ODE can be approximated by a linear ODE. For the linear ODE approximation not to diverge we must impose the condition

$$\tilde{h}^{(1)}(0) < 0. \quad (3.18)$$

Therefore, under this condition the approximate bias will tend to an approximation of the equilibrium point  $\varepsilon_\infty$ .

Before obtaining the asymptotic bias given by the equilibrium point  $\varepsilon_\infty$ , we will verify condition (3.18). The derivative of  $\tilde{h}(\varepsilon)$  w.r.t.  $\varepsilon$  is given by

$$\tilde{h}^{(1)}(\varepsilon) = \frac{dh}{d\varepsilon} = \sum_{i=1}^{\frac{N_I}{2}} \eta_i \left[ \tilde{f}_d(i, \varepsilon) - \tilde{f}_d(-i, \varepsilon) \right], \quad (3.19)$$

where  $\tilde{f}_d(i, \varepsilon)$  is

$$\tilde{f}_d(i, \varepsilon) = \begin{cases} f(\tau'_i \Delta + \varepsilon) - f(\tau'_{i-1} \Delta + \varepsilon), & \text{if } i \in \left\{1, \dots, \frac{N_I}{2}\right\}, \\ f(\tau'_{i+1} \Delta + \varepsilon) - f(\tau'_i \Delta + \varepsilon), & \text{if } i \in \left\{-1, \dots, -\frac{N_I}{2}\right\}. \end{cases} \quad (3.20)$$

At point  $\varepsilon = 0$ ,  $\tilde{f}_d(i, \varepsilon) = \tilde{f}_d(i, 0)$  for  $i \in \left\{1, \dots, \frac{N_I}{2}\right\}$  is negative because  $\tau'_i > \tau'_{i-1}$  and the noise PDF is strictly decreasing by assumption. For  $-i$ ,  $\tilde{f}_d(-i, 0)$  has the same absolute value as  $\tilde{f}_d(i, 0)$  by the symmetry assumptions, but it is positive. Therefore,  $\tilde{f}_d(i, 0) - \tilde{f}_d(-i, 0) = 2\tilde{f}_d(i, 0)$  and this difference is always negative. The sum  $\tilde{h}^{(1)}(\varepsilon)$  is then given by

$$\tilde{h}^{(1)}(0) = 2 \sum_{i=1}^{\frac{N_I}{2}} \eta_i \tilde{f}_d(i, 0) \quad (3.21)$$

and it is also negative, as the output quantizer levels  $\eta_i$  are positive for positive  $i$  by assumption. This means that condition (3.18) is satisfied and the ODE linear approximation will converge to an equilibrium point. For simplifying the notation, we will use  $h_\varepsilon$  in the place of  $\tilde{h}^{(1)}(0)$  from now on.

As the system is linear, the equilibrium point will be unique and independent of the initial condition. We can obtain its expression by setting  $\frac{d\varepsilon}{dt}$  to zero in the ODE approximation. This leads to the following equation:

$$h_\varepsilon \varepsilon_\infty - \frac{u}{\gamma} = 0,$$

for which the solution is

$$\varepsilon_\infty = \frac{u}{\gamma h_\varepsilon}.$$

As the bias ODE is an approximation of the true bias, this is equivalent to say that for small  $u$

$$\mathbb{E} \left( \hat{X}_k - X_k \right) \underset{k \rightarrow \infty}{\approx} \frac{u}{\gamma h_\varepsilon}. \quad (3.22)$$

Note that differently from the constant and Wiener process cases, the estimator is not asymptotically unbiased. Observe also that if  $u_k$  is not a small constant, but a small amplitude slowly varying sequence, we could replace  $u$  by  $u(t)$  in the ODE approximation above and for each time step ( $t \in [t_k, t_k + \gamma)$ ) approximate the varying  $u(t)$  by the constant  $u_k$ . This would lead to replace  $u$  by  $u_k$  in the bias approximate expression above (3.22) and instead of considering it as a valid expression for  $k \rightarrow \infty$ , we would say that it is valid for a large  $k$ .

### 3.3.2 Asymptotic MSE

After characterizing the mean behavior of the algorithm, we must quantify its random fluctuations. For doing so, we will mainly use asymptotic results on the variance of the algorithm. With the asymptotic bias and the asymptotic variance we can obtain the asymptotic MSE. The asymptotic MSE is a function of the parameter  $\gamma$ , thus by minimizing it through  $\gamma$ , we will obtain expressions for the MSE independent of  $\gamma$ .

#### Asymptotic variance for estimating a constant

Under the condition that the algorithm is asymptotically unbiased, it can be shown using a central limit theorem, that the normalized estimation error is asymptotically distributed as a Gaussian r.v. [Benveniste 1990, p. 109]

$$\sqrt{k} \left( \hat{X}_k - x \right) \underset{k \rightarrow \infty}{\rightsquigarrow} \mathcal{N} \left( 0, \sigma_\infty^2 \right), \quad (3.23)$$

the symbol  $\rightsquigarrow$  means convergence in distribution. The asymptotic variance  $\sigma_\infty^2$  is given by

$$\sigma_\infty^2 = \frac{\gamma^2 R}{-2\gamma h_\varepsilon - 1}, \quad (3.24)$$

where the term  $\tilde{h}_\varepsilon$  is the derivative of  $\tilde{h}(\varepsilon)$  w.r.t.  $\varepsilon$  at  $\varepsilon = 0$ , as it was defined before. The term  $R$  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 the mean of the algorithm, which is approximated by the ODE solution  $\hat{x}$ , is equal to  $x$ . From the symmetry assumptions on the noise and on the quantizer, the normalized mean of the increments  $h(\hat{x})$  is zero when  $\hat{x} = x$ . Thus, this variance is given by the second order moment of the quantizer output levels:

$$\begin{aligned} R &= \text{Var} \left[ \eta \left( Q \left( \frac{x - \hat{x} + V}{\Delta} \right) \right) \right] \Big|_{\hat{x}=x} \\ &= \sum_{i=1}^{\frac{N_I}{2}} \left( \eta_i^2 F_d(i, x, x) + \eta_{-i}^2 F_d(-i, x, x) \right) = 2 \sum_{i=1}^{\frac{N_I}{2}} \eta_i^2 F_d(i, x, x) \\ &= 2 \sum_{i=1}^{\frac{N_I}{2}} \eta_i^2 \tilde{F}_d(i, 0), \end{aligned} \quad (3.25)$$

where the third equality comes from the symmetry of the quantizer and the noise distribution and the last equality is obtained using the  $\tilde{F}_d$  notation.

For minimizing the asymptotic variance w.r.t.  $\gamma$ , we must find the positive  $\gamma$  for which  $\frac{d\sigma_\infty^2(\gamma)}{d\gamma} = 0$ . The expression for the derivative is

$$\frac{d\sigma_\infty^2(\gamma)}{d\gamma} = R \left[ \frac{2\gamma}{-2\gamma h_\varepsilon - 1} + \frac{2\gamma^2 h_\varepsilon}{(-2\gamma h_\varepsilon - 1)^2} \right] = \frac{R}{(-2\gamma h_\varepsilon - 1)^2} (-2\gamma^2 h_\varepsilon - 2\gamma),$$

which equals zero for  $\gamma = -\frac{1}{h_\varepsilon}$ . Note that this gain is positive as  $h_\varepsilon$  is negative. By rewriting the derivative above as

$$\frac{d\sigma_\infty^2(\gamma)}{d\gamma} = \frac{-2R\gamma h_\varepsilon}{(-2\gamma h_\varepsilon - 1)^2} \left( \gamma + \frac{1}{h_\varepsilon} \right),$$

we can see that for  $\gamma > -\frac{1}{h_\varepsilon}$ , the derivative is positive and for  $\gamma < -\frac{1}{h_\varepsilon}$ , the derivative is negative, thus  $\gamma = -\frac{1}{h_\varepsilon}$  gives a minimum  $\sigma_\infty^2$ . The optimum gain  $\gamma^*$  and its corresponding variance are

$$\gamma^* = -\frac{1}{h_\varepsilon}, \quad (3.26)$$

$$\sigma_\infty^2 = \frac{R}{h_\varepsilon^2}. \quad (3.27)$$

Note that this result is valid under the condition that the estimator is asymptotically unbiased, a condition that was shown to be true in the previous subsection.

### Asymptotic variance for estimating a Wiener process

The MSE for a varying parameter and a constant adaptive gain can be expressed as a sum of three terms

$$\begin{aligned} \text{MSE}_k &= \mathbb{E}^2[\hat{x}(t_k) - x(t_k)] + \mathbb{E} \left\{ \left\{ \left[ \hat{X}_k - \hat{x}(t_k) \right] - [X_k - x(t_k)] \right\}^2 \right\} + o(\gamma) \\ &= \varepsilon^2(t_k) + \mathbb{E}(\xi_k^2) + o(\gamma), \end{aligned} \quad (3.28)$$

where  $\varepsilon^2(t_k) = \mathbb{E}^2[\hat{x}(t_k) - x(t_k)]$  and  $\xi_k = \left[ \hat{X}_k - \hat{x}(t_k) \right] - [X_k - x(t_k)]$ . The first term  $\varepsilon^2(t_k)$  is an approximation of the squared bias  $\mathbb{E}^2[\varepsilon_k]$ . The second term is an approximation of the error variance, which can be obtained by evaluating the second order moment of the total fluctuation of the error  $\xi_k$ . The last term is the error due to the approximations and if  $\gamma$  is small this term is negligible. As  $\sigma_w$  is small by assumption,  $\gamma$  must be small for tracking  $X_k$  without large fluctuations, thus this last term is expected to be negligible.

It was shown in the last subsection that the algorithm is asymptotically unbiased, thus the first term of the decomposition tends to zero as  $k$  tends to infinity. As a consequence, the asymptotic MSE, that we denote  $\text{MSE}_{q,\infty}$ , depends mainly on the asymptotic characterization of  $\xi_k$ . Under the conditions that the estimator is asymptotically unbiased and that  $h_\varepsilon < 0$ , which were both shown to be true in the previous subsection, it can be shown [Benveniste 1990, pp. 130–131] that  $\xi_k$  tends to be a stationary Gaussian process with marginal distribution  $\mathcal{N}(0, \sigma_\xi^2)$ . The asymptotic variance  $\sigma_\xi^2$  is given as a sum of two terms, one produced by the fluctuations of the estimator itself and equal to  $\frac{\gamma R}{-2h_\varepsilon}$  and the other due to the fluctuations of the parameter and equal to  $\frac{\sigma_w^2}{-2\gamma h_\varepsilon}$ , thus giving

$$\sigma_\xi^2 = \frac{\gamma R}{-2h_\varepsilon} + \frac{\sigma_w^2}{-2\gamma h_\varepsilon}$$

and leading to the asymptotic MSE

$$\text{MSE}_{q,\infty} = \frac{\gamma R}{-2h_\varepsilon} + \frac{\sigma_w^2}{-2\gamma h_\varepsilon} + o(\gamma). \quad (3.29)$$

Neglecting the  $\circ(\gamma)$ , we can find the approximately optimal gain by equating to zero its derivative w.r.t.  $\gamma$ . This gives the equation

$$\frac{d\text{MSE}_{q,\infty}(\gamma)}{d\gamma} \approx \frac{1}{2h_\varepsilon} \left( -R + \frac{\sigma_w^2}{\gamma^2} \right) = 0,$$

which is zero for  $\gamma = \frac{\sigma_w}{\sqrt{R}}$ . The second derivative can be approximated by

$$\frac{d^2\text{MSE}_{q,\infty}(\gamma)}{d\gamma^2} \approx -\frac{\sigma_w^2}{h_\varepsilon\gamma^3},$$

as  $h_\varepsilon$  is negative and  $\sigma_w^2$  is positive, the second derivative is positive for positive  $\gamma$ . This means that choosing  $\gamma = \frac{\sigma_w}{\sqrt{R}}$  leads to a minimum MSE. Thus,

$$\gamma^* = \frac{\sigma_w}{\sqrt{R}} \quad (3.30)$$

and the corresponding asymptotic MSE is

$$\text{MSE}_{q,\infty} = \frac{\sigma_w\sqrt{R}}{-h_\varepsilon} + \circ(\gamma^*). \quad (3.31)$$

We can express  $\text{MSE}_{q,\infty}$  as a function of  $\sigma_\infty$  given in (3.27). This gives

$$\text{MSE}_{q,\infty} = \sigma_w\sigma_\infty + \circ(\gamma^*). \quad (3.32)$$

Observe that both the asymptotic MSE for estimating a Wiener process and for estimating a constant depend on the quantizer parameters ( $\eta_i$ ,  $\Delta$  and  $\tau'$ ) through an increasing function of  $\sigma_\infty^2$ , therefore the asymptotically optimal quantizer parameters is the same in both cases. The only difference in the adaptive algorithm for these two cases is the sequence of gains  $\gamma_k$ .

### Asymptotic MSE for estimating a Wiener process with drift

When the Wiener process has a drift, the MSE can still be written as the sum of three terms (3.28)

$$\text{MSE}_k = \varepsilon^2(t_k) + \mathbb{E}(\xi_k^2) + \circ(\gamma).$$

Even if  $\gamma \gg u$ , we still expect it to be small, so that the algorithm is able to reduce the effects of the measurement noise. Thus, we can still neglect the  $\circ(\gamma)$ .

We will proceed similarly as for the Wiener process without drift. We will evaluate the asymptotic MSE and then we will obtain the asymptotically optimal gain.

Differently, from the Wiener process without drift, the algorithm is not asymptotically unbiased and we must use the expression for the asymptotic bias approximation (3.22)

$$\varepsilon_\infty = \frac{u}{\gamma h_\varepsilon}$$

in the first term of  $\text{MSE}_{q,\infty}$ . As it is explained in [Benveniste 1990, p. 133], by using  $\gamma \gg u$ , the fluctuations of the parameter around its ODE approximation are negligible when compared

with the fluctuations of the algorithm. Therefore, we can approximate the asymptotic variance of the fluctuation by

$$\sigma_\xi^2 \approx \frac{\gamma R}{-2h_\varepsilon}. \quad (3.33)$$

Using the bias from (3.22) and the variance from (3.33), we obtain

$$\text{MSE}_{q,\infty} \approx \frac{u^2}{\gamma^2 h_\varepsilon^2} + \frac{\gamma R}{-2h_\varepsilon} + o(\gamma). \quad (3.34)$$

To obtain the minimum w.r.t.  $\gamma$ , we must find  $\gamma$  satisfying

$$\frac{d\text{MSE}_{q,\infty}(\gamma)}{d\gamma} \approx -\frac{2u^2}{\gamma^3 h_\varepsilon^2} + \frac{R}{-2h_\varepsilon} = 0.$$

The solution of this equation in the variable  $\gamma$  is  $\gamma = \left(\frac{4u^2}{-h_\varepsilon R}\right)^{\frac{1}{3}}$ . To verify that this value of  $\gamma$  corresponds to a minimum of  $\text{MSE}_{q,\infty}$ , we evaluate the second derivative

$$\frac{d^2\text{MSE}_{q,\infty}(\gamma)}{d\gamma^2} \approx \frac{6u^2}{\gamma^4 h_\varepsilon^2}.$$

We can verify that this quantity is positive. Therefore,

$$\gamma^* = \left(\frac{4u^2}{-h_\varepsilon R}\right)^{\frac{1}{3}} \quad (3.35)$$

and its corresponding asymptotic MSE is

$$\text{MSE}_{q,\infty} \approx 3 \left(\frac{uR}{4h_\varepsilon^2}\right)^{\frac{2}{3}} + o(\gamma^*). \quad (3.36)$$

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

The asymptotic MSE in (3.36) can also be rewritten as a function of  $\sigma_\infty^2$  with a dependence on  $u$

$$\text{MSE}_{q,\infty} \approx 3 \left(\frac{u}{4}\sigma_\infty^2\right)^{\frac{2}{3}} + o(\gamma^*). \quad (3.37)$$

Also in this case the asymptotic MSE is an increasing function of  $\sigma_\infty^2$ .

**Remark:** in the previous subsection, we remarked that if  $u_k$  is a small amplitude slowly varying parameter, the bias could be approximated by  $\varepsilon_k \approx \frac{u_k}{\gamma h_\varepsilon}$  for large  $k$ . Thus, following the same development and considering that the gains  $\gamma_k$  can be slowly variable, we have for large  $k$

$$\gamma_k^* = \left(\frac{4u_k^2}{-h_\varepsilon R}\right)^{\frac{1}{3}}$$

and the corresponding asymptotic MSE

$$\text{MSE}_k \approx 3 \left(\frac{u_k R}{4h_\varepsilon^2}\right)^{\frac{2}{3}} + o(\gamma_k^*).$$

### 3.4 Optimal algorithm parameters and performance

Now, we focus on the asymptotically optimal design of the quantizer parameters. From the previous results, we can see that the asymptotic performance for the three cases is dependent on an increasing function of  $\sigma_\infty^2$ . Also, for the three cases the asymptotic performance depends on the quantizer parameters ( $\eta_i$ ,  $\Delta$  and  $\boldsymbol{\tau}'$ ) only through  $\sigma_\infty^2$ . Therefore, the optimal parameters are the same for the three cases.

In the next subsections, first we will minimize  $\sigma_\infty^2$  w.r.t. to the quantizer update coefficients  $\eta_i$ , then we will discuss on the choice of the input gain  $\frac{1}{\Delta}$ . After that, we will present the optimal algorithm general form and its corresponding  $\sigma_\infty^2$ . We then discuss on how to optimize the performance w.r.t. the threshold variations set  $\boldsymbol{\tau}'$ . Finally, we will present the optimal gain and performance for each of the three parameter models, by considering the optimal update coefficients. In each case, we will also evaluate the performance loss due to quantization.

#### 3.4.1 Optimal algorithm parameters

##### Update coefficients (output levels)

Using the expressions for  $h_\varepsilon$  (3.21) and  $R$  (3.25) in the expression for  $\sigma_\infty^2$  (3.27), the optimization of the algorithm performance w.r.t. the update coefficients can be written as the following minimization problem:

$$\operatorname{argmin}_{\boldsymbol{\eta}} \frac{R}{h_\varepsilon^2} = \operatorname{argmin}_{\boldsymbol{\eta}} \frac{\boldsymbol{\eta}^\top \mathbf{F}_d \boldsymbol{\eta}}{2(\boldsymbol{\eta}^\top \mathbf{f}_d)^2}, \quad (3.38)$$

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

$$\boldsymbol{\eta} = \left[ \eta_1 \cdots \eta_{\frac{N_I}{2}} \right]^\top, \quad (3.39)$$

$\mathbf{F}_d$  is a diagonal matrix given by

$$\mathbf{F}_d = \operatorname{diag} \left[ \tilde{F}_d(1,0), \cdots, \tilde{F}_d\left(\frac{N_I}{2}, 0\right) \right], \quad (3.40)$$

with  $\operatorname{diag}[\cdot]$  the function that creates a matrix with the input sequence added to the diagonal of a zero matrix.  $\mathbf{f}_d$  is the following vector

$$\mathbf{f}_d = \left[ \tilde{f}_d(1,0) \cdots \tilde{f}_d\left(\frac{N_I}{2}, 0\right) \right]^\top. \quad (3.41)$$

The minimization problem is equivalent to the following maximization problem:

$$\operatorname{argmax}_{\boldsymbol{\eta}} \frac{(\boldsymbol{\eta}^\top \mathbf{f}_d)^2}{\boldsymbol{\eta}^\top \mathbf{F}_d \boldsymbol{\eta}}. \quad (3.42)$$

Using the fact that  $\mathbf{F}_d$  is a positive semidefinite matrix (it is a diagonal matrix with nonzero diagonal elements), we can rewrite (3.42) as

$$\operatorname{argmax}_{\boldsymbol{\eta}} \frac{\left[ \left( \mathbf{F}_d^{\frac{1}{2}} \boldsymbol{\eta} \right)^\top \left( \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \right) \right]^2}{\left( \mathbf{F}_d^{\frac{1}{2}} \boldsymbol{\eta} \right)^\top \left( \mathbf{F}_d^{\frac{1}{2}} \boldsymbol{\eta} \right)},$$

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

$$\left\{ \frac{\left[ \left( \mathbf{F}_d^{\frac{1}{2}} \boldsymbol{\eta} \right)^\top \left( \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \right) \right]^2}{\left( \mathbf{F}_d^{\frac{1}{2}} \boldsymbol{\eta} \right)^\top \left( \mathbf{F}_d^{\frac{1}{2}} \boldsymbol{\eta} \right)} \right\} \leq \mathbf{f}_d^\top \mathbf{F}_d^{-1} \mathbf{f}_d$$

and the equality happens for

$$\mathbf{F}_d^{\frac{1}{2}} \boldsymbol{\eta} \propto \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d.$$

Under the assumption that the update coefficients are positive for positive  $i$  AQ3 (p. 112), the optimal  $\boldsymbol{\eta}$  can be chosen to be

$$\boldsymbol{\eta}^* = -\mathbf{F}_d^{-1} \mathbf{f}_d. \quad (3.43)$$

The minimum  $\sigma_\infty^2$  w.r.t.  $\boldsymbol{\eta}$  is

$$\sigma_\infty^2 = \frac{1}{2 \left( \mathbf{f}_d^\top \mathbf{F}_d^{-1} \mathbf{f}_d \right)} = \left[ 2 \sum_{i=1}^{\frac{N_I}{2}} \frac{\tilde{f}_d^2(i, 0)}{\tilde{F}_d(i, 0)} \right]^{-1}. \quad (3.44)$$

We can recognize that the sum above is exactly equal to the FI given in (1.13) when the central threshold is placed exactly at the parameter  $x$ ,  $I_q(0)$

$$I_q(0) = 2 \sum_{i=1}^{\frac{N_I}{2}} \frac{\tilde{f}_d^2(i, 0)}{\tilde{F}_d(i, 0)}. \quad (3.45)$$

### Choice of the input gain

To simplify the choice of the constant  $\Delta$ , we can consider that the noise CDF is parametrized by a known scale parameter  $\delta$ , which means that

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

where  $F_n$  is the CDF for  $\delta = 1$ . In this case the key quantity that appears in the evaluation of the quantizer output levels is  $\frac{\Delta}{\delta}$ . Thus, the evaluation of the output levels can be simplified by setting

$$\Delta = c_\Delta \delta, \quad (3.46)$$

where  $c_\Delta$  is a constant used to adjust the input gain when the quantizer threshold variation range is fixed or to adjust the quantization step-length when the threshold variations are uniform and fixed to a value that cannot be changed.

For given  $\delta$ ,  $c_\Delta$  and  $F_n$ , the coefficients do not depend on the true parameter value, neither on the estimator value, so that they can be pre-calculated and stored in a table. In scalar form the coefficients are

$$\eta_i^* = -\frac{\tilde{f}_d(i, 0)}{\tilde{F}_d(i, 0)}. \quad (3.47)$$

Note that for  $\Delta$  given by (3.46),  $\eta_i$  depends on  $\delta$  only through a  $\frac{1}{\delta}$  multiplicative factor, the other factor can be written as a function of the normalized PDF and CDF, thus it can be pre-calculated based only on the normalized distribution.

An interesting observation is that  $\eta_i^*$  is 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.

### Optimal algorithm and general performance for the three cases

Using the  $\eta_i^*$  from (3.47) and the assumption on the symmetry of the output levels AQ3, the adaptive estimator is

$$\hat{X}_k = \hat{X}_{k-1} + \gamma_k \text{sign}(i_k) \eta_{|i_k|}^*, \quad (3.48)$$

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

The asymptotic  $(\gamma, \eta_i)$ -optimized adaptive algorithm performance is approximated for all the three cases (for the constant case it is exact) by

$$\text{MSE}_{q,\infty} \approx \psi [I_q(0)], \quad (3.49)$$

where  $\psi$  is a decreasing function of  $I_q(0)$ :

- *constant*:  $\text{MSE}_k \approx \frac{1}{kI_q(0)}$ .
- *Wiener process*:  $\text{MSE}_{q,\infty} \approx \frac{\sigma_w}{\sqrt{I_q(0)}}$ .
- *Wiener process with drift*:  $\text{MSE}_{q,\infty} \approx 3 \left(\frac{u}{4I_q(0)}\right)^{\frac{2}{3}}$ .

### Optimal threshold variations

In the performance given in (3.49), the threshold variations set  $\tau'$  is influent only through  $I_q(0)$ . Therefore, for optimizing the algorithm through  $\tau'$ , we will have the same optimization problem discussed in Ch. 1, namely (1.47)

$$I_q^* = \underset{\tau'}{\text{argmax}} I_q(0).$$

In Ch. 1, we saw that this problem is difficult in general. Two alternatives were proposed: the first one would be to constrain the quantizer to be uniform and then obtain the optimal quantizer interval step-length. The second would be to consider a general quantizer but with a very large (tending to infinity) number of quantizer intervals. For the simulated results to be presented later, Sec. 3.5, we will use the first approach. We consider that the positive threshold variations are uniform and fixed to be

$$\boldsymbol{\tau}' = \left[ -\tau'_{\frac{N_I}{2}} = -\infty \cdots -\tau'_1 = -1 \quad 0 \quad +\tau'_1 = +1 \cdots +\tau'_{\frac{N_I}{2}} = +\infty \right]^\top. \quad (3.50)$$

Then in this case, only  $c_\Delta$  need to be maximized and, as it was stated before, this can be done using a grid method.

### 3.4.2 Algorithm performance for optimal gain and coefficients

We now present for each parameter model the optimal adaptive gain  $\gamma_k^*$  and the asymptotic MSE for the update coefficients  $\boldsymbol{\eta}^*$ . In each case, after evaluating the asymptotic MSE, we will also evaluate the effect of quantization on the estimation performance. This will be done by evaluating the performance loss due to quantization  $L_q$  defined by

$$L_q = 10 \log_{10} \left( \frac{\text{MSE}_{q,\infty}}{\text{MSE}_{c,\infty}} \right), \quad (3.51)$$

where  $\text{MSE}_{q,\infty}$  is the asymptotic MSE for the adaptive algorithm based on quantized measurements and  $\text{MSE}_{c,\infty}$  is a quantity related to the asymptotic performance of estimation based on continuous measurements.  $\text{MSE}_{c,\infty}$  will be specified later for each case. Observe that the loss  $L_q$  is a relative measure and it is expressed in decibels (dB).

Before proceeding to the performance evaluation for each case, we still need to determine the quantities  $h_\varepsilon$  and  $R$  for the optimal update coefficients. Using the expression for  $\eta_i^*$  (3.47) in the expression for  $h_\varepsilon$  (3.21) and  $R$  (3.25), we have

$$h_\varepsilon = -2 \sum_{i=1}^{\frac{N_I}{2}} \frac{\tilde{f}_d^2(i,0)}{\tilde{F}_d(i,0)} = -I_q(0), \quad (3.52) \quad R = 2 \sum_{i=1}^{\frac{N_I}{2}} \frac{\tilde{f}_d^2(i,0)}{\tilde{F}_d(i,0)} = I_q(0). \quad (3.53)$$

#### 3.4.2.1 Constant case: gain and performance

Replacing  $h_\varepsilon$  given by (3.52) in (3.26) and then the result in (3.10), we have the following gains

$$\gamma_k^* = \frac{1}{kI_q(0)}. \quad (3.54)$$

Also, replacing (3.52) and (3.53) in the expression for  $\sigma_\infty^2$  (3.27), we get

$$\sigma_\infty^2 = \frac{1}{I_q(0)}. \quad (3.55)$$

In practice, this means that, for large  $k$ , the MSE will be

$$\text{MSE}_k \approx \frac{1}{kI_q(0)}. \quad (3.56)$$

The continuous asymptotic performance  $\tilde{\text{MSE}}_{c,\infty}$  can be obtained through the CRB. As the measurements are independent, the FI for  $k$  continuous measurements is  $k$  times the FI for continuous measurements  $I_c$ , thus the continuous measurement bound  $\text{CRB}_c$  is

$$\text{CRB}_c = \frac{1}{kI_c}. \quad (3.57)$$

The expression for  $I_c$  can be obtained by evaluating the expectation  $\mathbb{E}[S_c^2]$ , where the score is given by (1.15)

$$S_c(y) = \frac{\partial \log f(y-x)}{\partial x}.$$

Changing variables,  $I_c$  is given by the following integral:

$$I_c = \int_{\mathbb{R}} \left( \frac{f^{(1)}(x)}{f(x)} \right)^2 f(x) dx. \quad (3.58)$$

The ratio  $\frac{\text{MSE}_{q,\infty}}{\text{MSE}_{c,\infty}}$  is then given by  $\lim_{k \rightarrow \infty} \frac{\text{MSE}_k}{\text{CRB}_c} = \frac{I_c}{I_q(0)}$ , leading to the loss

$$L_q = -10 \log_{10} \left( \frac{I_q(0)}{I_c} \right). \quad (3.59)$$

We have the following solution to problem (a) (p. 27):

**Solution to (a) - Adaptive algorithm with decreasing gain**

**(a3) 1) Estimator**

For each time  $k$ , the estimate and threshold update is given by (3.48)

$$\hat{X}_k = \tau_{0,k} = \hat{X}_{k-1} + \gamma_k \mathbf{sign}(i_k) \eta_{|i_k|}^*,$$

with  $i_k = Q\left(\frac{Y_k - \hat{X}_{k-1}}{c_\Delta \delta}\right)$ ,  $\gamma_k = \frac{1}{k I_q(0)}$  and  $\eta_i^* = -\frac{\tilde{f}_d(i,0)}{\tilde{F}_d(i,0)}$ .

**2) Performance (asymptotic)**

$\hat{X}_k$  is asymptotically unbiased and its bias for large  $k$  can be approximated by  $\varepsilon(t_k)$ , which is the solution of the ODE (3.11)

$$\frac{d\varepsilon}{dt} = \tilde{h}(\varepsilon),$$

where  $\tilde{h}(\varepsilon) = h(\varepsilon + x)$ ,  $h$  is given by (3.7) and the time is  $t_k = \sum_{j=1}^k \gamma_j$ . Its asymptotic MSE or variance is given by (3.56)

$$\text{MSE}_k \underset{k \rightarrow \infty}{\sim} \frac{1}{k I_q(0)},$$

where  $I_q(0)$  is given by (1.13) with  $\varepsilon = 0$ , representing a loss of performance w.r.t. the asymptotically optimal estimator based on continuous measurements of (3.59)

$$L_q = -10 \log_{10} \left( \frac{I_q(0)}{I_c} \right),$$

with  $I_c$  the continuous measurement FI given by (3.58).

### 3.4.2.2 Wiener process case: gain and performance

Using (3.53) in (3.30), we obtain the optimal constant gain

$$\gamma^* = \frac{\sigma_w}{\sqrt{I_q(0)}} \quad (3.60)$$

and for this gain, the asymptotic MSE is given by substituting (3.55) in (3.32)

$$\text{MSE}_{q,\infty} = \frac{\sigma_w}{\sqrt{I_q(0)}} + o(\sigma_w). \quad (3.61)$$

Note that we used the fact that  $\gamma^*$  in this case depends linearly on  $\sigma_w$  for writing the  $o$  term.

The comparison with the continuous case can be done by using the asymptotic BCRB for continuous measurements as  $\text{MSE}_{c,\infty}$ . The evaluation of the asymptotic BCRB follows in the same line as the one presented in Ch. 2 for estimation based on quantized measurements. The main difference is that in the continuous case, the FI  $I_c$  is independent of the parameter value, thus  $\mathbb{E}[I_c] = I_c$  and we do not need to consider a lower bound on the BCRB. For small  $\sigma_w$  (small compared with  $I_c$ ) the asymptotic BCRB can be approximated exactly in the same way as for the lower bound on the MSE for quantized measurements

$$\text{BCRB}_{c,\infty} = \frac{\sigma_w}{\sqrt{I_c}} + o(\sigma_w). \quad (3.62)$$

The loss of performance, in this case denoted  $L_q^W$ , is given as follows

$$L_q^W = 10 \log_{10} \left( \frac{\text{MSE}_{q,\infty}}{\text{BCRB}_{c,\infty}} \right) = 10 \log_{10} \left( \frac{\frac{\sigma_w}{\sqrt{I_q(0)}} + o(\sigma_w)}{\frac{\sigma_w}{\sqrt{I_c}} + o(\sigma_w)} \right). \quad (3.63)$$

We multiply the numerator and the denominator inside the logarithm of (3.63) by  $\frac{\sqrt{I_c}}{\sigma_w}$ . This gives

$$L_q^W = 10 \log_{10} \left( \frac{\sqrt{\frac{I_c}{I_q(0)}} + \frac{o(\sigma_w)}{\sigma_w}}{1 + \frac{o(\sigma_w)}{\sigma_w}} \right),$$

where we have assimilated the  $\sqrt{I_c}$  in the  $o(\sigma_w)$  term. Using the first order Taylor expansion around  $x = 0$ ,  $\frac{1}{1+x} = 1 - x + o(x)$ , we can obtain

$$L_q^W = 10 \log_{10} \left( \sqrt{\frac{I_c}{I_q(0)}} + \frac{o(\sigma_w)}{\sigma_w} \right).$$

Then factorizing  $\sqrt{\frac{I_c}{I_q(0)}}$  and using the first order Taylor expansion around  $x = 0$ ,  $\log_{10}(1+x) = \frac{x}{\ln(10)} + o(x)$ , where  $\ln$  is the natural logarithm, we have

$$L_q^W = 10 \log_{10} \left( \sqrt{\frac{I_c}{I_q(0)}} \right) + \frac{o(\sigma_w)}{\sigma_w} = -5 \log_{10} \left( \frac{I_q(0)}{I_c} \right) + \frac{o(\sigma_w)}{\sigma_w}.$$

Note that the first term is half the loss of performance for the constant case

$$L_q^W = \frac{1}{2} L_q + \frac{o(\sigma_w)}{\sigma_w}. \quad (3.64)$$

From the definition of the  $o$  term we also have

$$\lim_{\sigma_w \rightarrow 0} L_q^W = \frac{1}{2} L_q.$$

This gives the following solution to problem (b) (p. 29) when the parameter is modeled by a Wiener process without drift:

**Solution to (b) - Adaptive algorithm with constant gain for tracking a Wiener process with small  $\sigma_w$ .**

**(b3.1) 1) Estimator**

For each time  $k$ , the estimate and threshold update is given by (3.48)

$$\hat{X}_k = \tau_{0,k} = \hat{X}_{k-1} + \gamma \mathbf{sign}(i_k) \eta_{|i_k|}^*,$$

$$\text{with } i_k = Q\left(\frac{Y_k - \hat{X}_{k-1}}{c_{\Delta} \delta}\right), \quad \gamma = \frac{\sigma_w}{\sqrt{I_q(0)}} \quad \text{and} \quad \eta_i^* = -\frac{\tilde{f}_d(i,0)}{\tilde{F}_d(i,0)}.$$

**2) Performance (approximated and asymptotic)**

$\hat{X}_k$  is asymptotically unbiased and its bias can be approximated by  $\varepsilon(t_k)$ , which is the solution of the ODE (3.11)

$$\frac{d\varepsilon}{dt} = \tilde{h}(\varepsilon),$$

where  $\tilde{h}(\varepsilon) = h(\varepsilon + x)$ ,  $h$  is given by (3.7) and the time is  $t_k = k\gamma$ . Its asymptotic MSE or variance is given by (3.61)

$$\text{MSE}_{q,\infty} = \frac{\sigma_w}{\sqrt{I_q(0)}} + o(\sigma_w),$$

where  $I_q(0)$  is given by (1.13) with  $\varepsilon = 0$ , representing a loss of performance w.r.t. the asymptotically optimal estimator based on continuous measurements of (3.64)

$$L_q^W = -5 \log_{10} \left( \frac{I_q(0)}{I_c} \right) + \frac{o(\sigma_w)}{\sigma_w} = \frac{1}{2} L_q + \frac{o(\sigma_w)}{\sigma_w},$$

with  $I_c$  the continuous measurement FI given by (3.58) and  $L_q$  the loss of the adaptive algorithm for estimating a constant.

### 3.4.2.3 Wiener process with drift case: gain and performance

Replacing the expressions for  $h_\varepsilon$  (3.52) and  $R$  (3.53) in the expressions for  $\gamma^*$  (3.35) and  $\text{MSE}_{q,\infty}$  (3.36), we obtain

$$\gamma^* = \left( \frac{4u^2}{I_q^2(0)} \right)^{\frac{1}{3}}, \quad (3.65) \quad \text{MSE}_{q,\infty} \approx 3 \left( \frac{u}{4I_q(0)} \right)^{\frac{2}{3}} + o(\gamma^*). \quad (3.66)$$

If  $u$  is unknown, it might be estimated. It can be estimated by smoothing the differences

between successive estimates

$$\hat{U}_k = \hat{U}_{k-1} + \gamma_k^u \left[ \left( \hat{X}_k - \hat{X}_{k-1} \right) - \hat{U}_{k-1} \right]. \quad (3.67)$$

where  $\gamma_k^u$  is a sequence of small positive gains. The estimator  $\hat{U}_k$  can replace  $u$  in the evaluation of the gain and of the asymptotic MSE. If the drift is not constant but slowly varying, the adaptive algorithm above can also be used. In this case, additional information on the evolution of the drift might be incorporated in (3.67) to have more precise estimates and get an adaptive gain closer to the optimal.

For the evaluation of the loss due to quantization, we could use  $\text{BCRB}_{c,\infty}$  for the continuous measurement performance. However, this would result in an unfair comparison, as the imposition of using  $\hat{X}_{k-1}$  instead of the prediction is known to be suboptimal. Therefore, the evaluation of the loss will be done using the approximate performance for an adaptive algorithm of the same form, but using continuous measurements instead of quantized measurements. The algorithm has the following form:

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

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

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

$$\gamma_k^{c,*} = \left( \frac{4u^2}{I_c^2} \right)^{\frac{1}{3}}, \quad \eta_c(x) = -\frac{f'(x)}{f(x)},$$

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

$$\text{MSE}_{c,\infty} \approx 3 \left[ \frac{u}{4I_c} \right]^{\frac{2}{3}}. \quad (3.68)$$

This asymptotic MSE can be used as  $\tilde{\text{MSE}}_{c,\infty}$  in the evaluation of the loss. Using similar Taylor expansions as in the previous Wiener model and denoting the loss in this case by  $L_q^{WD}$ , we have

$$L_q^{WD} \approx -\frac{20}{3} \log_{10} \left( \frac{I_q(0)}{I_c} \right) + \frac{\circ \left( u^{\frac{2}{3}} \right)}{u^{\frac{2}{3}}} = \frac{2}{3} L_q + \frac{\circ \left( u^{\frac{2}{3}} \right)}{u^{\frac{2}{3}}}. \quad (3.69)$$

Note that here the limit result is on  $u$

$$\lim_{u \rightarrow 0} L_q^{WD} = \frac{2}{3} L_q.$$

However, note also that hidden in the approximation is the fact that  $\sigma_w$  must also tend to zero.

We have the following solution to problem (b) (p. 29) when the parameter is modeled by a Wiener process with deterministic drift:

**Solution to (b) - Adaptive algorithm with constant gain for tracking a Wiener process with small  $\sigma_w$  and small  $u$ .**

**(b3.2) 1) Estimator**

For each time  $k$ , the estimate and threshold update is given by (3.48)

$$\hat{X}_k = \tau_{0,k} = \hat{X}_{k-1} + \gamma \mathbf{sign}(i_k) \eta_{|i_k|}^*,$$

$$\text{with } i_k = Q\left(\frac{Y_k - \hat{X}_{k-1}}{c_\Delta \delta}\right), \gamma = \left(\frac{4u^2}{I_q^2(0)}\right)^{\frac{1}{3}} \text{ and } \eta_i^* = -\frac{\tilde{f}_d(i,0)}{\tilde{F}_d(i,0)}.$$

**2) Performance**

(approximated and approximated asymptotic)

The estimation bias can be approximated by  $\varepsilon(t_k)$ , which is the solution of the ODE (3.16)

$$\frac{d\varepsilon}{dt} = \tilde{h}(\varepsilon) - \frac{u}{\gamma},$$

where  $\tilde{h}(\varepsilon) = h(\varepsilon + x)$ ,  $h$  is given by (3.7),  $x$  is the mean of the Wiener process and the time is  $t_k = k\gamma$ . Its asymptotic MSE or variance is approximated as follows (3.66)

$$\text{MSE}_{q,\infty} \approx 3 \left(\frac{u}{4I_q(0)}\right)^{\frac{2}{3}} + o\left(u^{\frac{2}{3}}\right),$$

where  $I_q(0)$  is given by (1.13) with  $\varepsilon = 0$ , representing a loss of performance w.r.t. the asymptotically optimal adaptive estimator based on continuous measurements of (3.69)

$$L_q^{WD} \approx -\frac{20}{3} \log_{10} \left(\frac{I_q(0)}{I_c}\right) + \frac{o\left(u^{\frac{2}{3}}\right)}{u^{\frac{2}{3}}} = \frac{2}{3} L_q + \frac{o\left(u^{\frac{2}{3}}\right)}{u^{\frac{2}{3}}},$$

with  $I_c$  the continuous measurement FI given by (3.58) and  $L_q$  the loss of the adaptive algorithm for estimating a constant.

Observe that 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 thresholds set when comparing to the equivalent continuous measurements based algorithm.

### 3.5 Simulations

Now, we are going to check the validity of the results through simulation. We will mainly focus on obtaining a simulated version of the loss of performance for the three parameter models and then we will compare the simulated loss with the theoretical one. After that, we will compare the adaptive algorithm performance with the algorithms presented in the previous chapters, namely the adaptive MLE scheme for estimating a constant and the PF with dynamical central threshold for estimating a Wiener process. This comparison will allow us to know if we lose in estimation performance and what we lose in estimation performance, when we use the low complexity adaptive algorithm presented in this chapter, instead of the algorithms presented in the previous chapters.

#### 3.5.1 General considerations

**Threshold variations.** In what follows the threshold variations are considered to be uniform and given by (3.50)

$$\boldsymbol{\tau}' = \left[ -\tau'_{\frac{N_I}{2}} = -\infty \cdots -\tau'_1 = -1 \quad 0 \quad +\tau'_1 = +1 \cdots +\tau'_{\frac{N_I}{2}} = +\infty \right]^\top.$$

#### Evaluation of $I_q(0)$ and the algorithm parameters

For a given type of noise, supposing that its noise scale parameter  $\delta$  is known, for a fixed  $N_I$ ,  $I_q(0)$  can be evaluated by using the normalized CDF and PDF,  $F_n$  and  $f_n$  (CDF and PDF for  $\delta = 1$ ), in (3.45) (or (1.46)). Using the parametrization  $\Delta = c_\Delta \delta$  and the fact that  $f(x) = \frac{1}{\delta} f_n\left(\frac{x}{\delta}\right)$ , we have

$$I_q(0) = \frac{2}{\delta^2} \sum_{i=1}^{\frac{N_I}{2}} \frac{\{f_n[(i-1)c_\Delta] - f_n[ic_\Delta]\}^2}{\{F_n[ic_\Delta] - F_n[(i-1)c_\Delta]\}}. \quad (3.70)$$

As  $I_q(0)$  is now a function of  $c_\Delta$  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  $c_\Delta$ ). After finding the optimal  $c_\Delta^*$ , the coefficients  $\eta_i = \eta_i^*$  can be evaluated using the normalized CDF and PDF in (3.47). This gives

$$\eta_i^* = \frac{1}{\delta} \frac{f_n[(i-1)c_\Delta^*] - f_n[ic_\Delta^*]}{F_n[ic_\Delta^*] - F_n[(i-1)c_\Delta^*]}. \quad (3.71)$$

Then, with  $\delta$ , the optimal  $I_q(0)$  and depending on the model,  $\sigma_w$  or  $u$ , we can evaluate  $\frac{1}{\Delta}$ ,  $\gamma_k$  and then all the algorithm parameters are defined.

#### Discussion on the signal model

Note that it is supposed that the model for  $X_k$  is known, as setting  $\gamma_k$  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,  $\sigma_w$  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 that the  $\gamma_k$  depends on  $u$  is not very useful and prior information on the variations of  $X_k$  are not normally as detailed as knowing  $u$  itself, making it necessary to accept a small loss of performance to estimate  $u$  jointly. The estimation of  $u$  can be done using (3.67) where prior knowledge on the variations of  $u_k$  can be integrated in the gain  $\gamma_k^u$ . If precise knowledge on the evolution of  $u_k$  is known through dynamical models, it might be more useful to use other forms of adaptive estimators known as multi-step algorithms [Benveniste 1990, Ch. 4].

### Discussion on the noise model

The evaluation of the loss and the verification of the results will be done considering two different classes of noise that verify assumptions AN1, AN2 and AN3, namely, generalized Gaussian (GGD) noise and noise distributed according to the **Student's-t distribution (STD)**. The motivation for the use of these two distributions 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 GGD [Varanasi 1989]. This distribution not only contains the Gaussian case as a specific example, but also by changing one of its parameters, one can model the impulsive Laplacian distribution as well as distributions close to uniform. In robust statistics, when the additive noise is considered to be impulsive, a general class for the distribution of the noise is the STD [Lange 1989]. STD includes as a specific case the Cauchy distribution, known to be heavy-tailed and used intensively 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 prior is used for the noise, noise models must be as random as possible to ensure that the noise is an uninformative part of the measurement. Thus, noise models must maximize some criterion of randomness. Commonly used criteria for randomness are entropy measures and both distributions considered above are entropy maximizers. The GGD maximizes the Shannon entropy under constraints on the moments [Cover 2006, Ch. 12] and the STD maximizes the Rényi entropy under constraints on the second order moment [Costa 2003].

Both families of distributions are parametrized by a shape parameter  $\beta \in \mathbb{R}_+$  and a scale parameter  $\delta$ . The CDF and PDF of the GGD were given in Ch. 1 by (1.39) and (1.40)

$$f_{GGD}(x) = \frac{\beta}{2\delta\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left|\frac{x}{\delta}\right|^\beta\right),$$

$$F_{GGD}(x) = \frac{1}{2} \left[ 1 + \text{sign}(x) \frac{\gamma\left(\frac{1}{\beta}, \left|\frac{x}{\delta}\right|^\beta\right)}{\Gamma\left(\frac{1}{\beta}\right)} \right],$$

while for the STD, the CDF and PDF are respectively

$$f_{STD}(x) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\delta\sqrt{\beta\pi}\Gamma\left(\frac{\beta}{2}\right)} \left[1 + \frac{1}{\beta} \left(\frac{x}{\delta}\right)^2\right]^{-\frac{\beta+1}{2}}, \quad (3.72)$$

$$F_{STD}(x) = \frac{1}{2} \left\{ 1 + \text{sign}(x) \left[ 1 - I_{\left(\frac{x}{\delta}\right)^2 + \beta} \left( \frac{\beta}{2}, \frac{1}{2} \right) \right] \right\}, \quad (3.73)$$

$I_*(\cdot, \cdot)$  is the incomplete beta function

$$I_w(x, y) = \int_0^w z^{x-1} (1-z)^{y-1} dz.$$

### 3.5.2 Theoretical performance loss due to quantization

The main quantity that must be evaluated before simulating the algorithm is the theoretical loss  $L_q$ . This quantity will not only be useful to check the simulation results, but will also be useful to observe how the performance evolves as we change the number of quantization intervals and as we change the noise model.

To evaluate  $L_q$ , after evaluating  $I_q(0)$  based on the CDF and PDF given above, we also need to evaluate  $I_c$ . The continuous measurement FI for the GGD can be obtained by using (1.39) in the integral expression (3.58), this gives (Why? - App. A.1.7)

$$I_{c,GGD} = \frac{1}{\delta^2} \frac{\beta(\beta-1)\Gamma\left(1 - \frac{1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)}. \quad (3.74)$$

For the STD the continuous measurement FI is given by using (3.72) also in (3.58). Integrating, we obtain (Why? - App. A.1.8)

$$I_{c,STD} = \frac{1}{\delta^2} \frac{\beta+1}{\beta+3}. \quad (3.75)$$

We evaluated the theoretical loss for  $N_I \in \{2, 4, 8, 16, 32\}$ , which corresponds to  $N_B = \log_2(N_I) \in \{1, 2, 3, 4, 5\}$  numbers of bits, for shape parameters  $\beta \in \{1.5, 2, 2.5, 3\}$  for GGD noise and  $\beta \in \{1, 2, 3\}$  for STD noise. The results are shown in Fig. 3.4. As it was intuitively 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 bit 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 one bit seems to be related to the noise tail. Note that smaller losses were obtained for distributions with heavier tail (STD in general and GGD with  $\beta = 1.5$ ). This is due to the fact that for large tail distributions a small region around the median of the

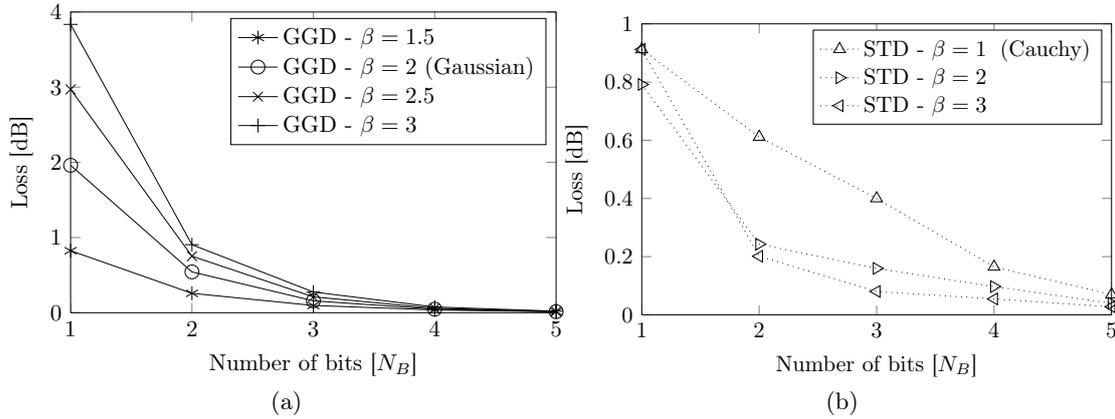


Figure 3.4: Adaptive algorithm loss of estimation performance due to quantization of measurements corresponding to the constant case  $L_q$  (theoretical). The loss is evaluated for different types of noise, GGD noise in (a) and STD noise in (b), and different numbers of quantization bits. For the other models of parameter studied here, the loss is proportional to  $L_q$ .

distribution is very informative, thus (as most of the information is contained there) when the only threshold available is placed close to the median, 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 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.

The loss was not shown in Fig. 3.4 for the Laplacian distribution, because for this distribution the adaptive optimal estimator in the continuous case is already an adaptive estimator with a binary quantizer. One can see this by evaluating the coefficients  $\eta_i$ , which in this case are constant for positive  $i$  showing that only the sign of the difference between the measurement and the last estimate is important. This behavior of optimality for binary quantization was already observed in Ch. 1, where we showed that the CRB for binary quantized measurements can be equal to the CRB for continuous measurements in the Laplacian case. Consequently, the loss in this case is zero dB for all  $N_B$ .

### 3.5.3 Simulated loss

To validate the results, we will simulate the loss of performance. The simulation results will be presented in the same order as the theoretical results presented in the previous sections. First the constant case, then the Wiener process case and finally the Wiener process with drift. All the simulations are done for  $N_B \in \{2, 3, 4, 5\}$ .

#### Simulated loss: constant case

In the constant case, the 7 types of noise with previously evaluated  $L_q$  were tested, the value of  $X_0 = x$  was set to 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 5000 to ensure convergence. The algorithm was simulated  $2.5 \times 10^6$  times and the error results were averaged yielding a

simulated MSE. Based on the simulated MSE a simulated loss was calculated. GGD noise was simulated using a transformation of gamma variates (How? - App. A.3.2), while STD noise was simulated using a transformation of independent uniform variates similar to the transformation used for generating Gaussian variates (How? - App. A.3.5). The results are shown in Fig. 3.5 for GGD noise and in Fig. 3.6 for STD noise.

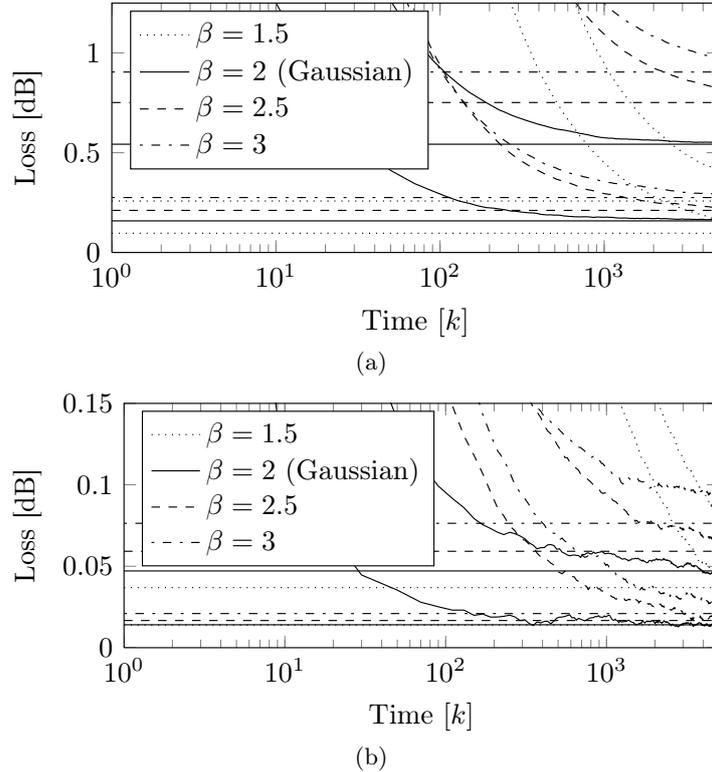


Figure 3.5: Quantization loss of performance for GGD noise and  $N_B \in \{2, 3, 4, 5\}$  when  $X_k$  is constant. For each type of noise there are 4 curves, the constant losses are the theoretical results and the decreasing losses are the simulated results, thus producing pairs of curves of the same type, for each pair the higher results represent lower number of quantization bits. In (a) results for  $N_B = 2$  and  $3$  are shown. In (b) the results for  $N_B = 4$  and  $5$  are shown. The simulated results were obtained through Monte Carlo simulation using  $2.5 \times 10^6$  realizations of blocks of 5000 error samples, the true parameter value in all simulations was set to zero, while  $\hat{X}$  was set to have a small initial error ( $\hat{X}_0 \in \{0, 10\}$ ). We used  $\delta = 1$  in all simulations.

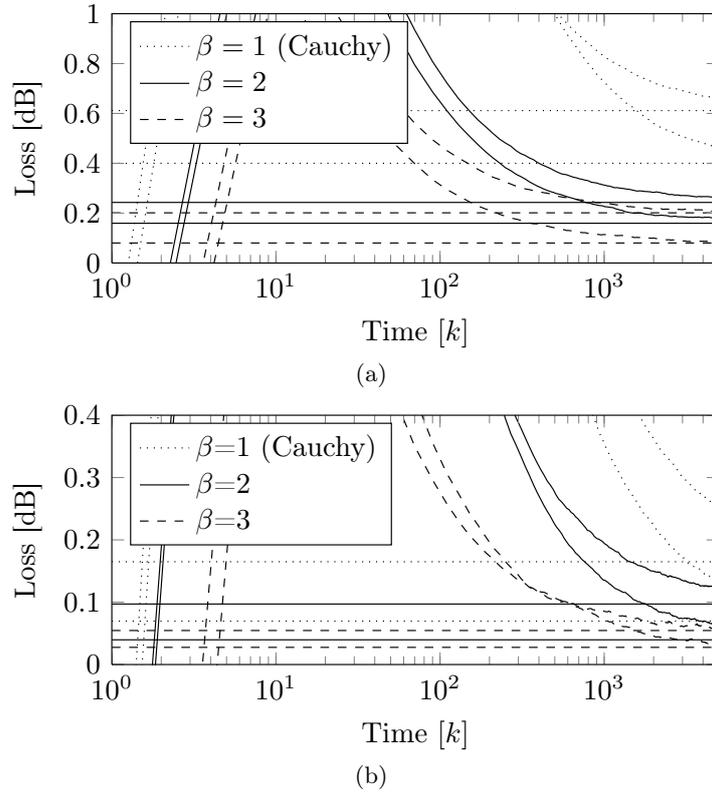


Figure 3.6: Quantization loss of performance for STD noise and  $N_B \in \{2, 3, 4, 5\}$  when  $X_k$  is constant. For each type of noise there are 4 curves, the constant losses are the theoretical results and the decreasing losses are the simulated results, thus producing pairs of curves of the same type, for each pair the higher results represent lower number of quantization bits. In (a) results for  $N_B = 2$  and  $3$  are shown. In (b) results for  $N_B = 4$  and  $5$  are shown. The simulated results were obtained through Monte Carlo simulation using  $2.5 \times 10^6$  realizations of blocks of 5000 error samples, the true parameter value in all simulations was set to zero, while  $\hat{X}$  was set to have a small initial error ( $\hat{X}_0 \in \{0, 10\}$ ). We used  $\delta = 1$  in all simulations.

#### Remarks:

- note that the losses are independent of  $\delta$  as both  $I_q(0)$  and  $I_c$  depend on it through the same multiplicative constant  $\frac{1}{\delta^2}$ .
- The simulated results seem 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 observations  $\frac{1}{kI_q(0)}$ , showing that the algorithm is asymptotically optimal.
- The convergence time seems to be related to  $N_B$  (when  $N_B$  increases, the time to get closer to the optimal performance decreases).

### Simulated loss: Wiener process case

For a Wiener process,  $L_q^W$  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. As it was observed that the error was approximately stationary after  $k = 1000$ , the sample MSE 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. 3.7. As expected, the results have the same form of the theoretical loss given in Fig. 3.4.

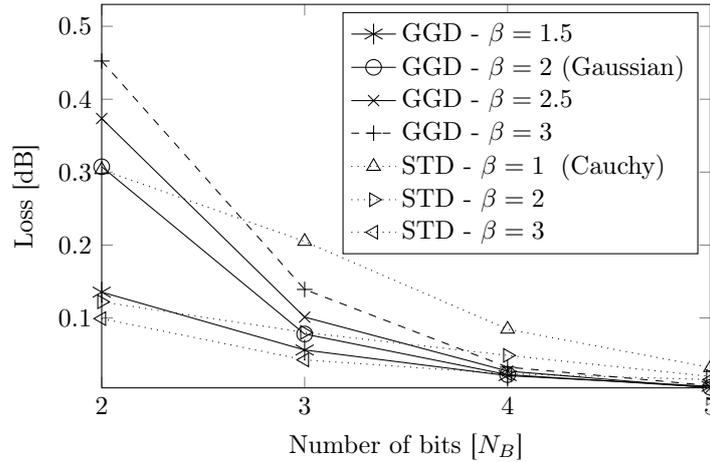


Figure 3.7: Simulated quantization performance loss for a Wiener process  $X_k$  with  $\sigma_w = 0.001$ , different types of noise and numbers of quantization bits. The simulated losses were obtained through Monte Carlo simulation. For each evaluated loss (each symbol on the curves)  $10^4$  realizations with  $10^5$  samples were simulated. As it was observed that the error is stationary after  $k = 1000$ , the sample MSE was also averaged leading to an estimate of the asymptotic MSE and consequently of the loss. The simulations were done by setting the initial estimate randomly around zero (with a Gaussian distribution) and also by setting  $X_0 = 0$ . In all simulations, we considered  $\delta = 1$ .

To verify the results for different values of  $\sigma_w$ , the loss was evaluated through simulation also for  $\sigma_w = 0.1$  in the Gaussian (GGD with  $\beta = 2$ ) and Cauchy cases (STD with  $\beta = 1$ ). The results are shown in Fig. 3.8, where the theoretical losses for these cases are also shown. These results clearly show that  $X_k$  may move slowly to give a performance close to the theoretical results. However, it is also interesting to note that the simulated loss seems to have the same decreasing rate as a function of  $N_B$  when compared with the theoretical results. This means that the dependence on  $I_q(0)$  of the MSE seems to be still correct. Moreover, it indicates that even in a faster regime for  $X_k$ , the threshold variations can be set by maximizing  $I_q(0)$ .

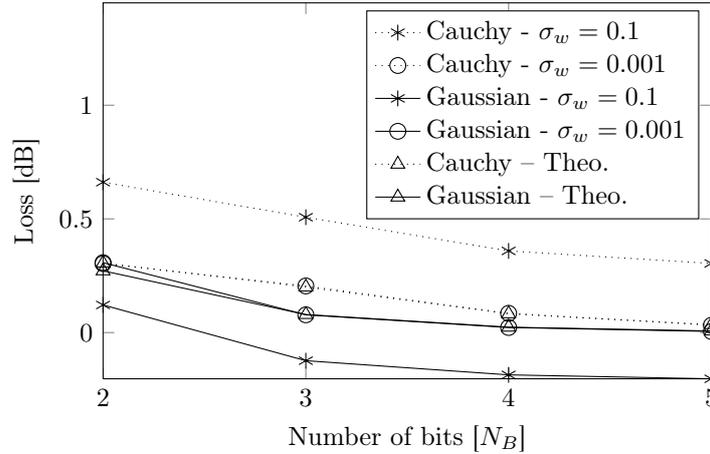


Figure 3.8: Comparison of simulated and theoretical losses in the Gaussian and Cauchy noise cases when estimating a Wiener process with  $\sigma_w = 0.1$  or  $\sigma_w = 0.001$ . The simulated losses were obtained through Monte Carlo simulation. For each evaluated loss (each symbol on the curves)  $10^4$  realizations with  $10^5$  samples were simulated. As it was observed that the error is stationary after  $k = 1000$ , the sample MSE was also averaged leading to an estimate of the asymptotic MSE and consequently of the loss. The simulations were done by setting the initial estimate randomly around zero (with a Gaussian distribution) and also by setting  $X_0 = 0$ . In all simulations, we considered  $\delta = 1$ .

### Simulated loss: Wiener process with drift case

For a Wiener process  $X_k$  with drift,  $W_k$  was simulated with mean and standard deviations  $u = \sigma_w = 10^{-4}$ , which represents a slow drift with small random fluctuations. The initial conditions were set to  $X_0 = \hat{X} = 0$  and the drift estimator was set with constant gain  $\gamma_k^u = 10^{-5}$ . Its initial condition was set to the true  $u$  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, after averaging the squared error through realizations and time. The results for the Gaussian and Cauchy cases are shown in Fig. 3.9.

The small offset between the simulated and theoretical results is justified by the joint estimation of  $u$  and  $X_k$ . Note that keeping  $\gamma_k^u$  small allows one to adaptively follow slow variations in the drift. The convergence to the simulated loss in Fig. 3.9 was also obtained for simulations including errors in the initial conditions. However, in this case, the transient regime was very long, indicating that other schemes might be considered when the theoretical performance is needed in a short period of time.

Note also that if the drift is known, the procedure simulated for tracking  $X_k$  is clearly suboptimal. In this case, we can obtain better asymptotic results by using the prediction (which includes the drift) in the adaptive algorithm. However, in practice, as we have to estimate jointly the unknown drift, the simulated algorithm normally has a shorter transient than the version using the prediction. This is an advantage when the drift can vary in time.

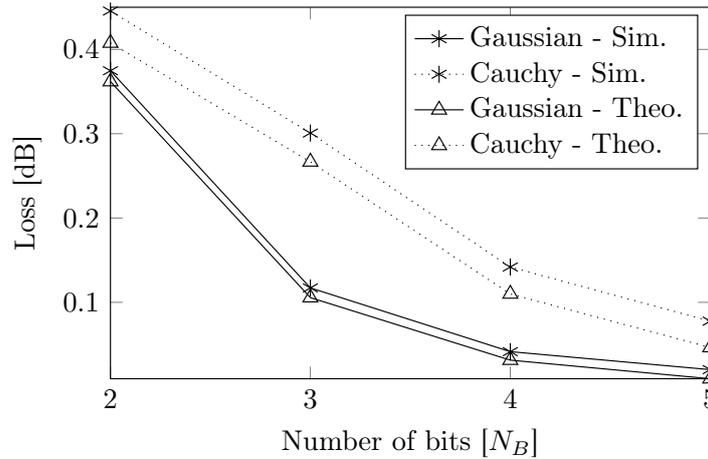


Figure 3.9: Comparison of simulated and theoretical losses in the Gaussian and Cauchy noise cases for estimating a Wiener process with constant mean drift  $u_k = 10^{-4}$  and standard deviation  $\sigma_w = 10^{-4}$ . The simulation results were obtained with  $10^4$  realizations of  $10^5$  samples, for evaluating the simulated asymptotic MSE, the squared error samples were averaged through the realizations and through the time samples after the transient time (for  $k > 1000$ ). The initial estimate value and initial parameter value were both set to zero. The initial value of the estimate of the drift was also set to the true parameter value to reduce the transient time.

### 3.5.4 Comparison with the high complexity algorithms

The adaptive algorithms that we propose will be compared with their equivalent counterparts given in previous chapters. When the parameter is constant, we will compare the adaptive algorithm with decreasing gain (a3) with the adaptive algorithm based on the MLE (a2.2) presented in Ch. 1 (p. 69). We will discuss the main differences in terms of performance and computational complexity.

#### Adaptive algorithm *vs* adaptive MLE

**Asymptotic performance.** Asymptotically both algorithms are equivalent, since they are asymptotically unbiased and their asymptotic variance is equivalent to  $\frac{1}{kI_q(0)}$ . This means that for commonly used noise distributions both algorithms are asymptotically optimal under the unbiasedness constraint. Thus, if there is a difference in performance, this difference might be found in the transient, before getting close to the asymptotic performance.

**Transient performance.** The transient for both algorithms is difficult to study analytically. For the adaptive scheme with decreasing gain, the first few steps will be mainly characterized by the bias. Unfortunately, the bias approximation given by the ODE approximation cannot be used in the initial transient as the size of the steps is too large. For the adaptive scheme based on the MLE, we cannot obtain any result either, as the general behavior of the MLE is known only asymptotically. Therefore, we will analyze the transient through simulations.

We simulated both algorithms for  $N_I = 8$  and two different types of noises, Gaussian and Cauchy noises. The threshold variations were considered to be uniform with step-length

chosen in the same way as for the evaluation and simulation of the losses. For evaluating the simulated MSE for the transient, we simulated 1000 realizations of the algorithms, each realization with 50 samples. The noise scale factor used for both cases was  $\delta = 1$  and the parameter and initial estimate were  $x = 0$  and  $\hat{X}_0 = 1$ . For starting the adaptive scheme based on the MLE, 10 samples with fixed thresholds were used for obtaining the first estimate. The algorithm used in the maximization procedure of the MLE was a search algorithm<sup>1</sup>. The results are shown in Fig. 3.10, where we also show the CRB for quantized measurements when the central threshold is placed at the true parameter  $CRB_q^* = \frac{1}{kI_q(0)}$ .

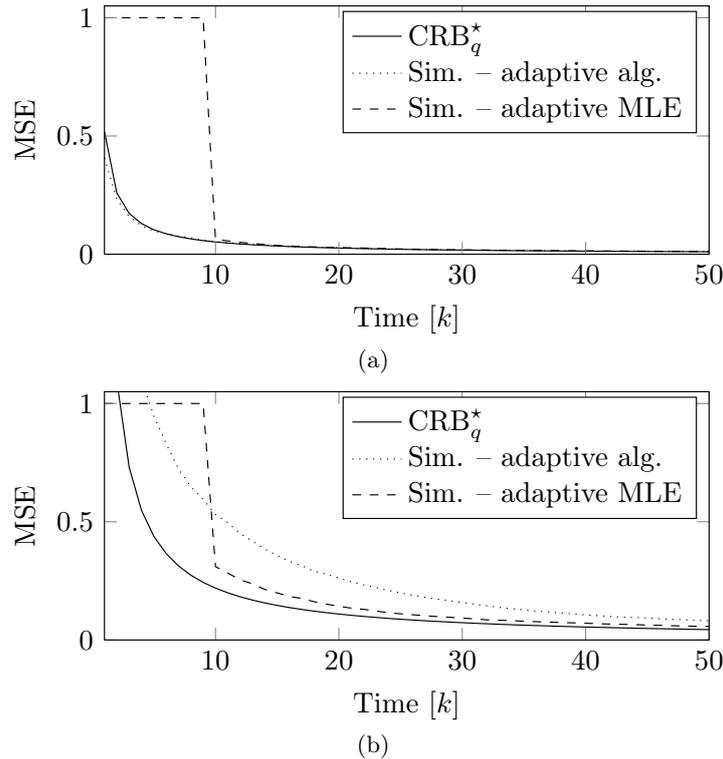


Figure 3.10: Minimum CRB and simulated MSE for the adaptive algorithm with decreasing gain and for the adaptive algorithm based on the MLE. Both algorithms were simulated with  $N_I = 8$ , optimal uniform thresholds, Gaussian and Cauchy noise with  $\delta = 1$ ,  $x = 0$  and  $\hat{X}_0 = 1$ . For evaluating the transient MSE, 50 samples were simulated 1000 times for each algorithm. The scheme based on the MLE is started by applying the MLE with samples obtained with fixed thresholds. The maximization in the MLE is done with a search algorithm<sup>1</sup>. In (a), results for Gaussian noise are shown. In (b), the results for Cauchy noise are shown.

We would expect that the MLE based algorithm would produce better results, as it seems that we treat the data in an intuitively better way (we maximize the likelihood of the data). This is indeed the case when we consider Cauchy noise, but the opposite happens when we test it with Gaussian noise. The decreasing gain algorithm is even slightly below the bound initially (which is possible only because the algorithm is initially biased). Thus, we cannot

<sup>1</sup>More precisely we used the MATLAB<sup>®</sup> function *fminsearch*. We chose this function instead of Newton's method because it can handle non-convex problems. For Cauchy noise the likelihood is not convex.

say that one of the algorithms is better than the other.

As the algorithms performance seems equivalent, a practical choice can be done in terms of complexity.

**Complexity.** At time  $k$ , the adaptive scheme based on MLE must solve a maximization problem using the last  $k$  measurements  $i_{1:k}$ . Each measurement produces an additional term on the log-likelihood to be maximized, thus at time  $k$ , the evaluation of the log-likelihood function itself requires  $k$  evaluations of the logarithm of the marginal likelihood. Note that the marginal likelihood can be very costly to be evaluated as it is a difference of CDF.

For the adaptive algorithm with decreasing gains, the gains can be precalculated and stored in a table, or they can be obtained by using one division, the update coefficients can also be precalculated and stored in a table. To generate one estimate the adaptive algorithm then requires: one search in a table to have the update coefficient, one division or one search in a table to have the gain, one multiplication to obtain the total correction and one sum to have the final estimate.

One can conclude that the adaptive algorithm with decreasing gains has far lower complexity requirements when compared with the scheme based on the MLE. Note also that the adaptive algorithm based on MLE needs a certain number of measurements with fixed (or not adaptive) thresholds to start. This is due to the fact that the MLE for one measurement is ill defined and produces estimates equal to  $+\infty$  or  $-\infty$ . Note that it can also happen with more than one measurement, if all measurements are equal to  $+1$  or if they are all equal to  $-1$ . Thus, for the adaptive algorithm based on MLE we can have realizations with unbounded values and this will happen especially in the cases when the initial quantizer dynamics is far away from the parameter. Such behavior will not happen for the adaptive algorithm with decreasing gains as the update coefficients are bounded above (considering PDF with upper bounded  $\tilde{f}_d$  and lower bounded from zero  $\tilde{F}_d$ ). Therefore, for practical purposes the choice between them is clear, we should choose the algorithm with decreasing gains (a3).

### Adaptive algorithm vs PF

We compare now the adaptive algorithm (with fixed gain) and the PF procedure for tracking a Wiener process.

**Asymptotic performance for fast parameter evolution.** In this case, for any  $\sigma_w$ , the PF is known to be optimal if the number of particles tends to infinity. Thus, for a very large number of particles we expect the PF procedure to be as good as the adaptive algorithm.

**Asymptotic performance for slow parameter evolution.** When  $\sigma_w$  is small, the procedures have equivalent asymptotic performance. The PF is approximately unbiased, if we choose a sufficiently large number of particles and the adaptive procedure is asymptotically unbiased. Their asymptotic MSE is approximately  $\frac{\sigma_w}{\sqrt{I_q(0)}}$ . Thus, when  $\sigma_w$  is small, the differences, if they exist, will also occur in the transient performance.

**Transient performance.** Similarly to the constant case, we analyze the transient performance through simulation. We simulated both the adaptive algorithm and the PF for  $N_I = 8$  and asymptotically optimal uniform quantization. The parameter model was a Wiener process with increment standard deviation  $\sigma_w = 0.001$ , with initial standard deviation  $\text{Var}(X_0) = 0.1$  and with initial mean equal to zero. We simulated the algorithms both for Gaussian and Cauchy noise with  $\delta = 1$ . For obtaining the simulated transient MSE, 1000 samples were simulated 2500 times for each algorithm and each noise distribution. The initial estimate for both algorithms  $\hat{X}_0$  was set to zero in all the cases. We used 5000 particles in the PF and its resampling procedure was triggered each time the number of effective particles was below 50. The results are shown in Fig. 3.11 where the asymptotically optimal performance ( $\frac{\sigma_w}{\sqrt{I_q(0)}}$ ) for small  $\sigma_w$  is also presented.

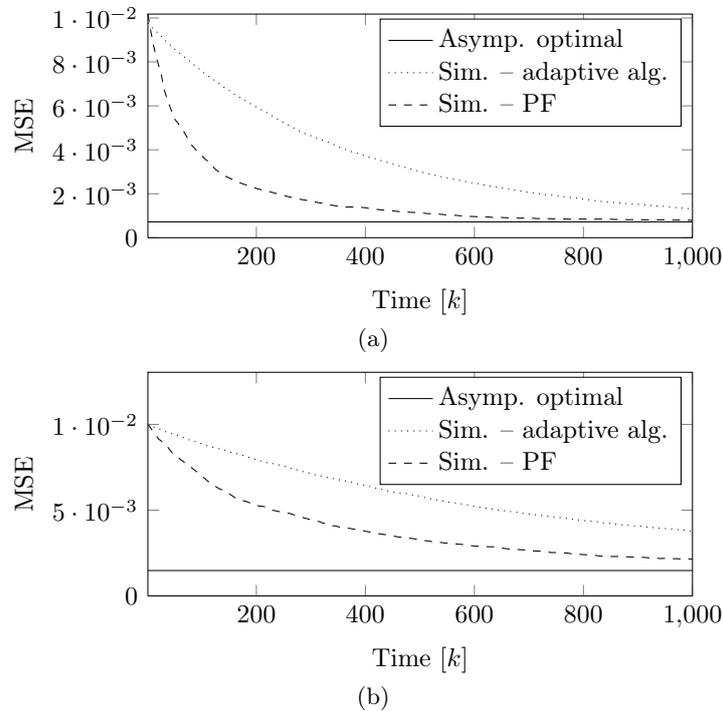


Figure 3.11: Asymptotic MSE for the optimal estimator of a Wiener process with small  $\sigma_w$  and simulated MSE for the adaptive algorithm with constant gain and for the PF with dynamic central threshold. Both algorithms were simulated with  $N_I = 8$ , optimal uniform thresholds, Gaussian and Cauchy noise with  $\delta = 1$ ,  $\sigma_w = 0.001$ ,  $\mathbb{E}(X_0) = 0$ ,  $\text{Var}(X_0) = 0.1$  and  $\hat{X}_0 = 1$ . The evaluation of the transient MSE was done with 2500 simulations of the algorithms for blocks with 1000 samples. The PF was simulated with 5000 particles and its threshold for the resampling procedure was set at  $N_{\text{thresh}} = 50$ . In (a) results for Gaussian noise are shown, while in (b) we have the results for Cauchy noise.

In this case the expected results are obtained. The PF, which might be close to optimal when the number of particles is large, is clearly faster to converge when compared with the adaptive algorithm.

**Complexity.** When comparing the complexity of the algorithms the difference is impressive.

At time  $k$ , for each particle in the PF, a Gaussian r.v. has to be simulated in the prediction step and its likelihood has to be evaluated. After that, the weighted mean of the particles is computed. It is then followed by the evaluation of the effective number of particles with a possible resampling step.

For the adaptive algorithm the complexity is one search in a table, to obtain the update coefficient, one multiplication with the constant gain and one sum with the previous estimate.

Therefore, one might choose the PF whenever there is no restriction on the complexity of the algorithm<sup>2</sup>. If there is a strong complexity restriction, by paying the price of a slower convergence, the adaptive algorithm can be a good solution.

### 3.5.5 Discussion on the results

We summarize the main points observed until now and we will discuss some of them.

- We proposed a low complexity adaptive algorithm to track 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 the quantizer is also symmetric with its center placed on the previous parameter estimate, we could prove by using Lyapunov theory that the algorithm is asymptotically unbiased for the estimation of a constant and of a Wiener process. We showed that the asymptotic performance for the optimal update coefficients is a function of the FI  $I_q(0)$ , which shows that this function plays an important role in the choice of the threshold variations, as it was also observed in Ch. 1 and 2.
- For the optimal update coefficients, the adaptive algorithm that is obtained is a generalization of the recursive algorithm found at the end of Ch. 1, being exactly equal if we constrain  $N_I = 2$ .

In the case of estimating a Wiener process, the adaptive algorithm with optimal update coefficients is equal to the asymptotic recursive algorithm presented at the end of Ch. 2. Therefore, the adaptive algorithm is a low complexity alternative to the algorithms presented in Ch. 1 and 2 with equivalent asymptotic performance.

- For testing the results, we considered two different families of noises, generalized Gaussian noises and Student's-t noises, both tested with uniform quantization. First, we evaluated the theoretical loss of performance due to quantization w.r.t. the continuous measurement equivalent estimator 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 measurement 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.

---

<sup>2</sup>Note that the number of particles necessary to have close to optimal performance can be reduced by using the optimal proposal distribution, thus reducing complexity. This can have an impact on the choice of the algorithm when the restriction on complexity is not strong.

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 MSE it is possible to estimate jointly the drift.

- As the algorithms are asymptotically equivalent in performance to the adaptive scheme based on the MLE in the constant case and to the PF in the Wiener process case, we simulated their transient performance, to see if we lose in performance and how much we lose by using the low complexity approach.

In the constant case, we cannot say that the adaptive scheme based on the MLE is better, thus in practice, the adaptive algorithm with decreasing gain might be used as it requires far lower complexity.

In the Wiener process case, the PF is superior to the adaptive algorithm with constant gain, thus if no complexity constraints are considered, we might use the PF. If we have strong complexity constraints, by accepting a slower convergence, the adaptive algorithm gives a good solution.

- An interesting link between standard quantization and the adaptive algorithm for tracking the Wiener process can be observed. In the binary case, the adaptive algorithm proposed here is similar to delta modulation [Gersho 1992, p. 214], the difference is that here we do not use the quantization noise approach for obtaining its performance and we also consider the effect of the measurement noise on the final performance.

When  $N_I > 2$  the algorithm that we propose can be seen as a form of predictive quantization intended for estimation and not for reconstruction of the measurements.

- Another interesting result is that a varying parameter has a loss of performance due to quantization smaller than the loss for a constant parameter, thus a type of dithering effect seems to be present. In this case, the variations of the input signal brings the tracking performance of the estimator closer 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. An approach of this type will be presented in Subsec. 3.6.2.
- 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.

We will study the joint estimation of the constant  $x$  and the scale parameter in Subsec. 3.6.1. The joint estimation of  $\sigma_w$  in the Wiener process case will lead to a scheme similar to delta modulation with variable gain, this is left for future work.

### 3.6 Adaptive quantizers for estimation: extensions to location-scale estimation and to the multiple sensor approach

We present now the two extensions discussed in the previous section.

The first extension that will be presented is the joint estimation of the unknown noise scale factor. We will see that the adaptive estimation of  $x$  does not change, the only thing that changes is the addition of the adaptive estimator of the scale parameter  $\delta$ . We will also see that the fact that we do not know the scale parameter value does not degrade the asymptotic estimation performance when compared with the location-only estimation problem.

We then present the multiple sensor approach based on a fusion center architecture. We will see that the optimal correction of the adaptive algorithm based on multiple quantized measurements from different sensors will be simply a weighted sum of their corrections in the single sensor case.

#### 3.6.1 Joint estimation of location and scale parameters

We start by stating the problem and defining the adaptive estimator. In a second step, we look for its performance and we optimize the algorithm in a similar way as it was done previously. We find the optimal adaptive gain, *i.e.* the optimal adaptive gain matrix. The optimal update coefficients are obtained in a third step. At the end of the section, we present some simulations and we discuss the results.

##### Problem statement and estimator

We consider that a sequence of i.i.d. r.v.  $Y_k$  with marginal CDF  $F_n\left(\frac{y-x}{\delta}\right)$  are quantized with an adjustable quantizer ( $F_n(\cdot)$  is the noise CDF for  $\delta = 1$ ), resulting in a sequence of discrete measurements  $i_{1:k}$ . The pair of parameters  $(x, \delta)$  is unknown and the objective is to estimate it based on the quantized measurements. This is equivalent to the following modification of problem (a) (p. 27):

**(a') Solve problem (a) when the noise scale parameter  $\delta$  is unknown and must be estimated jointly with  $x$ .**

Observe that this problem is a joint location-scale estimation based on quantized measurements.

The adjustable quantizer is given by (3.1), where for enhancing the estimation performance, we set the offset and the input gain to be

$$b_k = \hat{X}_{k-1}, \quad \Delta_k = c_{\Delta} \hat{\delta}_{k-1}. \quad (3.76)$$

Note that the main difference with the adjustable quantizer used previously is the use of the last scale parameter estimate for setting the input gain.

The adaptive estimation algorithm can be extended to include the joint estimation of the scale parameter. The extended version is

$$\begin{bmatrix} \hat{X}_k \\ \hat{\delta}_k \end{bmatrix} = \begin{bmatrix} \hat{X}_{k-1} \\ \hat{\delta}_{k-1} \end{bmatrix} + \frac{\mathbf{\Gamma}}{k} \hat{\delta}_{k-1} \begin{bmatrix} \eta_x(i_k) \\ \eta_\delta(i_k) \end{bmatrix}, \quad (3.77)$$

where  $\mathbf{\Gamma}$  is a  $2 \times 2$  matrix of gains,  $\eta_x[i]$  and  $\eta_\delta[i]$  are sequences of  $N_I$  update coefficients  $\left\{ \eta_x \left[ -\frac{N_I}{2} \right], \dots, \eta_x \left[ \frac{N_I}{2} \right] \right\}$  and  $\left\{ \eta_\delta \left[ -\frac{N_I}{2} \right], \dots, \eta_\delta \left[ \frac{N_I}{2} \right] \right\}$ .

The advantages of this extended version are the following:

- it is still a low complexity algorithm, requiring only a few operations more than the initial adaptive algorithm.
- It is an online algorithm. Making it possible for real-time applications to have access to the recent estimates at any time  $k$ .
- Its performance can also be studied using the general results from [Benveniste 1990].

The noise and quantizer follow the assumptions AN1–AQ1, AQ2' and AN3. For simplification purposes and to have a stable algorithm, we will assume that both  $\eta_x[i]$  and  $\eta_\delta[i]$  are symmetric,  $\eta_\delta[i]$  have even symmetry with negative<sup>3</sup>  $\eta_\delta[1] = \eta_\delta[-1]$ , while  $\eta_x[i]$  are defined with odd symmetry and they are positive for positive  $i$ , similarly as stated in AQ3.

**Assumption (on the quantizer output levels):**

**AQ3'** The quantizer output levels  $\eta_x[i]$  are odd and the output levels  $\eta_\delta[i]$  are even.

$$\eta_x[i] = -\eta_x[-i], \quad \eta_\delta[i] = \eta_\delta[-i], \quad (3.78)$$

with  $\eta_x[i] > 0$  for  $i > 0$  and  $\eta_\delta[1] < 0$ .

The estimation scheme is depicted in Fig. 3.12, where the UPDATE block is the estimation algorithm.

---

<sup>3</sup>This constraint on  $\eta_\delta[1]$  is imposed to guarantee the convergence of  $\hat{\delta}_k$ . The idea here is that when the quantized measurements are small, it means asymptotically (when  $\hat{X}_k$  is close to  $x$ ) that the quantizer range is too large, thus the range and, consequently,  $\hat{\delta}_k$  must be reduced. If we set the coefficients with the opposite sign,  $\hat{\delta}_k$  will diverge.

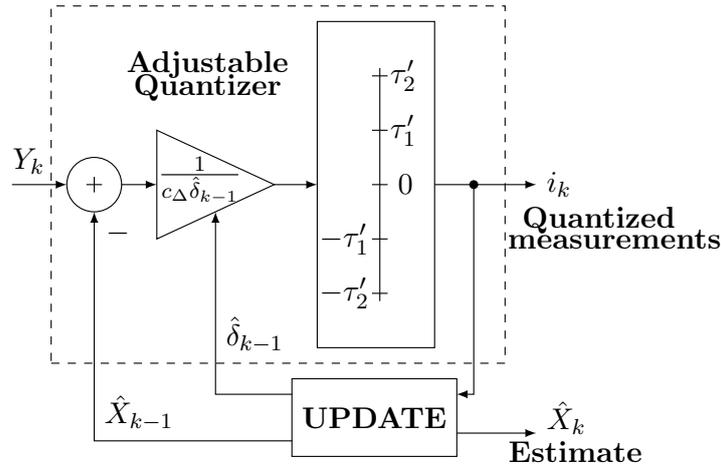


Figure 3.12: Scheme representing the adjustable quantizer. The offset and gain are adjusted dynamically using the estimates while the quantizer thresholds (the threshold variations) are fixed.

### Optimal parameters and performance

The analysis of the algorithm will be done using the results from [Benveniste 1990, Ch. 3]. We will analyze the bias and the asymptotic covariance matrix of the estimation error.

Similarly to the estimation of the constant location parameter, the algorithm mean can be approximated by the solution of an ODE. However, in this case, we have a vectorial ODE with one component for  $\hat{x}$  and one component for  $\hat{\delta}$ :

$$\frac{d}{dt} \begin{bmatrix} \hat{x} \\ \hat{\delta} \end{bmatrix} = \mathbf{G} \mathbf{h}(\hat{x}, \hat{\delta}). \quad (3.79)$$

The relation between continuous and discrete time is  $t_k = \sum_{j=1}^k \frac{1}{j}$  and  $\mathbf{h}$  is the following mean vector field:

$$\begin{aligned} \mathbf{h}(\hat{x}, \hat{\delta}) &= \mathbb{E} \begin{bmatrix} \hat{\delta} \eta_x \left( Q \left( \frac{y - \hat{x}}{c_{\Delta} \hat{\delta}} \right) \right) \\ \hat{\delta} \eta_{\delta} \left( Q \left( \frac{y - \hat{x}}{c_{\Delta} \hat{\delta}} \right) \right) \end{bmatrix} = \\ &= \hat{\delta} \begin{bmatrix} \sum_{i=1}^{\frac{N_I}{2}} \eta_x [i] \left\{ \tilde{F}_d(i, \hat{x}, x, \hat{\delta}, \delta) - \tilde{F}_d(-i, \hat{x}, x, \hat{\delta}, \delta) \right\} \\ \sum_{i=1}^{\frac{N_I}{2}} \eta_{\delta} [i] \left\{ \tilde{F}_d(i, \hat{x}, x, \hat{\delta}, \delta) + \tilde{F}_d(-i, \hat{x}, x, \hat{\delta}, \delta) \right\} \end{bmatrix}, \end{aligned} \quad (3.80)$$

where the expectation is w.r.t. to the noise marginal probability measure, the second equality comes from the symmetry assumptions and  $\tilde{F}_d$  is

$$\tilde{F}_d = \begin{cases} F_n \left( \frac{\tau_i c_{\Delta} \hat{\delta}}{\delta} + \frac{\hat{x} - x}{\delta} \right) - F_n \left( \frac{\tau_{i-1} c_{\Delta} \hat{\delta}}{\delta} + \frac{\hat{x} - x}{\delta} \right), & \text{if } i \in \left\{ 1, \dots, \frac{N_I}{2} \right\}, \\ F_n \left( \frac{\tau_{i+1} c_{\Delta} \hat{\delta}}{\delta} + \frac{\hat{x} - x}{\delta} \right) - F_n \left( \frac{\tau_i c_{\Delta} \hat{\delta}}{\delta} + \frac{\hat{x} - x}{\delta} \right), & \text{if } i \in \left\{ -1, \dots, -\frac{N_I}{2} \right\}. \end{cases} \quad (3.81)$$

The conditions on the mean convergence of the algorithm are then conditions on the global asymptotic stability of the point  $\hat{x} = x$  and  $\hat{\delta} = \delta$ . One necessary condition for asymptotic stability is that the true parameters must be an equilibrium point of the ODE, which means that  $\mathbf{h}(\hat{x} = x, \hat{\delta} = \delta)$  must be zero. From the symmetry assumptions:

$$\mathbf{h}(\hat{x} = x, \hat{\delta} = \delta) = \begin{bmatrix} 0 \\ 2\boldsymbol{\eta}_\delta^\top \mathbf{F}_d^{vec} \end{bmatrix},$$

where the vector  $\mathbf{F}_d^{vec}$  is

$$\mathbf{F}_d^{vec} = \left[ \tilde{F}_d[1] \cdots \tilde{F}_d \left[ \frac{N_I}{2} \right] \right]^\top,$$

with elements  $F_d[i] = F_d(i, x, x, \delta, \delta)$  independent of the parameters. Then, the condition for the parameters to be the equilibrium point is

$$\boldsymbol{\eta}_\delta^\top \mathbf{F}_d^{vec} = 0. \quad (3.82)$$

Other conditions are necessary for the mean convergence of the algorithm. These conditions can be found by the analysis of the ODE using Lyapunov theory. The analysis of these other conditions will not be detailed here and under the assumptions already stated and the constraint on  $\boldsymbol{\eta}_\delta$  given in (3.82), it will be assumed that the algorithm converges in the mean to the true parameters.

We turn our attention now to the asymptotic fluctuation of the algorithm, which is given by its asymptotic covariance matrix. Under the assumptions stated previously (assumptions AN1–AQ3' and the assumption that the algorithm is asymptotically unbiased), it can be shown [Benveniste 1990, pp. 110–113] that the normalized estimation error  $\sqrt{k}\boldsymbol{\varepsilon}_k$  tends in distribution to a zero mean Gaussian random variable as follows

$$\sqrt{k}\boldsymbol{\varepsilon}_k \underset{k \rightarrow \infty}{\rightsquigarrow} \mathcal{N}(0, \mathbf{P}), \quad (3.83)$$

where  $\mathbf{P}$  is the covariance matrix given by the optimal gain  $\boldsymbol{\Gamma}^*$ . The matrices  $\mathbf{P}$  and  $\boldsymbol{\Gamma}^*$  are the following:

$$\mathbf{P} = \frac{\delta^2}{2} \begin{bmatrix} \frac{\boldsymbol{\eta}_x^\top \mathbf{F}_d \boldsymbol{\eta}_x}{(\boldsymbol{\eta}_x^\top \mathbf{f}_d^{(x)})^2} & 0 \\ 0 & \frac{\boldsymbol{\eta}_\delta^\top \mathbf{F}_d \boldsymbol{\eta}_\delta}{(\boldsymbol{\eta}_\delta^\top \mathbf{f}_d^{(\delta)})^2} \end{bmatrix} \quad (3.84)$$

and

$$\boldsymbol{\Gamma}^* = -\frac{1}{2} \begin{bmatrix} \frac{1}{\boldsymbol{\eta}_x^\top \mathbf{f}_d^{(x)}} & 0 \\ 0 & \frac{1}{\boldsymbol{\eta}_\delta^\top \mathbf{f}_d^{(\delta)}} \end{bmatrix}, \quad (3.85)$$

where  $\mathbf{F}_d$  is a diagonal matrix  $\mathbf{F}_d = \text{diag}[\mathbf{F}_d^{vec}]$ ,  $\mathbf{f}_d^{(x)} = [\tilde{f}_d^{(x)}[1] \cdots \tilde{f}_d^{(x)}[\frac{N_I}{2}]]^\top$  and  $\mathbf{f}_d^{(\delta)} = [\tilde{f}_d^{(\delta)}[1] \cdots \tilde{f}_d^{(\delta)}[\frac{N_I}{2}]]^\top$  are the derivatives in vector form of the quantizer output probabilities  $\tilde{F}_d(i, \hat{x}, x, \hat{\delta}, \delta)$  multiplied by  $\hat{\delta}$  when  $\hat{x} = x$  and  $\hat{\delta} = \delta$ :

$$\tilde{f}_d^{(x)}[i] = f_n(\tau_i) - f_n(\tau_{i-1}), \quad (3.86)$$

$$\tilde{f}_d^{(\delta)}[i] = c_\Delta [\tau_i f_n(\tau_i) - \tau_{i-1} f_n(\tau_{i-1})]. \quad (3.87)$$

These results are obtained in an equivalent way as the results presented for the estimation of  $x$ . But in this case  $\mathbf{\Gamma}^*$  is not the inverse of the scalar derivative of  $-h$ , but instead is the inverse of the Jacobian matrix of  $-\mathbf{h}$  evaluated at the point  $(\hat{x} = x, \hat{\delta} = \delta)$ . In the same way, the normalized covariance for the optimal gain is the normalized covariance of the vector of corrections  $\begin{bmatrix} \eta_x(i_k) \\ \eta_\delta(i_k) \end{bmatrix}$  pre and post-multiplied by the inverse of the Jacobian of  $\mathbf{h}$ , with all the factors being evaluated at  $(\hat{x} = x, \hat{\delta} = \delta)$ . The specific diagonal pattern of the  $\mathbf{\Gamma}^*$  and  $\mathbf{P}$  comes from the symmetry assumptions on the noise and the quantizer.

Minimization of the estimation variance can be done through the minimization of the diagonal terms of  $\mathbf{P}$  w.r.t.  $\boldsymbol{\eta}_x$  and  $\boldsymbol{\eta}_\delta$ . The two minimization problems can be solved separately. In the case of the optimization w.r.t.  $\boldsymbol{\eta}_\delta$ , the equilibrium constraint (3.82) has to be taken into account. The optimal  $\boldsymbol{\eta}_x$  can be found by using the Cauchy-Schwarz inequality, while the optimal  $\boldsymbol{\eta}_\delta$  are obtained by casting the constrained minimization problem as a modified eigenvalue problem solved in [Golub 1973] (Why? - App. A.1.9).

The optimal coefficients are

$$\begin{aligned} \boldsymbol{\eta}_x &\propto \mathbf{F}_d^{-1} \mathbf{f}_d^{(x)}, \\ \boldsymbol{\eta}_\delta &\propto \mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)} - \mathbf{1} \mathbf{f}_d^{(\delta)} = \mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)}, \end{aligned}$$

where  $\mathbf{1}$  is a squared matrix with ones. The second equality comes from the fact that the sum of  $\mathbf{f}_d^{(\delta)}$  is zero. To respect the assumptions we can set

$$\boldsymbol{\eta}_x = -\mathbf{F}_d^{-1} \mathbf{f}_d^{(x)}, \quad (3.88)$$

$$\boldsymbol{\eta}_\delta = -\mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)}. \quad (3.89)$$

Therefore, the optimal  $\mathbf{P}$  and  $\mathbf{\Gamma}^*$  are

$$\mathbf{P} = \delta^2 \mathbf{\Gamma}^* = \frac{\delta^2}{2} \begin{bmatrix} \frac{1}{\mathbf{f}_d^{(x)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(x)}} & 0 \\ 0 & \frac{1}{\mathbf{f}_d^{(\delta)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)}} \end{bmatrix}. \quad (3.90)$$

Note that the asymptotic variances are equal to the CRB for estimating the parameters based on the quantized measurements, when the quantizer offset and input gain are placed exactly at  $x$  and  $\frac{1}{c_\Delta \delta}$ .

We have the following solution to problem (a') (p. 149):

**Solution to (a') - Adaptive algorithm with decreasing gain  
for estimating  $x$  and  $\delta$**

**(a'1) 1) Estimator**

For each time  $k$ , the estimate, the quantizer offset and the quantizer input gain are obtained using (3.77)

$$\begin{bmatrix} \hat{X}_k \\ \hat{\delta}_k \end{bmatrix} = \begin{bmatrix} \tau_{0,k} \\ \frac{\Delta_k}{c_\Delta} \end{bmatrix} = \begin{bmatrix} \hat{X}_{k-1} \\ \hat{\delta}_{k-1} \end{bmatrix} + \frac{\mathbf{\Gamma}}{k} \hat{\delta}_{k-1} \begin{bmatrix} \eta_x(i_k) \\ \eta_\delta(i_k) \end{bmatrix},$$

with  $i_k = Q\left(\frac{Y_k - \hat{X}_{k-1}}{c_\Delta \hat{\delta}_{k-1}}\right)$ ,  $\mathbf{\Gamma} = \frac{1}{2} \begin{bmatrix} \frac{1}{\mathbf{f}_d^{(x)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(x)}} & 0 \\ 0 & \frac{1}{\mathbf{f}_d^{(\delta)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)}} \end{bmatrix}$  and

$$\begin{bmatrix} \eta_x(i_k) \\ \eta_\delta(i_k) \end{bmatrix} = \begin{bmatrix} -\frac{\tilde{f}_d^{(x)}[i_k]}{\tilde{F}_d[i_k]} \\ -\frac{\tilde{f}_d^{(\delta)}[i_k]}{\tilde{F}_d[i_k]} \end{bmatrix}.$$

**2) Performance (assumed and asymptotic)**

The estimator is assumed to be asymptotically unbiased. When  $k \rightarrow \infty$  the normalized estimation error vector  $\sqrt{k}\boldsymbol{\varepsilon}_k$  is Gaussian distributed with covariance matrix  $\mathbf{P}$  given by (3.90)

$$\mathbf{P} = \frac{\delta^2}{2} \begin{bmatrix} \frac{1}{\mathbf{f}_d^{(x)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(x)}} & 0 \\ 0 & \frac{1}{\mathbf{f}_d^{(\delta)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)}} \end{bmatrix}.$$

Observe also that the asymptotic performance can still be optimized through  $\boldsymbol{\tau}'$  and  $c_\Delta$ . As optimization through  $\boldsymbol{\tau}'$  is difficult, in the simulation section we will consider again that the thresholds variations are uniform as in (3.50)

$$\boldsymbol{\tau}' = \left[ -\tau'_{\frac{N_I}{2}} = -\infty \cdots -\tau'_1 = -1 \quad 0 \quad +\tau'_1 = +1 \cdots +\tau'_{\frac{N_I}{2}} = +\infty \right]^\top,$$

thus the only free parameter for optimization is  $c_\Delta$ .

### Simulations

The algorithm will be simulated to validate the theoretical results. The simulation will be focused on the performance for the estimation of  $x$ . As it was mentioned, the quantizer is uniform and  $c_\Delta$  will be chosen so as to minimize the variance of estimation of  $x$ . As this is a scalar problem, it can be solved by an exhaustive search using a fine grid. After finding the

optimal  $c_\Delta$ , the other parameters of the algorithm  $\mathbf{\Gamma}$ ,  $\boldsymbol{\eta}_x$  and  $\boldsymbol{\eta}_\delta$  can be evaluated using the information from the noise distribution.

The Gaussian and Cauchy distribution will be used for modeling the noise. The algorithm will be simulated for  $5 \times 10^5$  blocks with  $4 \times 10^4$  samples each. The simulated MSE for the estimation of the location parameter will be evaluated by calculating the mean of the squared error for each sample. Other simulation parameters are  $\delta = 1$ ,  $\hat{\delta}_0 = 2$ ,  $x = 0$ ,  $\hat{X}_0 = 1$  and  $N_I \in \{4, 8, 16, 32\}$ . For comparison purposes, the CRB for the estimation of  $x$  based on continuous measurements  $\text{CRB}_c$  will be also evaluated for Gaussian and Cauchy distributions. Using the fact that the measurements are independent and the expressions for  $I_c$  for the GGD given in (3.74) with  $\beta = 2$  and for the STD given in (3.75) with  $\beta = 1$ , the  $\text{CRB}_c$  for Gaussian and Cauchy noise are respectively  $\frac{1}{2} \frac{\delta^2}{k}$  and  $2 \frac{\delta^2}{k}$ .

The results of the simulation are shown in Fig. 3.13, where we also plotted the CRB for the estimation with quantized measurements when the offset and gain are static and set with the true parameter values. The MSE was normalized by  $k$  and the logarithm scale is used in both axis for better visualization.

It can be observed that after a transient time, the simulated performance becomes very close to the asymptotic theoretical results, also it can be seen that the gain in performance when increasing  $N_I$  is very small even for a small number of quantization intervals ( $N_I = 8$  or 16) and that the gap between the performance given by  $N_I = 32$  and the continuous measurement bound is negligible.

### Discussion on the results

Despite the very low complexity of the algorithm, its asymptotic performance for estimating the parameters is not only decoupled (the covariance is diagonal) but it is also optimal. The normalized asymptotic variance for estimating  $x$  is  $\frac{1}{I_q(0)}$  and the variance for estimating  $\delta$  is also the inverse of the corresponding FI. This optimal decoupling means that no degradation of performance is brought by estimating jointly the scale parameter. As no degradation is present, the asymptotic performance of the estimator of  $x$  has the same behavior as it was shown previously, if we choose  $N_I = 4$  or 5 the estimation performance is very close to the optimal continuous measurement performance. This indicates that even when  $\delta$  is unknown there is no need to use high resolution quantizers, if we have a large number of samples.

### 3.6.2 Fusion center approach with multiple sensors

We present now the adaptive algorithm for estimating a constant parameter, when a fusion center has access to quantized measurements from multiple sensors. We will define first, the problem, the architecture to be used and the adaptive estimator. Then, similarly to the joint location-scale problem, we obtain the algorithm performance and its optimal parameters. We close this section with simulations and a discussion about our results.

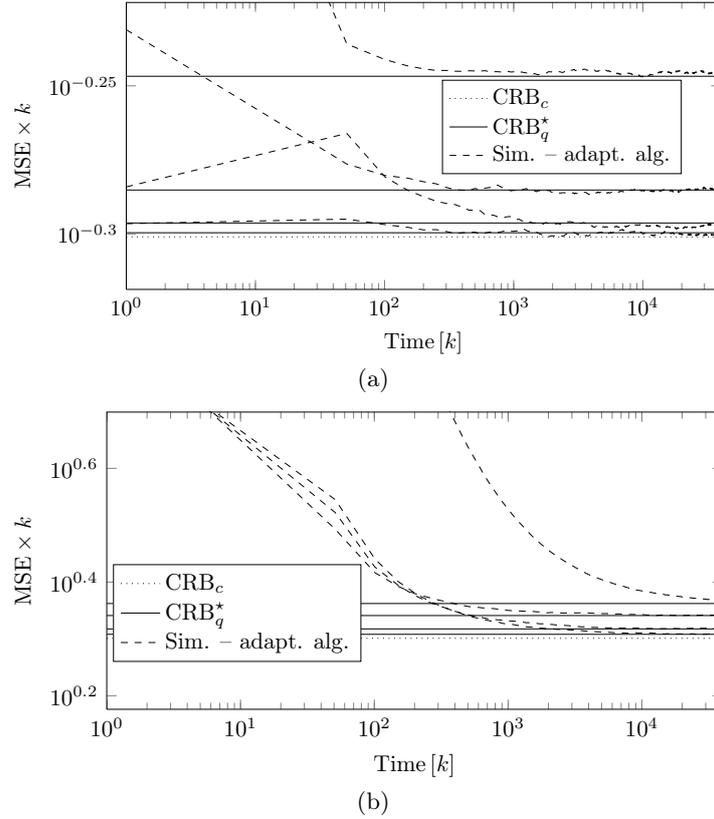


Figure 3.13: CRB for estimating a location parameter of Gaussian and Cauchy distributions based on quantized and continuous measurements and simulated MSE for the estimation of the location parameter with the adaptive location-scale parameter estimator. In all cases, we considered the true scale parameter and its initial estimate  $\delta = 1$ ,  $\hat{\delta}_0 = 2$ , for the location parameter we considered  $x = 0$ ,  $\hat{X}_0 = 1$ . The numbers of quantization intervals simulated were  $N_I \in \{4, 8, 16, 32\}$ . For obtaining the simulated MSE for the location parameter, the algorithm was simulated for  $5 \times 10^5$  blocks with  $4 \times 10^4$  samples each. The curves that are asymptotically lower are related to a higher number of quantization intervals.

### Problem statement and estimator

The scalar parameter is supposed to be a constant  $x$  and it is measured by  $N_s$  sensors. Each sensor measures the parameter with additive noise

$$Y_k^{(j)} = x + V_k^{(j)}, \quad \text{for } j \in \{1, \dots, N_s\}, \quad (3.91)$$

where  $V_k^{(j)}$  is the noise r.v. for the sample  $k$  obtained at the sensor  $j$ . The sensor noises are independent and each sensor noise is i.i.d.. The noise r.v. also respects assumptions AN1, AN2 and AN3. Its marginal CDF for sample  $k$  of sensor  $j$  will be denoted as  $F^{(j)}(v)$  and its PDF as  $f^{(j)}(v)$ .

The measurements at each sensor are quantized by a scalar adjustable quantizer, similar to the quantizer used in the previous sections. The quantizers for the sensors are then characterized by their input gains  $\frac{1}{\Delta_k^{(j)}}$ , input offsets  $b_k^{(j)}$  and the vector of threshold variations

(considered to be static) that defines the  $N_I^{(j)}$  quantizer intervals

$$\boldsymbol{\tau}^{(j)} = \left[ \tau_{-\frac{N_I}{2}}^{(j)} \cdots \tau_{-1}^{(j)} \tau_0^{(j)} \tau_1^{(j)} \cdots \tau_{\frac{N_I}{2}}^{(j)} \right].$$

We will consider again the following assumptions:

- AQ1 on the quantizer outputs: the set of possible quantizer outputs of the sensor  $j$  is  $\mathcal{I}^{(j)} = \left\{ -\frac{N_I^{(j)}}{2}, \dots, -1, 1, \dots, \frac{N_I^{(j)}}{2} \right\}$ .
- AQ2 on the quantizer threshold variations: the quantizers have symmetric threshold variations  $\tau_i^{(j)} = -\tau_{-i}^{(j)}$  with  $\tau_0^{(j)} = 0$  and  $\tau_{\frac{N_I}{2}}^{(j)} = +\infty$ .

The output of quantizer  $j$  is then given by

$$i_k^{(j)} = Q^{(j)} \left( \frac{Y_k^{(j)} - b_k^{(j)}}{\Delta_k^{(j)}} \right) = i \operatorname{sign} \left( Y_k^{(j)} - b_k^{(j)} \right), \text{ for } \frac{|Y_k^{(j)} - b_k^{(j)}|}{\Delta_k^{(j)}} \in \left[ \tau_{i-1}^{(j)}, \tau_i^{(j)} \right). \quad (3.92)$$

The noise CDF are considered also to have a known scale parameter  $\delta^{(j)}$ . Therefore, similarly to what was done before, we can use the noise scale factor to normalize the input of the quantizer

$$\Delta_k^{(j)} = c_\Delta^{(j)} \delta^{(j)}, \quad (3.93)$$

where  $c_\Delta$  is a free parameter which, as it was explained before, can be used to adjust the quantizer input range or to optimize quantization performance when the threshold variations are fixed.

After obtaining the quantized measurements, the sensors send their measurements to a fusion center. The transmission of the quantized measurements is supposed to be perfect, as it was explained in the Introduction. The fusion center can feedback information to the sensors through perfect continuous amplitude channels. Thus, we want to solve the following modification of problem (a) (p. 27):

**(a'') Solve problem (a) with independent quantized measurements from  $N_s$  sensors. The measurements from the  $N_s$  sensors are available at a fusion center that can process these measurements and feedback information to the sensors through perfect continuous amplitude channels.**

Note that the simplifying assumption of perfect feedback channels means that the fusion center has enough power and/or band for feedbacking real (or very finely quantized) noiseless estimates.

To solve problem (a''), the fusion center generates an online estimate  $\hat{X}_k$  that will be broadcasted to the quantizers through the feedback channels, so that they can use it as their next input offset for enhancing estimation performance. At time  $k$ , this means that

$$b_k^{(j)} = \hat{X}_{k-1}. \quad (3.94)$$

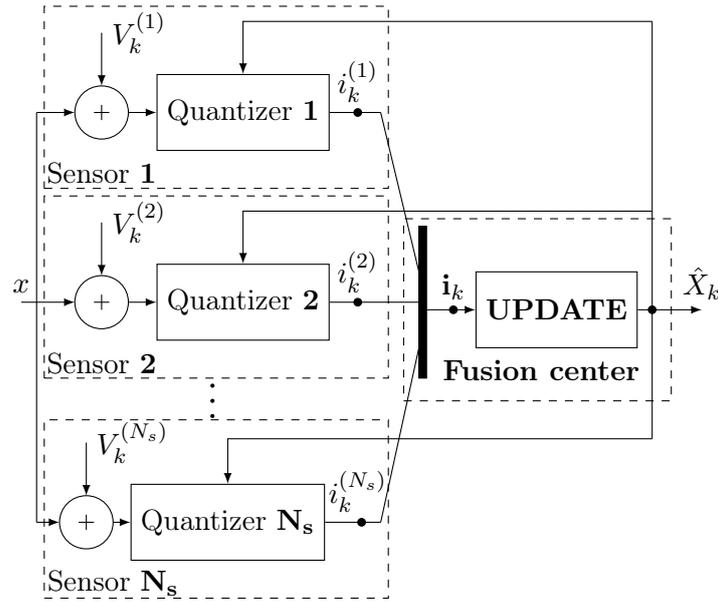


Figure 3.14: Scheme representing the sensor network. The fusion center updates the estimate of the parameter and broadcasts it through a perfect channel to the sensors. The sensors then use the new estimate as their quantizer input offset (their quantizer central threshold).

The general scheme is depicted in Fig. 3.14, where the **UPDATE** block contains an online estimator of the parameter.

For estimating the parameter, we can use an extension of the adaptive algorithm with decreasing gains

$$\hat{X}_k = \hat{X}_{k-1} + \frac{\gamma}{k} \eta(\mathbf{i}_k), \quad (3.95)$$

where  $\gamma$  is a positive gain,  $\mathbf{i}_k$  is the vector of quantized observations  $[i_k^{(1)} \dots i_k^{(N_s)}]^T$  and  $\eta[\mathbf{i}]$  is the update coefficient (or the quantizer output level) defined as a function from  $\{\mathcal{I}^{(1)}, \dots, \mathcal{I}^{(N_s)}\}$  to  $\mathbb{R}$ . The main advantage of this algorithm when compared with an adaptive scheme based on the MLE is its low complexity both in terms of processing and memory requirements.

### Optimal parameters and performance

Using the results from [Benveniste 1990, Ch. 3], the asymptotic variance of the estimation error can be obtained under the condition that the mean error converges to zero as  $k \rightarrow \infty$ . To prove this convergence, it would be sufficient to use the ODE approximation of the mean of  $\hat{X}_k$  and then prove global convergence properties for the ODE using Lyapunov theory. Such analysis is left for future work. Here, only the mean behavior of the algorithm at equilibrium ( $\hat{X}_k = x$ ) will be studied.

When  $\hat{X}_{k-1} = x$ , the normalized mean increment  $\frac{k}{\gamma} E \left( \hat{X}_k - \hat{X}_{k-1} \right)$  is given by

$$\frac{k}{\gamma} E \left( \hat{X}_k - \hat{X}_{k-1} \right) = \mathbb{E} [\eta(\mathbf{i})] = \boldsymbol{\eta}^T \mathbf{F}_d^{vec}, \quad (3.96)$$

where  $\boldsymbol{\eta}$  is a vector regrouping all possible values of the output coefficients

$$\boldsymbol{\eta} = \left[ \eta \left( i_{-\frac{N_I^{(1)}}{2}}, \dots, i_{-\frac{N_I^{(N_s)}}{2}} \right) \cdots \eta \left( i_{\frac{N_I^{(1)}}{2}}, \dots, i_{\frac{N_I^{(N_s)}}{2}} \right) \right]^T$$

and  $\mathbf{F}_d^{vec} = \left[ \cdots \tilde{F}_d[\mathbf{i}] \cdots \right]^T$  with

$$\tilde{F}_d[\mathbf{i}] = \prod_{j=1}^{N_s} \tilde{F}_d^{(j)} [i^{(j)}], \quad (3.97)$$

where  $\tilde{F}_d^{(j)} [i^{(j)}]$  is the probability of having the output  $i^{(j)}$  at the sensor  $j$  when  $\hat{X}_k = x$ :

$$\tilde{F}_d^{(j)} [i^{(j)}] = \begin{cases} F^{(j)} \left( \tau_i^{(j)} c_\delta^{(j)} \delta^{(j)} \right) - F^{(j)} \left( \tau_{i-1}^{(j)} c_\delta^{(j)} \delta^{(j)} \right), & \text{if } i^{(j)} \in \left\{ 1, \dots, \frac{N_I^{(j)}}{2} \right\}, \\ F^{(j)} \left( \tau_{i+1}^{(j)} c_\delta^{(j)} \delta^{(j)} \right) - F^{(j)} \left( \tau_i^{(j)} c_\delta^{(j)} \delta^{(j)} \right), & \text{if } i^{(j)} \in \left\{ -1, \dots, -\frac{N_I^{(j)}}{2} \right\}. \end{cases} \quad (3.98)$$

Thus, the following condition is needed to have an equilibrium point at the true parameter:

$$\boldsymbol{\eta}^T \mathbf{F}_d^{vec} = 0. \quad (3.99)$$

Note that this is a necessary condition for asymptotic unbiasedness of the algorithm.

Assuming that the algorithm is asymptotically unbiased, similarly to the single sensor case, we can use the results in [Benveniste 1990, pp. 110–113] to obtain the asymptotic distribution of the estimation error, the optimal gain  $\gamma^*$  and the minimum normalized asymptotic estimation error variance  $\sigma_\infty^2$ . The asymptotic estimation error is Gaussian distributed and it is given as follows

$$\sqrt{k} \varepsilon_k \underset{k \rightarrow \infty}{\rightsquigarrow} \mathcal{N} \left( 0, \sigma_\infty^2 \right). \quad (3.100)$$

The optimal  $\gamma$  and minimum  $\sigma_\infty^2$  are then given by

$$\gamma^* = -\frac{1}{\boldsymbol{\eta}^T \mathbf{f}_d} \quad (3.101)$$

and

$$\sigma_\infty^2 = \frac{\boldsymbol{\eta}^T \mathbf{F}_d \boldsymbol{\eta}}{(\boldsymbol{\eta}^T \mathbf{f}_d)^2}. \quad (3.102)$$

The matrix  $\mathbf{F}_d$  is a diagonal matrix  $\text{diag}[\mathbf{F}_d^{vec}]$  and  $\mathbf{f}_d$  is the vector form (as  $\boldsymbol{\eta}$  and  $\mathbf{F}_d^{vec}$ ) regrouping the elements

$$\tilde{f}_d[\mathbf{i}] = \sum_{j=1}^{N_s} \tilde{f}_d [i^{(j)}] \prod_{\substack{j'=1 \\ j' \neq j}}^{N_s} \tilde{F}_d^{(j')} [i^{(j')}], \quad (3.103)$$

where

$$\tilde{f}_d^{(j)} [i^{(j)}] = \begin{cases} f^{(j)} \left( \tau_i^{(j)} c_\delta^{(j)} \delta^{(j)} \right) - f^{(j)} \left( \tau_{i-1}^{(j)} c_\delta^{(j)} \delta^{(j)} \right), & \text{if } i^{(j)} \in \left\{ 1, \dots, \frac{N_I^{(j)}}{2} \right\}, \\ f^{(j)} \left( \tau_{i+1}^{(j)} c_\delta^{(j)} \delta^{(j)} \right) - f^{(j)} \left( \tau_i^{(j)} c_\delta^{(j)} \delta^{(j)} \right), & \text{if } i^{(j)} \in \left\{ -1, \dots, -\frac{N_I^{(j)}}{2} \right\}. \end{cases} \quad (3.104)$$

The asymptotic performance can also be optimized through the choice of  $\boldsymbol{\eta}$ , this can be done by minimizing (3.102) w.r.t.  $\boldsymbol{\eta}$  under the equilibrium constraint (3.99). This problem can be solved in the same way as it was done for finding the optimal vector  $\boldsymbol{\eta}_\delta$  in the joint estimation of location and scale parameters. Consequently, we find the following optimal vector  $\boldsymbol{\eta}$  (Why? - App. A.1.9):

$$\boldsymbol{\eta} \propto \mathbf{F}_d^{-1} \mathbf{f}_d - \mathbf{1} \mathbf{f}_d = \mathbf{F}_d^{-1} \mathbf{f}_d.$$

The second equality comes from the fact that the sum of the elements of  $\mathbf{f}_d$  is zero. For proving this, note that for each possible  $\mathbf{i}$  there is  $-\mathbf{i}$ . As the function  $\tilde{f}_d [i]$  is odd and  $\tilde{F}_d [i]$  is even, we have  $\tilde{f}_d [\mathbf{i}] = -\tilde{f}_d [-\mathbf{i}]$ . Therefore, when adding  $\tilde{f}_d [\mathbf{i}]$  for all the possible  $\mathbf{i}$ , the pairs  $(\tilde{f}_d [\mathbf{i}], \tilde{f}_d [-\mathbf{i}])$  will cancel each other, resulting in a zero sum. Similarly to the previous cases we will choose

$$\boldsymbol{\eta} = -\mathbf{F}_d^{-1} \mathbf{f}_d. \quad (3.105)$$

For the update coefficients given by (3.105), the asymptotic normalized variance and the optimal gain are

$$\sigma_\infty^2 = \gamma^* = \frac{1}{\mathbf{f}_d^T \mathbf{F}_d^{-1} \mathbf{f}_d}. \quad (3.106)$$

Using the expressions for  $\tilde{F}_d [\mathbf{i}]$  (3.97) and for  $\tilde{f}_d [\mathbf{i}]$  (3.103), for a given measurement vector  $\mathbf{i}$  the update coefficients are<sup>4</sup>

$$\boldsymbol{\eta} (\mathbf{i}) = - \sum_{j=1}^{N_s} \frac{\tilde{f}_d^{(j)} [i^{(j)}]}{\tilde{F}_d^{(j)} [i^{(j)}]}. \quad (3.107)$$

If we use the symmetry assumptions, the expression for the asymptotic normalized variance and for the optimal gain (3.106) becomes (Why? - App. A.1.10)

$$\sigma_\infty^2 = \gamma^* = \frac{1}{\sum_{j=1}^{N_s} \sum_{i^{(j)} \in \mathcal{I}^{(j)}} \frac{\tilde{f}_d^{(j)2} [i^{(j)}]}{\tilde{F}_d^{(j)} [i^{(j)}]}}. \quad (3.108)$$

Observe that the update coefficients are the sum of the update coefficients obtained in the single sensor approach. The asymptotic normalized variance is equal to the inverse of the sum of the FI  $I_q (0)$  for each sensor, which means that the algorithm is asymptotically efficient.

---

<sup>4</sup>Using this specific form for the update coefficients, we can prove, similarly as it was done for the single sensor case, that the algorithm is asymptotically unbiased.

We have then the following solution to problem (a'') (p. 157):

**Solution to (a'') - Adaptive algorithm with decreasing gain for estimating  $x$  using multiple sensors and a fusion center**

**(a'') 1) Estimator**

For each time  $k$ ,

- the sensors send  $i_k^{(j)} = Q^{(j)} \left( \frac{Y_k^{(j)} - \hat{X}_{k-1}}{c_\Delta^{(j)} \delta^{(j)}} \right)$  to the fusion center.
- The fusion center estimates the parameter using (3.95)

$$\hat{X}_k = \hat{X}_{k-1} + \frac{\gamma}{k} \eta(\mathbf{i}_k),$$

where  $\gamma^* = \frac{1}{\sum_{j=1}^{N_s} \sum_{i^{(j)} \in \mathcal{I}^{(j)}} \frac{\tilde{f}_d^{(j)2}[i^{(j)}]}{\tilde{F}_d^{(j)}[i^{(j)]}}$  and  $\eta(\mathbf{i}) = - \sum_{j=1}^{N_s} \frac{\tilde{f}_d^{(j)}[i^{(j)}]}{\tilde{F}_d^{(j)}[i^{(j)]}$ .

- The fusion center then broadcasts the estimate to the sensors through perfect channels to be used as the next quantizers input offset.

**2) Performance (assumed and asymptotic)**

The estimator is assumed to be asymptotically unbiased. When  $k \rightarrow \infty$  the normalized estimation error  $\sqrt{k} \varepsilon_k$  is Gaussian distributed with variance  $\sigma_\infty^2$  given by (3.108)

$$\sigma_\infty^2 = \frac{1}{\sum_{j=1}^{N_s} \sum_{i^{(j)} \in \mathcal{I}^{(j)}} \frac{\tilde{f}_d^{(j)2}[i^{(j)}]}{\tilde{F}_d^{(j)}[i^{(j)]}}$$

Note that, again here, we can still optimize the performance through  $\tau'^{(j)}$  and  $c_\Delta^{(j)}$ . In the same way as it was done previously, in what follows we consider that the threshold variations are uniform with unitary step-length and that only  $c_\Delta^{(j)}$  are used for optimizing the performance.

### Simulations

The validity of the results will be verified through simulations. All the sensors within a simulation will be considered to have the same type of noise and the same noise scale factor  $\delta = 1$ . The noise considered will be Gaussian or Cauchy distributed. Optimization w.r.t.  $c_\Delta$  (the same gain for all sensors in this case, as the noise is identically distributed) will be done by searching the maximum of the corresponding FI in a fine grid. After finding the optimal  $c_\Delta$ , the coefficients  $-\frac{\tilde{f}_d[i]}{F_d[i]}$  and the gain  $\gamma^*$  can be calculated.

For all the following simulations, the length of the block of samples will be 5000 and for evaluating the MSE the average of the squared error will be calculated using  $5 \times 10^4$  blocks. The parameter value and initial estimator value are  $x = 0$  and  $\hat{X}_0 = 1$ .

In the first simulation, it will be considered that all the quantizers have  $N_I = 4$  and  $N_s$  will be 1, 2 or 3, the results can be observed in Fig. 3.15 in log scale both in time and MSE. The simulated results are compared with the theoretical approximations, for this algorithm they are asymptotically equal to the CRB for quantized measurements obtained from a number of sensors  $N_s$ ,  $\text{CRB}_q^{N_s, \star}$ .

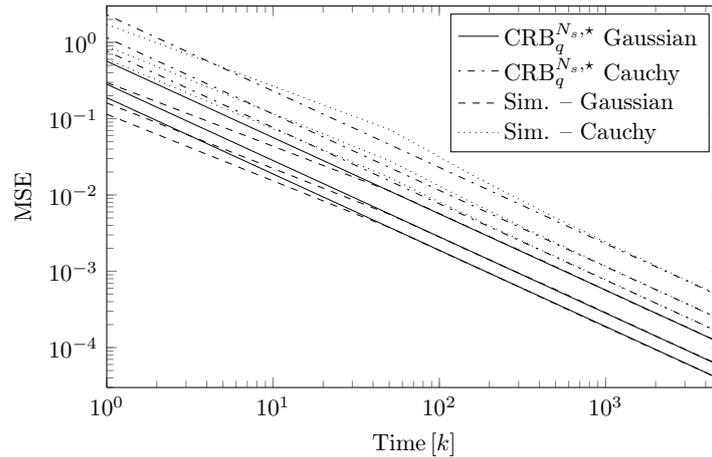


Figure 3.15: Cramér–Rao bound and simulated MSE for the adaptive algorithm when  $N_I = 4$ ,  $N_s = 1, 2, 3$  and the noise is Gaussian or Cauchy distributed, both with  $\delta = 1$ . For obtaining the simulated MSE, the algorithm was simulated  $5 \times 10^4$  times for blocks with 5000 samples. For all simulations the true parameter was set to zero and the initial estimate was  $\hat{X}_0 = 1$ . In each set of curves the results for the three different number of sensors are represented, the highest MSE curves represents the performance for  $N_s = 1$  and the lowest MSE represent  $N_s = 3$ . The curves are plotted in log log scales for better visualization.

As it was expected, the MSE decreases with the number of sensors and the simulated results are very close to the theoretical approximation for a large number of samples. To have a more appropriate comparison between different numbers of sensors, channel bandwidth constraints must be considered.

In the second simulation, the total rate will be fixed to 5 bits. Two possible settings will be considered, a single sensor approach using the 5 bits ( $N_I = 32$ ) and a multisensor approach

with one sensor quantizing the measurements with 2 ( $N_I = 4$ ) bits and the other with 3 bits ( $N_I = 8$ ). We keep all the other simulation parameters from the previous simulation. The results are shown in Fig. 3.16, also with a comparison with the asymptotic performance (which again is equal to the optimal CRB for quantized measurements).

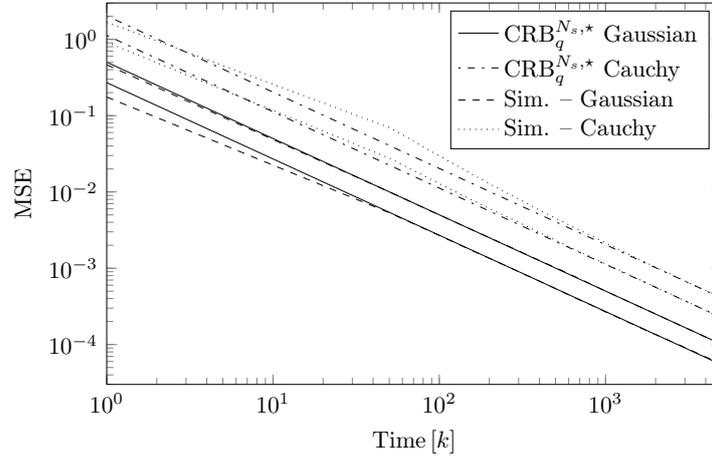


Figure 3.16: Cramér–Rao bound and simulated MSE for the adaptive algorithm for  $N_s = 1$  and  $N_B = 5$  and for  $N_s = 2$ , one sensor with  $N_{B1} = 2$  bits and the other with  $N_{B2} = 3$  bits. The noise was considered to be Gaussian or Cauchy distributed, both with  $\delta = 1$ . For obtaining the simulated MSE, the algorithm was simulated  $5 \times 10^4$  times for blocks with 5000 samples. For all simulations the true parameter was set to zero and the initial estimate was  $\hat{X}_0 = 1$ . In each set of results the higher curve represents the performance for  $N_s = 1$ . The curves are plotted in log log scales for better visualization.

For both types of noise, the theoretical and simulated results show that the multisensor approach is superior.

### Discussion on the results

The proposed algorithm shows that in practice, in a rate constrained context, a multiple sensor approach with low resolution quantizers might be superior to a high resolution single sensor approach. Such observation motivates the use of low resolution sensor networks for estimation purposes.

Note that in the case studied, we did not analyze the interaction between the noise scale factor (it is considered to be constant over the sensors) and the number of quantization bits used in each sensor. When the total number of bits to be transmitted to the fusion center is constrained, an interesting problem for further investigation is the problem of optimal allocation of number of bits to sensors as a function of their noise scale factors. This problem will be studied in an approximate form in Part II.

The adaptive algorithm that is implemented in the fusion center has very low complexity. The complexity is roughly linear in the number of sensors, as the optimal correction  $\eta$  is equivalent to a weighted sum of the corrections given by the single sensor algorithm. Despite

this fact, it can be very costly to implement this algorithm due to the perfect feedback channels requirement. Thus in future work, we can consider that the feedback channels are not perfect, for example, by considering that the estimates are feedback after being quantized and that they are corrupted with additive noise.

### 3.7 Chapter summary and directions

We summarize now the main points observed in this chapter and we also present some subjects that are interesting for further research.

- We presented an adaptive algorithm that can estimate three types of parameter: constant, slowly varying with a Wiener process model without drift or slowly varying with a Wiener process model with drift. The adaptive algorithm can be used for any even number of quantization intervals and under the assumption that the noise is symmetric, unimodal and has a regular CDF (locally Lipschitz continuous) it was shown that
  - using decreasing gains, when the parameter is a constant, the algorithm converges asymptotically in the mean to the true parameter value and its asymptotic performance in terms of variance attains the minimum CRB for common noise distributions  $\text{CRB}_q^*$ . Thus, the answers to the two initial questions (p. 105) are positive: the algorithm with gain proportional to  $\frac{1}{k}$  converges and it can be extended to a multibit setting.
  - Using a constant gain, when the parameter is modeled by a slowly varying Wiener process without drift, the algorithm also converges in the mean to the true parameter and the algorithm is approximately asymptotically optimal. This answers the third question also in a positive way.
  - Using a constant gain, when the parameter is modeled by a slowly varying Wiener process with drift, the algorithm is biased and its asymptotic MSE can be minimized by setting the gain as a function of the drift.
- Using the asymptotic performance results, we evaluated the loss of estimation performance due to quantization for the algorithm. We observed the following:
  - the loss in all cases is a function of  $I_q(0)$ , showing one more time the importance of studying the behavior of this quantity as a function of the threshold variations. We remind that this problem will be studied with an asymptotic approach  $N_I \rightarrow \infty$  in Part II.
  - When the parameter varies, the loss due to quantization is smaller than when the parameter is constant. Thus, when using quantized measurements for estimation, it seems that a type of dithering effect is present.
  - The loss of performance is almost negligible in all cases for 4 or 5 quantization bits. In a rate constrained scenario this seems to be a strong motivation for using a low to medium resolution multiple sensor approach instead of a high resolution single sensor approach. This was validated using an extension of the adaptive algorithm

designed for multiple sensors that can communicate their measurements to a fusion center.

- When comparing the adaptive algorithm with its equivalent counterparts studied in Ch. 1 and 2, the following was observed:
  - for estimating a constant, the adaptive algorithm has a very low complexity when compared with the adaptive scheme based on the MLE and their performance is equivalent.
  - for estimating a slowly varying Wiener process, the algorithm has also a very low complexity when compared with the PF scheme using dynamical central threshold. In this case the only drawback of the adaptive algorithm is that it has a longer transient time.

Therefore, if complexity constraints are present, the adaptive algorithm seems to be the best analyzed solution.

If no constraints on complexity are considered, then the adaptive algorithm is still the best choice for estimating a constant, but it should be replaced by the PF for estimating the slowly varying parameter. An interesting point for future work would be to look for ways of choosing the quantizer update coefficients during the transient, so that the adaptive algorithm performance would be similar to the PF performance.

- We presented two extensions of the algorithm, both for estimating a constant parameter. They are the following:
  - the joint location-scale adaptive estimator, for which we showed that even if we do not know the noise scale parameter it is possible to estimate it with the same asymptotic performance obtained for the case with known scale parameter.
  - The fusion center approach with multiple sensors. In this extension of the algorithm, we considered that measurements from multiple sensors are sent to a fusion center. The role of the fusion center is to estimate the parameter and then broadcast the estimate to the sensors, so that it can be used for setting the quantizers offset. As it was mentioned above, with this approach we showed that a low to medium resolution multiple sensor approach might be better for estimation purposes than a high resolution single sensor approach. We remind that this was shown for sensors with the same type of noise distribution and the same noise scale parameter value. Thus, an interesting subject to study is the bit allocation problem among sensors, when the total bandwidth is constrained and the sensors have the same type of noise distribution but different scale parameters. This will be done in Part II, in the case of a weak bandwidth constraint.
- Many other extensions can be the subject of future work. They are
  - the joint estimation of the drift when we track a Wiener process with drift. Actually, this was already done by adding a simple adaptive estimator of the drift. However, in some cases we can have a more detailed dynamical model for the drift. By

using adaptive multistep algorithms [Benveniste 1990, Sec. 4.2], we can use this additional information to have a better estimate of the Wiener process.

- The joint estimation of the Wiener process increment standard deviation and the Wiener process itself. This will lead to a robust multibit generalization of delta modulation with varying gain.
  - The joint estimation of a location parameter and the shape of the noise distribution. In this case, we can consider that the noise CDF has an unknown shape but a known structure, for example that it is locally polynomial, and that we want to estimate jointly the location parameter and the parameters of the noise distribution.
  - The nonparametric estimation of the location parameter. We can consider for example that we only know that the noise distribution is symmetric, without any specific parametrization. Then we can try to define a nonparametric adaptive algorithm based on adaptive histograms for getting as close as possible to the parametric performance.
  - The joint estimation of location-scales parameters when the multiple sensor fusion center approach is considered. This extension can be directly implemented by joining the features of the adaptive location-scale estimator with the fusion center approach. The main difference in this case is that for reducing the communication complexity, the sensors will have to estimate their individual scale parameters for setting their quantizers.
- We can also consider some extensions of the estimation problem itself for which modifications of the adaptive algorithm would be a good solution. Some examples are:
    - a fusion center approach where the quantized measurements from the sensors are transmitted through noisy channels. This is a problem that we decided not to treat but it is an interesting and more realistic point for further development.
    - A fusion center approach where the information that is feedback is quantized and passed through noisy channels. For this extension, we can consider an additional adaptive algorithm at the sensors for smoothing out the noise from the feedback channels. For dealing with quantization of the estimates we can consider including a dither signal. With this extension we will be able to assess the importance of the feedback channel quality, thus giving a more realistic global estimate of the sensor network cost.

The main issue that makes these two extensions far more difficult to be studied is that the output quantizer indexes cannot be defined arbitrarily, as they are corrupted by the channel noise.

    - The estimation of a scalar parameter following an autoregressive model

$$X_k = aX_{k-1} + W_k$$

instead of the Wiener process model. This will lead to a robust generalization of scalar predictive quantization.

- 
- Compression of a Wiener process with drift. We can consider that at sensor level we can store continuous measurements (or very finely quantized) and then we can apply the adaptive algorithm for a block of measurements in both time directions (forward and backward) and average the results to have final estimates with reduced bias. By storing the initial and final continuous measurements and both the forward and backward quantized sequence, we equivalently have stored a compressed estimate of the true parameter sequence.



# Conclusions of Part I

The main objective of Part I was to propose and study the performance of algorithms for estimation based on quantized measurements. We assumed simple parameter and measurement models:

- Parameter model – a scalar parameter that can be either constant or varying with a Wiener process model.
- Measurement model (noise model) – the scalar constant is measured with independent, unimodal and symmetrically distributed noise.
- Measurement model (quantizer) – the quantizer is symmetric.

Under these settings, we obtained the following conclusions:

- **Adaptiveness is important.** The performance of estimation based on quantized measurements is mainly dependent on the FI for quantized measurements and this dependence is direct. Increased FI is equivalent to increased estimation performance.

For the noise distributions considered, the FI is increased if we set the quantizer dynamic range close to the parameter to be estimated and for commonly used noise distributions (Gaussian, Laplacian, Cauchy), we must put the quantizer central threshold exactly at the true parameter value. As we do not know the value of the parameter, for obtaining optimal estimation performance, we must resort to adaptive algorithms that place the quantizer range close to the true parameter value, for example by placing the quantizer central threshold at the most recent estimate of the parameter. Therefore, this indicates that adaptiveness of the quantizer is a main requirement for optimal estimation.

- **Low complexity is possible and it might be even asymptotically optimal.** It is possible to estimate a constant and a slowly varying parameter with a low complexity adaptive algorithm. The adaptive algorithm is not only convergent in the mean (with a small bias in the drift case) but its parameters can be chosen in such a way that it is exactly equivalent to the asymptotically optimal estimator. This observation goes in the exact opposite direction of some proposed solutions (adaptive scheme based on the MLE and the PF) which requires high complexity both in terms of memory and processing.
- **Low to medium resolution is enough.** Both for a constant and slowly varying parameter model, the loss of performance that is incurred by using quantized measurements instead of continuous amplitude measurements seems to be negligible for a number of quantization bits larger than 4 or 5. In a rate constrained context this means that using more sensors with less resolution may be better than using less sensors with more resolution.



## Part II

Estimation based on quantized  
measurements:

high-rate

approximations



---

*"Finite – to fail, but infinite to venture" - part of a poem of Emily Dickinson.*

## Motivation

The introduction of this part will also be done using a motivational example.

To maintain their economic growth, emerging economies will have to look for new mineral and material sources. This will generate a potential increase of exploration in unusual places, for example the seafloor. Sulfur and metal base rich mineral deposits can be found at seafloor in hydrothermal vent sites [Hoagland 2010].

Hydrothermal vents, also called black smokers, occur when seawater penetrates the oceanic crust through fissures. The water penetrates so deeply in the oceanic crust that it enters in contact with upper parts of magma chambers. A large increase in temperature (from  $\approx 2^\circ\text{C}$  to  $\approx 400^\circ\text{C}$ ) is observed, along with a decrease in pH and Eh. The hot corrosive liquid rises then through fissures carrying metal and sulfur from the rocks. The mineral rich water is released in the seawater as hot black smoke. Precipitation of the elements present in the smoke happens when the hot water is mixed with ocean cold water. As a result, a mineral rich chimney and a massive sulfide deposit are formed around the hot water releasing point [Herzig 2002].

To mine the sulfide deposit, first, it must be located. One possible way to locate it is by measuring the concentration of chemical compounds and elements in the seawater. The chemical plume generated by the hydrothermal vent can be detected using  $\text{CH}_4$ , Fe, H, He or Mn concentration measurements [Baker 2004]. After detecting the plume, for example using sensor measurements from multiple **autonomous underwater vehicles (AUV)** that communicate with a fusion center, the source location must be found. This can be done by following the ascent gradient direction of a chemical compound concentration. The gradient direction can be obtained by exploiting the local information measured by the AUV.

Underwater communication is challenging as bandwidth is severely constrained in this environment. To overcome this problem, quantization of the concentration measurements from the AUV can be considered. As a consequence, to calculate an approximated gradient at the fusion center we will have to deal with the same problem treated in Part I, which is the following:

- **How to estimate a scalar constant (the concentration) based on noisy quantized measurements?**

Algorithms for doing this were presented in Part I, where it was noted that estimation performance is given (at least asymptotically) by the FI for quantized measurements. However, a question remained without answer:

- **How to set the quantizer thresholds to have an optimized estimation performance?**

Only in some cases the answer for this question was given:

1. In the **binary case** it was observed that for commonly used noise models, the quantization threshold should be placed exactly at the parameter.
2. In the **multibit uniform quantization case**, after setting the central threshold at the parameter, the corresponding performance maximization is a one-dimensional optimization problem, which can be solved using exhaustive search.

In the general nonuniform case, setting the threshold was observed to be a complicated optimization problem.

Similarly to standard quantization [Gersho 1992, pp. 185–186], where analytical characterization of quantization performance is difficult for a finite number of quantization intervals, when the number of quantization intervals is large, the set of intervals can be approximated by an interval density. The interval density is a function whose integral over an interval gives the fraction of the number of quantization intervals contained in that interval. By using small interval approximations of the FI, we can obtain an asymptotic expression ( $N_I \rightarrow \infty$ ) for the FI as a function of the interval density. The resulting FI can be maximized w.r.t. the interval density to get an approximation of the optimal interval set. After that, the optimal interval density can be used to have an approximated analytical expression for the optimal FI, thus giving a complete asymptotic characterization of the estimation algorithms.

As the interval density is an asymptotic quantity, a main issue that must also be solved is how to do a practical approximation of this density with a finite number of intervals. An interesting question would be to find an analytical expression for the approximately optimal quantization thresholds as a function of the number of intervals.

Writing it in a more detailed form, we want to do the following:

- **Find an asymptotic (in terms of number of quantization bits) approximation of the Fisher information for estimating a constant parameter embedded in noise as a function of an interval density.**
- (c) • **Find the asymptotically optimal interval density.**
- **Give an analytical expression approximating the maximum FI for the optimal interval density.**
- **Obtain a practical approximation for the optimal quantization thresholds.**

Now, with the optimal thresholds given by the asymptotic approximation, the adaptive estimation algorithms from Part I work (at least asymptotically) in an optimal way. We can imagine that for reliability issues or for reducing measurement latency, multiple concentration sensors are installed in each AUV. Due to deterioration, the sensors do not have exactly the same noise levels. Thus, under a given rate constraint, another question arises:

- **How many quantization bits do we allocate for each sensor?**

For an array of sensors with the same type of noise and considering the independence between sensor noise, the only parameter that can change from sensor to sensor is the noise scale factor. Thus, what we want to do precisely is

**(d) Find the optimal or approximately optimal number of bits per sensor as a function of the noise scale factors under a maximum constraint on the total number of bits.**

The problems presented above are quite general and they can appear in the performance analysis of any optimal estimation algorithm in a constrained rate context. In what follows, we will obtain insight on how we can solve these problems. As we have only one chapter in this part, its outline will be given directly at the chapter introduction.



# High rate approximations of the FI

---

To obtain insight on how we solve problems (c) and (d), we will resort to an asymptotic approach, that is, we will make the number of quantization intervals goes to infinity  $N_I \rightarrow \infty$  and we will see how the FI behaves as a function of the quantizer. This approach can also be found under the names high resolution or high-rate (in the title), the former is used to emphasize that the quantizer intervals are supposed to be very small and the latter is used to make explicit that the communication rate must be high, as the number of quantization bits is large.

Note that making  $N_I \rightarrow \infty$  seems to contradict one of the conclusions of Part I, that states that with only a few quantization bits we have a negligible loss of performance due to quantization. However, even if we make  $N_I \rightarrow \infty$ , we will see that the asymptotic approximations still depend on  $N_I$ , so that, as we stated above, we can use it to gain some insight on the estimation performance for finite  $N_I$ . Actually, we will see that for the location parameter estimation problem studied in Part I, the asymptotic approximations are valid even for small numbers of quantization bits ( $N_B = 4$  and  $N_B = 5$ ), which is very fortunate, as these cases were observed to be the practical useful limit in quantization for estimation and also they are the cases with lowest number of bits for which the maximization of the estimation performance w.r.t. the quantizer thresholds is difficult to be done in a direct way.

When  $N_I \rightarrow \infty$ , the quantizer can be characterized by its density of quantization intervals, thus asymptotically, the behavior of the FI as a function of the quantizer can be characterized by studying its behavior as a function of the intervals density. As a consequence, one of the main objectives of this chapter is to obtain an asymptotic analytic expression of the FI as a function of the interval density.

- *Fixed rate encoding.* We will obtain this expression for scalar quantization and we will not impose any strong constraints on the type of estimation problem that is treated (for example, we will not constrain it to be a location estimation problem).
- *Variable rate encoding.* Additionally to the fixed rate encoding scheme, where all the quantizer outputs use the same number of bits for encoding, we will also obtain the optimal interval density maximizing the FI for the variable rate encoding scheme, where we can use different numbers of bits for different quantizer outputs. We will also discuss on the difficulties of implementing the variable rate encoding scheme in practice.
- *Practical implementation.* We will describe how to implement in practice an approximation of the optimal interval density.

We will check the validity of the results in the location parameter case by comparing the theoretical results for the maximum FI obtained with the optimal interval density with the FI obtained with the practical approximation of the optimal density and with the FI for optimal uniform quantization. We will show that in practice we can obtain the asymptotic performance results by using the adaptive algorithm presented in Ch. 3. We will also look in some detail the location and scale parameter estimation problems for GGD and STD measurements.

In the single sensor location parameter case, we will study the problem of deciding how many quantization bits we might allocate to each sensor in a sensor network, when the total rate is constrained and all the sensors have the same type of noise distribution but different noise scale parameters. Approximate solutions for this problem will be given using the asymptotic approximations.

To show the connections between the results found here and asymptotic results for other inference problems, we will study the asymptotic approximation of a generalized inference performance measure known as the generalized  $f$ -divergence. The asymptotic results for this divergence were proposed in [Poor 1988], mainly for the uniform vector quantization fixed rate encoding case and they were stated but not proved in the non uniform case. Here, we will give a simple derivation of the asymptotic approximation for this divergence in the scalar case using the same procedure as the one that is used for the FI, we will also extend the results to the variable rate encoding case. After obtaining the general optimal density of thresholds, we will point out the similarities and differences between the way quantization must be done for three different inference problems: classic estimation (considered in this thesis), Bayesian estimation and detection.

At the end of the chapter we will summarize the main results and we will indicate some possible points for future work.

#### **Contributions presented in this chapter:**

- *Asymptotic approximation of the optimal interval density for classic parameter estimation.* The asymptotic analysis presented in [Poor 1988] is only detailed for uniform quantization, differently from the development that is presented here, where we consider non uniform quantization with the interval density approach.
- *Practical implementation of the optimal quantizer in the location parameter estimation problem.* In this chapter, we show that the asymptotically optimal quantizer depends on the true parameter value and we also show that in practice we can achieve the asymptotically optimal performance using the adaptive algorithm with decreasing gain presented in Ch. 3. This shows the importance of the adaptive approach. No result of this type seems to be present on the literature.
- *Approximate bit allocation for the multiple sensor approach.* The approximations of the optimal bit allocation among sensors seem to be new in the context of classic location parameter estimation.

---

## Contents

---

<b>4.1</b>	<b>Asymptotic approximation . . . . .</b>	<b>180</b>
4.1.1	General setting . . . . .	180
4.1.2	Loss of estimation performance due to quantization . . . . .	180
4.1.3	Asymptotic approximation of the loss . . . . .	181
4.1.4	Optimal fixed rate encoding . . . . .	183
4.1.5	Variable rate encoding . . . . .	187
4.1.6	Estimation of GGD and STD location and scale parameters . . . . .	190
4.1.7	Location parameter estimation . . . . .	196
<b>4.2</b>	<b>Bit allocation for scalar location parameter estimation . . . . .</b>	<b>200</b>
4.2.1	Unconstrained numbers of bits . . . . .	201
4.2.2	Positive numbers of bits . . . . .	205
<b>4.3</b>	<b>Generalization with the <math>f</math>-divergence . . . . .</b>	<b>207</b>
4.3.1	Definition of the generalized $f$ -divergence . . . . .	207
4.3.2	Generalized $f$ -divergence in inference problems . . . . .	207
4.3.3	Asymptotic results . . . . .	209
4.3.4	Interval densities for inference problems . . . . .	211
<b>4.4</b>	<b>Chapter summary and directions . . . . .</b>	<b>213</b>

---

## 4.1 Asymptotic approximation

### 4.1.1 General setting

The general setting considered here is the estimation of a scalar deterministic parameter  $x \in \mathbb{R}$  of a continuous distribution based on  $N$  independent measurements from this distribution  $\mathbf{Y} = [Y_1 \ Y_2 \ \cdots \ Y_N]^T$ . Again here, we will consider that the estimation of  $x$  is not based on  $\mathbf{Y}$ . Instead, it is based on a scalar quantized version of  $\mathbf{Y}$  denoted

$$\mathbf{i} = [i_1 \ i_2 \ \cdots \ i_N]^T = [Q(Y_1) \ Q(Y_2) \ \cdots \ Q(Y_N)]^T.$$

The function  $Q$  represents the scalar quantizer and is given by

$$Q(Y) = i, \quad \text{if } Y \in q_i = [\tau_{i-1}, \tau_i), \quad (4.1)$$

where  $i \in \{1, \dots, N_I\}$ ,  $N_I$  is the number of quantization intervals  $q_i$  and  $\tau_i$  are the quantizer thresholds. The first and last thresholds will be set to  $\tau_0 = \tau_{\min}$  and  $\tau_{N_I} = \tau_{\max}$ . Note that the setting considered here is more general than the setting presented in Part I, as we do not restrict the estimation problem to be a location estimation problem and we do not impose any symmetry on the quantizer. Observe also that the quantizer interval indexes now go from 1 to  $N_I$ .

It will be assumed that the marginal CDF of the continuous measurements parametrized by  $x$   $F(y; x)$  admits a PDF  $f(y; x)$  that is positive, smooth in both  $x$  and  $y$  and defined on a bounded support. The bounded support assumption is needed to simplify the derivation of the asymptotic results.

### 4.1.2 Loss of estimation performance due to quantization

For estimating a constant with quantized or continuous noisy measurements, we saw in Ch. 1 that the asymptotic performance of an optimal unbiased estimator attains the corresponding CRB. This asymptotic characterization is not restricted to location parameter estimation. Under regularity conditions on the likelihood, it can be applied to any situation where we want to estimate a constant embedded in noisy measurements. Thus, for a general parameter  $x$  and for a large number of samples, the estimation performance is still linked to the FI as follows

$$\text{Var} [\hat{X}] \sim \text{CRB}_q = \frac{1}{NI_q}, \quad (4.2)$$

where  $I_q$  is the FI for a quantized measurement that was already presented in Ch. 1 (for a location parameter). Rewriting the FI with the notation from this part, we have

$$\begin{aligned} I_q &= \mathbb{E} [S_q^2] = \mathbb{E} \left\{ \left[ \frac{\partial \log \mathbb{P}(i; x)}{\partial x} \right]^2 \right\} \\ &= \sum_{i=1}^{N_I} \left[ \frac{\partial \log \mathbb{P}(i; x)}{\partial x} \right]^2 \mathbb{P}(i; x), \end{aligned} \quad (4.3)$$

$S_q$  is again the score function for quantized measurements and  $\mathbb{P}(i; x)$  is the probability of having the quantizer output  $i$  (parametrized by  $x$ ):

$$\mathbb{P}(i; x) = F(\tau_i; x) - F(\tau_{i-1}; x). \quad (4.4)$$

The FI for quantized measurements can be written as a function of the FI for continuous measurements and the score functions, exactly in the same way as it was done in Ch. 1 (1.16 p. 42) (Why? - App. A.1.1)

$$I_q = I_c - \mathbb{E} \left[ (S_c - S_q)^2 \right], \quad (4.5)$$

where  $S_c = \frac{\partial \log f(y; x)}{\partial x}$  is the score function for continuous measurements,  $L = \mathbb{E} \left[ (S_c - S_q)^2 \right]$  is the loss of FI and consequently of estimation performance due to quantization. The main objective from now on will be to minimize  $L$  through the choice of the quantizer intervals when  $N_I$  is large. Notice that minimizing  $L$  defined here is equivalent to minimizing  $L_q$  defined in Ch. 3.

### 4.1.3 Asymptotic approximation of the loss

Similarly to standard quantization for measurement reconstruction, where optimal nonuniform quantization intervals can be approximated for large  $N_I$ , an approximation for  $I_q$  will now be developed.

The loss  $L$  which is an expectation under the measure  $F$  can be rewritten as a sum of integrals, each term of the integral corresponding to the loss produced by a quantization interval:

$$L = \sum_{i=1}^{N_I} \int_{q_i} \left[ \frac{\partial \log f(y; x)}{\partial x} - \frac{\partial \log \mathbb{P}(i; x)}{\partial x} \right]^2 f(y; x) dy. \quad (4.6)$$

**First term**  $\frac{\partial \log f(y; x)}{\partial x}$

For the interval with index  $i$ , the PDF can be approximated with a Taylor series around the central point  $y_i = \frac{\tau_i + \tau_{i-1}}{2}$ :

$$f(y; x) = f_i + f_i^{(y)} (y - y_i) + \frac{f_i^{(yy)}}{2} (y - y_i)^2 + o(y - y_i)^2, \quad (4.7)$$

where the superscripts indicate the variables for which the function is differentiated. The subscript represents that the function (after differentiation) is evaluated at  $y_i$ . It will be assumed that the sequences of intervals for increasing  $N_I$  are chosen such that, for any  $\varepsilon > 0$ , it is possible to find a  $N_I^*$  for which

$$\frac{o(y - y_i)^2}{(y - y_i)^2} < \varepsilon, \quad \text{for } N_I > N_I^*, y \in q_i. \quad (4.8)$$

Under the assumption that  $f > 0$ , the logarithm of  $f$  at interval  $q_i$  can be approximated also using a Taylor series:

$$\log f(y; x) = \log f_i + (\log f)_i^{(y)} (y - y_i) + (\log f)_i^{(yy)} \frac{(y - y_i)^2}{2} + o(y - y_i)^2$$

and the derivative w.r.t.  $x$  is

$$\frac{\partial \log f(y; x)}{\partial x} = (\log f)_i^{(x)} + (\log f)_i^{(yx)} (y - y_i) + (\log f)_i^{(yyx)} \frac{(y - y_i)^2}{2} + o(y - y_i)^2, \quad (4.9)$$

which is an expression for the continuous score function on  $q_i$  to be used in (4.6).

**Second term**  $\frac{\partial \log \mathbb{P}(i; x)}{\partial x}$

Now, the other term in the squared factor must be calculated. Integrating the PDF in (4.7) on the interval  $q_i$ , which has length denoted by  $\Delta_i$ , one gets

$$\mathbb{P}(i, x) = f_i \Delta_i + f_i^{(yy)} \frac{\Delta_i^3}{24} + o(\Delta_i^3). \quad (4.10)$$

Note that the term in  $\Delta_i^2$  is zero as  $y_i$  is the interval central point and the integral of  $(y - y_i)$  around it is zero. The logarithm of  $\mathbb{P}(i, x)$  can be obtained by dividing the second and third terms of the right hand side of (4.10) by the first term and then using the Taylor series for  $\log(1 + x) = x + o(x)$ . Differentiating the resulting expression w.r.t.  $x$  gives

$$\frac{\partial \log \mathbb{P}(i, x)}{\partial x} = (\log f)_i^{(x)} + \left( \frac{f^{(yy)}}{f} \right)_i^{(x)} \frac{\Delta_i^2}{24} + o(\Delta_i^2). \quad (4.11)$$

**Loss  $L$**

Subtracting (4.11) from (4.9) and squaring makes the leading term with least power in  $(y - y_i)$  or in  $\Delta_i$  to be  $(\log f)_i^{(yx)} (y - y_i)$ . When we square this difference and multiply by the Taylor series of  $f$ , we have a leading term  $\left[ (\log f)_i^{(yx)} \right]^2 f_i (y - y_i)^2$  and all other terms have larger powers of  $(y - y_i)$  and/or  $\Delta_i$ . Therefore, after integrating the squared difference multiplied by the Taylor series of  $f$ , we get

$$\begin{aligned} L &= \sum_{i=1}^{N_I} \left\{ \left[ (\log f)_i^{(yx)} \right]^2 f_i \frac{\Delta_i^3}{12} + o(\Delta_i^3) \right\} \\ &= \sum_{i=1}^{N_I} \left\{ \left( S_{c,i}^{(y)} \right)^2 f_i \frac{\Delta_i^3}{12} + o(\Delta_i^3) \right\}, \end{aligned} \quad (4.12)$$

where we have used the fact that  $f$  is smooth enough so that we can change the derivative order between  $y$  and  $x$  to get  $(\log f)_i^{(yx)} = S_{c,i}^{(y)}$ .

To obtain a characterization w.r.t. the quantization intervals, an interval density function  $\lambda(y)$  is defined

$$\lambda(y) = \lambda_i = \frac{1}{N_I \Delta_i}, \quad \text{for } y \in q_i. \quad (4.13)$$

The interval density when integrated in an interval gives, roughly, the fraction of the number of quantization intervals contained in that interval. It is a positive function that always sums

to one<sup>1</sup>. Rewriting (4.12) with this density gives

$$L = \sum_{i=1}^{N_I} \left\{ \left( S_{c,i}^{(y)} \right)^2 f_i \frac{\Delta_i}{12 N_I^2 \lambda_i^2} + o \left( \frac{1}{N_I^2} \right) \Delta_i \right\}. \quad (4.14)$$

As  $N_I \rightarrow \infty$ , it will be supposed that all  $\Delta_i$  converge uniformly to zero. Therefore,

$$\lim_{N_I \rightarrow \infty} N_I^2 L = \frac{1}{12} \int \frac{\left( \frac{\partial S_c(y;x)}{\partial y} \right)^2 f(y;x)}{\lambda^2(y)} dy. \quad (4.15)$$

This asymptotic expression for the loss gives the following approximation for the FI

$$I_q \approx I_c - \frac{1}{12 N_I^2} \int \frac{\left( \frac{\partial S_c(y;x)}{\partial y} \right)^2 f(y;x)}{\lambda^2(y)} dy, \quad (4.16)$$

which is valid for large  $N_I$ . Note that when  $N_I$  in (4.16) tends to infinity, if the quantizer intervals are chosen in a way such that all  $\Delta_i$  tend to zero uniformly, then the asymptotic estimation performance for quantized measurements will tend to the estimation performance for continuous measurements.

#### 4.1.4 Optimal fixed rate encoding

In the fixed rate encoding scheme, all the outputs of the quantizer are encoded with (binary) words that have the same binary size, namely,  $N_B = \log_2(N_I)$ . Thus, we can rewrite (4.16) using the number of bits  $N_B$  instead of the number of intervals  $N_I$ . This gives

$$I_q \approx I_c - \frac{2^{-2N_B}}{12} \int \frac{\left( \frac{\partial S_c(y;x)}{\partial y} \right)^2 f(y;x)}{\lambda^2(y)} dy. \quad (4.17)$$

This shows that the FI for quantized measurements under fixed rate encoding tends exponentially to the FI for continuous measurements with increasing number of bits. Moreover, the constant that multiplies the exponential depends not only on the measurement distribution and on the estimation problem, through  $f$  and  $S_c$ , but also on the quantizer intervals through  $\lambda$ .

#### Optimal interval density

We can characterize asymptotically the optimal quantizer for estimation by defining an optimization problem using (4.16) as the function to be maximized w.r.t.  $\lambda$ . To find the optimal

---

<sup>1</sup>The Riemann sum is equal to one  $\sum_{i=1}^{N_I} \frac{1}{N_I \Delta_i} \Delta_i = 1 \approx \int \lambda(y) dy$ .

$\lambda$  when  $N_B$  is large, we must solve the following optimization problem:

$$\begin{aligned} & \text{minimize} && \int \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^2 f(y;x)}{\lambda^2(y)} dy, \\ & \text{w.r.t. } \lambda(y) && \\ & \text{subject to} && \int \lambda(y) dy = 1, \\ & && \lambda(y) > 0, \end{aligned}$$

where the equality and inequality constraints on  $\lambda$  comes from its definition as a density.

This minimization problem can be solved using Hölder's inequality, which states [Hardy 1988, p. 140] that for two functions  $h(y)$  and  $g(y)$

$$\left(\int |h(y)|^p dy\right)^{\frac{1}{p}} \left(\int |g(y)|^q dy\right)^{\frac{1}{q}} \geq \int |h(y)g(y)| dy,$$

with equality happening when  $h^p(y) \propto g^q(y)$  and  $\frac{1}{p} + \frac{1}{q} = 1$ .

Setting  $p = 3$ ,  $q = \frac{3}{2}$ ,  $h(y) = \left[\frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^2 f(y;x)}{\lambda^2(y)}\right]^{\frac{1}{3}}$  and  $g(y) = \lambda^{\frac{2}{3}}(y)$  in Hölder's inequality and using the constraint that the integral of the density must sum to one, we have the following optimal interval density:

$$\lambda^*(y) = \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x)}{\int \left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x) dy} \propto \left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x) \quad (4.18)$$

and the corresponding maximum FI given by this density is

$$I_q^* \approx I_c - \frac{1}{12N_I^2} \left[ \int \left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x) dy \right]^3. \quad (4.19)$$

**Remark:** in standard quantization for minimum MSE measurement reconstruction the optimal interval density is given by [Gersho 1992, p. 186]

$$\lambda_{\text{rec}}^*(y) = \frac{f^{\frac{1}{3}}(y;x)}{\int f^{\frac{1}{3}}(y;x) dy} \propto f^{\frac{1}{3}}(y;x).$$

Therefore, the main difference from standard quantization is the additional factor depending on the derivative of the score function.

### Practical approximation of the interval density

From the definition of the interval density, the percentage of intervals until interval  $q_i$ ,  $\frac{i}{N_I}$  must be equal to the integral of the interval density from  $\tau_{\min}$  to  $\tau_i$ . Thus, a practical way of approximating the optimal thresholds is to set

$$\tau_i^* = F_\lambda^{-1} \left( \frac{i}{N_I} \right), \quad \text{for } i \in \{1, \dots, N_I - 1\}, \quad (4.20)$$

where  $F_\lambda^{-1}$  is the inverse of the cumulative distribution function (CDF) related to  $\lambda$ .

An important issue for evaluating the  $\tau_i$  is that they may depend explicitly on  $x$ , which is the parameter we want to estimate. A possible solution for this problem is to initially set  $\tau_i$  with an arbitrary guess of  $x$ , then estimate  $x$  using an initial set of measurements and finally update the thresholds with the estimate. This procedure can be performed in an adaptive way to get closer and closer to the optimal thresholds. We can use, for example, an adaptive scheme based on the MLE for doing at the same time estimation and thresholds setting. For the location parameter estimation problem, it was shown that this adaptive scheme converges, thus in this case, if we set  $\tau_i$  according to  $\tau_i^*$ , we expect to obtain the optimal asymptotic performance when  $N \rightarrow \infty$  and  $N_I$  is large. Also in the location parameter case, a low complexity alternative, which gives asymptotically the same performance as the scheme based on the MLE, is the adaptive algorithm presented in Ch. 3. We will see through simulation later that the low complexity adaptive algorithm with the thresholds chosen using  $\tau_i^*$  achieves asymptotically ( $N \rightarrow \infty$ ) a performance close to  $I_q^*$  given by (4.19), even for a moderate number of quantization intervals.

We have the following solution to problem (c) (p. 174):

**Solution to (c) - Asymptotic approximation of the FI  
for fixed rate encoding**

(c1) The asymptotic approximation of the FI is given by (4.16)

$$\begin{aligned} I_q &\approx I_c - \frac{1}{12N_I^2} \int \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^2 f(y;x)}{\lambda^2(y)} \mathbf{d}y \\ &\approx I_c - \frac{2^{-2N_B}}{12} \int \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^2 f(y;x)}{\lambda^2(y)} \mathbf{d}y, \end{aligned}$$

where  $I_c$  and  $S_c$  are the FI and the score function for continuous measurements and  $\lambda(y)$  is the interval density.

- Maximization of  $I_q$  gives the optimal interval density (4.18)

$$\lambda^*(y) = \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x)}{\int \left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x) \mathbf{d}y}.$$

- The corresponding asymptotic approximation of  $I_q$  is (4.19)

$$I_q^* \approx I_c - \frac{1}{12N_I^2} \left[ \int \left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x) \mathbf{d}y \right]^3.$$

- A practical approximation of the asymptotically optimal thresholds using a finite number of quantization intervals is (4.20)

$$\tau_i^* = F_\lambda^{-1} \left( \frac{i}{N_I} \right), \quad \text{for } i \in \{1, \dots, N_I - 1\},$$

where  $F_\lambda^{-1}$  is the inverse of the CDF related to the interval density. This CDF may be dependent on the true parameter  $x$ , therefore, it may be necessary to use an adaptive solution to obtain approximately optimal thresholds.

4.1.5 Variable rate encoding: dead end  $\ominus$ 

It is known from information theory that the minimum average length  $H$  required for describing a discrete r.v. with a binary word is obtained by encoding its possible values (index  $j$ ) with lengths  $l_j$  given by the negative logarithm of their probabilities  $p_j$  [Cover 2006, p. 111]

$$l_j = -\log_2(p_j).$$

For a r.v. with  $n$  possible values, this way of encoding the r.v. gives the following average length

$$H = -\sum_{j=1}^n p_j \log_2(p_j),$$

which is the minimum average length and it is also the entropy of the r.v..

For achieving rate requirements in the problem of estimation based on quantized measurements, instead of using the fixed rate encoding scheme, we can use a scheme with variable rate, where the outputs of the quantizer are coded with binary words with possibly different lengths. The lengths of the outputs can be defined as above, leading to the following minimum average length

$$H_q = -\sum_{i=1}^{N_I} \mathbb{P}(i, x) \log_2[\mathbb{P}(i, x)]. \quad (4.21)$$

Suppose that the communication channel imposes a constraint on the maximum  $H_q$ , so that for lower (or equal to the maximum)  $H_q$ , transmission through this channel occurs without any error, this constraint which is the capacity of the channel will be denoted  $R^2$ . The main objective now is to set the quantizer thresholds for a given  $N_I$  so that the FI  $I_q$  is maximized under the constraint  $H_q \leq R$ . As this problem is complicated to solve for finite  $N_I$ , we will use again the asymptotic approach to obtain the characterization of the optimal quantizer through  $\lambda$ .

The asymptotic expression for  $I_q$  was already developed above and it is given by (4.16)

$$I_q \approx I_c - \frac{1}{12N_I^2} \int \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^2 f(y;x)}{\lambda^2(y)} dy.$$

We need now to develop an asymptotic approximation for the entropy  $H_q$ . Using the Taylor series development for  $\mathbb{P}(i, x)$  given in (4.10) in the expression for  $H_q$  (4.21), we have

$$H_q = -\sum_{i=1}^{N_I} \left[ f_i \Delta_i + f_i^{(yy)} \frac{\Delta_i^3}{24} + o(\Delta_i^3) \right] \log_2 \left[ f_i \Delta_i + f_i^{(yy)} \frac{\Delta_i^3}{24} + o(\Delta_i^3) \right].$$

Separating the factor  $f_i \Delta_i$  inside the logarithm, using the Taylor expansion for  $\log_2(1+x)$  and multiplying the terms in the resulting expression gives

$$H_q = -\sum_{i=1}^{N_I} \left[ f_i \Delta_i \log_2(f_i) + f_i \Delta_i \log_2(\Delta_i) + o(\Delta_i^2) \right].$$

<sup>2</sup>We have supposed in the Introduction that efficient channel coding is used, so that we can assume no-error transmission for rates below channel capacity.

Using the interval density  $\Delta_i = \frac{1}{N_I \lambda_i}$  in the term with  $\log_2(\Delta_i)$  leads to

$$H_q = - \sum_{i=1}^{N_I} [f_i \Delta_i \log_2(f_i) - f_i \Delta_i \log_2(\lambda_i) - f_i \Delta_i \log_2(N_I) + o(\Delta_i^2)].$$

When  $N_I$  is large and  $\Delta_i$  are small, the sums can be approximated by integrals

$$H_q \approx - \int f(y; x) \log_2[f(y; x)] dy + \int f(y; x) \log_2[\lambda(y)] dy + \log_2(N_I),$$

where for obtaining the term  $\log_2(N_I)$ , we used the fact that  $\sum_{i=1}^{N_I} f_i \Delta_i$  is asymptotically close to one as it is approximately the integral of the PDF. The integral  $-\int f(y; x) \log_2[f(y; x)] dy$  is known [Cover 2006, p. 243] as the differential entropy of the r.v.  $Y$ , therefore, from now on we will denote it  $h_y$

$$H_q \approx h_y + \int f(y; x) \log_2[\lambda(y)] dy + \log_2(N_I). \quad (4.22)$$

For large  $N_I$ , using the integral in expression (4.16) and the approximation of the entropy (4.22), we can define the following optimization problem

$$\begin{aligned} & \text{minimize} && \int \frac{\left(\frac{\partial S_c(y; x)}{\partial y}\right)^2 f(y; x)}{\lambda^2(y)} dy, \\ & \text{w.r.t. } \lambda(y) && \\ & \text{subject to} && \int f(y; x) \log_2[\lambda(y)] dy \leq R - h_y - \log_2(N_I), \\ & && \int \lambda(y) dy = 1, \\ & && \lambda(y) > 0. \end{aligned}$$

The solution for this problem can be adapted from the development presented in [Li 1999]. First, we define the function  $p(y)$

$$p(y) = \frac{\left(\frac{\partial S_c(y; x)}{\partial y}\right)^2 f(y; x)}{\int \left(\frac{\partial S_c(y; x)}{\partial y}\right)^2 f(y; x) dy},$$

then the integral that must be minimized can be rewritten as

$$\int \frac{\left(\frac{\partial S_c(y; x)}{\partial y}\right)^2 f(y; x)}{\lambda^2(y)} dy = \left\{ \int \left(\frac{\partial S_c(y; x)}{\partial y}\right)^2 f(y; x) dy \right\} \left[ \int \frac{p(y)}{\lambda^2(y)} dy \right],$$

where we note that only the second factor depends on  $\lambda$ . Thus we can redefine the optimization

problem as

$$\begin{aligned}
& \text{minimize} && \int \frac{p(y)}{\lambda^2(y)} dy, \\
& \text{w.r.t. } \lambda(y) && \\
& \text{subject to} && \int f(y; x) \log_2[\lambda(y)] dy \leq R - h_y - \log_2(N_I), \\
& && \int \lambda(y) dy = 1, \\
& && \lambda(y) > 0.
\end{aligned}$$

To find the optimal  $\lambda$ , we take the logarithm of the integral to be minimized

$$\log_2 \left[ \int \frac{p(y)}{\lambda^2(y)} dy \right] = \log_2 \left[ \int \frac{p(y)}{\lambda^2(y) f(y; x)} f(y; x) dy \right]$$

and we apply Jensen's inequality (the logarithm is a concave function)

$$\log_2 \left[ \int \frac{p(y)}{\lambda^2(y) f(y; x)} f(y; x) dy \right] \geq \int \log_2 \left[ \frac{p(y)}{\lambda^2(y) f(y; x)} \right] f(y; x) dy,$$

now we exponentiate both sides of the inequality

$$\int \frac{p(y)}{\lambda^2(y) f(y; x)} f(y; x) dy \geq 2^{\left\{ \int \log_2 \left[ \frac{p(y)}{\lambda^2(y) f(y; x)} \right] f(y; x) dy \right\}}. \quad (4.23)$$

To obtain equality in the Jensen's inequality the term in the argument of the logarithm in the RHS of (4.23) must be a constant, thus

$$\lambda^*(y) \propto \left[ \frac{p(y)}{f(y; x)} \right]^{\frac{1}{2}} = \left[ \frac{\left( \frac{\partial S_c(y; x)}{\partial y} \right)^2}{\int \left( \frac{\partial S_c(y; x)}{\partial y} \right)^2 f(y; x) dy} \right]^{\frac{1}{2}}.$$

Integrating the constraint that  $\lambda(y)$  is a PDF makes the constant in the denominator of the expression above to disappear, thus giving

$$\lambda^*(y) = \frac{\left| \frac{\partial S_c(y; x)}{\partial y} \right|}{\int \left| \frac{\partial S_c(y; x)}{\partial y} \right| dy}. \quad (4.24)$$

The exponential in (4.23) can be written as a function of the rate constraint. We multiply the rate constraint by  $-2$  and we add  $h_y$  in both sides. We have

$$\int \log_2 \left[ \frac{1}{\lambda^2(y) f(y; x)} \right] f(y; x) dy \geq -2R + 3h_y + 2 \log_2(N_I).$$

Finally, we add  $\int \log_2[p(y)] f(y; x) dy$  to obtain

$$\int \log_2 \left[ \frac{p(y)}{\lambda^2(y) f(y; x)} \right] f(y; x) dy \geq -2R + 3h_y + 2 \log_2(N_I) + \int \log_2[p(y)] f(y; x) dy. \quad (4.25)$$

The integral in the RHS of (4.25) is

$$\begin{aligned}
\int \log_2 [p(y)] f(y; x) dy &= \int \log_2 \left\{ \left[ \frac{\partial S_c(y; x)}{\partial y} \right]^2 \right\} f(y; x) dy + \int \log_2 [f(y; x)] f(y; x) dy \\
&\quad - \int \log_2 \left\{ \int \left( \frac{\partial S_c(y; x)}{\partial y} \right)^2 f(y; x) dy \right\} f(y'; x) dy' \\
&= \int \log_2 \left\{ \left[ \frac{\partial S_c(y; x)}{\partial y} \right]^2 \right\} f(y; x) dy - h_y \\
&\quad - \log_2 \left\{ \int \left( \frac{\partial S_c(y; x)}{\partial y} \right)^2 f(y; x) dy \right\}.
\end{aligned}$$

Substituting the expression above in (4.25) and the result in (4.23), we obtain the minimum value of the integral in the optimization problem. This value is

$$2^{-2} \left\{ R - h_y - \int \log_2 \left[ \left| \frac{\partial S_c(y; x)}{\partial y} \right| \right] f(y; x) dy + \frac{1}{2} \log_2 \left\{ \int \left( \frac{\partial S_c(y; x)}{\partial y} \right)^2 f(y; x) dy \right\} - \log_2(N_I) \right\}.$$

Substituting this value in the approximation of the FI, we get

$$I_q \approx I_c - \frac{1}{12} 2^{-2} \left\{ R - h_y - \int \log_2 \left[ \left| \frac{\partial S_c(y; x)}{\partial y} \right| \right] f(y; x) dy \right\}. \quad (4.26)$$

Notice that again here the FI for quantized measurements tends exponentially to the FI for continuous measurements, the exponential decay rate is sensible to the randomness of the continuous measurements and to the derivative of the score function. The difference in the quantizer characterization w.r.t. to the fixed rate encoding scheme is that now the interval density is dictated only by the derivative of the score function, we must put more intervals around values of  $Y$  that have a larger score function variation.

Observe that the quantizer interval distribution may depend also on the true parameter value, as the score function may be a function of it. Thus, similarly to the fixed rate scheme, it will be necessary to set adaptively the thresholds. The main problem now is that we need to know the quantizer outputs probabilities to encode the outputs with their proper length, however as we do not know completely the measurement distribution, we cannot encode the words properly. As a solution, we can also use an adaptive solution for encoding, using as distribution for encoding, the distribution with the most recent estimate of the parameter. The problem with this solution is that we cannot encode correctly at the beginning of the adaptive estimation procedure, we will be penalized in terms of average length in the initial part of the procedure and as a consequence we will not respect the rate constraints. Thus, this solution is still not complete. Further work will be necessary, we can try to quantify the increase in rate at the beginning of the estimation procedure or we can try to find an encoding scheme with variable rate that quantize the measurements properly, without knowing the true parameter value.

#### 4.1.6 Estimation of GGD and STD location and scale parameters

We will apply the results given in solution (c1) (p. 186) for obtaining the approximately optimal quantization thresholds for estimation of location and scale parameters of the GGD

and the STD. Notice that even if their support is unbounded, as in standard quantization theory, it is expected that the error caused by neglecting the extremal regions (overload region) will be small.

### Results for the estimation of a GGD location parameter

The first step for obtaining the approximately optimal thresholds is to evaluate the optimal interval density given by (4.18). Thus, we start by calculating the derivative of the score function w.r.t.  $x$  and  $y$ . Differentiating the logarithm of the GGD PDF (1.39)

$$f(y; x) = \frac{\beta}{2\delta\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left|\frac{y-x}{\delta}\right|^\beta\right)$$

for  $\beta > 1$ , we obtain

$$\frac{\partial S_c^x(y; x)}{\partial y} = \frac{\beta(\beta-1)}{\delta^2} \left|\frac{y-x}{\delta}\right|^{\beta-2}.$$

Note that for  $\beta \leq 1$ , which includes the Laplacian case, the score function is not differentiable at  $x$ . Thus, we cannot evaluate the interval density for these cases. For  $\beta > 1$ , evaluating the power  $\frac{2}{3}$  of the expression above and multiplying it by  $f^{\frac{1}{3}}(y; x)$ , we have the following interval density:

$$\lambda_{GGD}^x(y) = \frac{\left|\frac{y-x}{\delta}\right|^{\frac{2\beta-4}{3}} \exp\left(-\frac{1}{3}\left|\frac{y-x}{\delta}\right|^\beta\right)}{C}, \quad (4.27)$$

where  $C$  is a constant normalizing the density. Using the symmetry of the density, this constant can be evaluated as the following integral:

$$C = 2 \int_x^{+\infty} \left(\frac{y-x}{\delta}\right)^{\frac{2\beta-4}{3}} \exp\left[-\frac{1}{3}\left(\frac{y-x}{\delta}\right)^\beta\right] dy.$$

An expression for this integral can be obtained by using the change of variables  $\varepsilon = \frac{1}{3}\left(\frac{y-x}{\delta}\right)^\beta$  and identifying the resulting integral factor with the gamma function. This gives

$$C = \frac{2\delta}{\beta} 3^{\frac{1}{3}\left(2-\frac{1}{\beta}\right)} \Gamma\left[\frac{1}{3}\left(2-\frac{1}{\beta}\right)\right].$$

Now, we can obtain the CDF related to the interval density. Exploiting again the symmetry of the distribution, we can obtain the CDF by integrating the PDF only for values of  $y$  larger than  $x$ . Also, by using the same change of variables used above for calculating  $C$ , we get

$$F_{\lambda,GGD}^x(y) = \frac{1}{2} + \frac{\text{sign}(y-x)}{2} \frac{\gamma\left[\frac{1}{3}\left(2-\frac{1}{\beta}\right), \frac{1}{3}\left|\frac{y-x}{\delta}\right|^\beta\right]}{\Gamma\left[\frac{1}{3}\left(2-\frac{1}{\beta}\right)\right]}.$$

Using the inverse of this function we can obtain the approximately optimal thresholds (4.20). For  $i \in \{1, \dots, N_I\}$

$$\tau_{i,GGD}^{*,x} = x + \delta \text{sign}\left(\frac{2i}{N_I} - 1\right) \left\{3\gamma^{-1}\left[\frac{1}{3}\left(2-\frac{1}{\beta}\right), \left|\frac{2i}{N_I} - 1\right| \Gamma\left[\frac{1}{3}\left(2-\frac{1}{\beta}\right)\right]\right]\right\}^{\frac{1}{\beta}}, \quad (4.28)$$

where  $\gamma^{-1}\{\cdot, \cdot\}$  is the inverse incomplete gamma function.

The interval densities for three GGD ( $\beta = 1.5, 2$  and  $2.5$ ) are shown in Fig. 4.1.

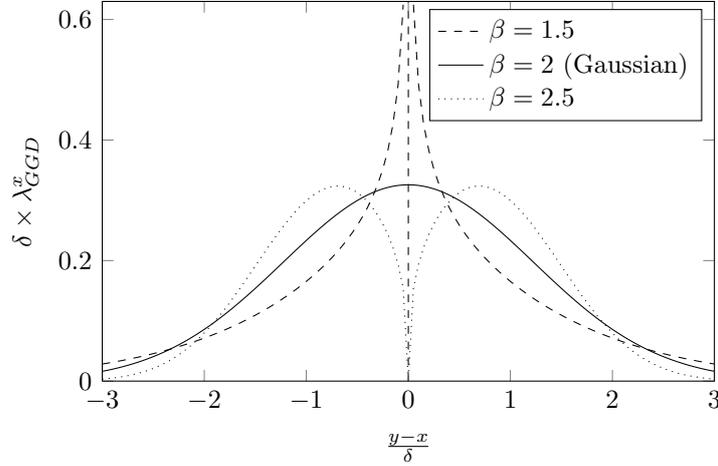


Figure 4.1: Interval densities for the estimation of a GGD location parameter. The GGD shape parameters are  $\beta = 1.5, 2$  and  $2.5$ . Both axis are normalized to have plots independent of  $x$  and  $\delta$ .

A few remarks can be done based on the results above:

- in Ch. 1 we saw that the binary quantization is optimal for the Laplacian distribution, as long as the quantizer threshold is placed at the true parameter. This singular behavior might be related to the difficulties on defining the optimal interval density in this case.
- Observe that for  $1 < \beta < 2$  (see Fig. 4.1 for  $\beta = 1.5$ ), the interval density tends to infinity at zero showing the importance of quantizing around this point for these distributions.
- Notice that within the subclass of GGD for which the density at  $x$  is finite, the Gaussian distribution is the distribution with the lower  $\beta$ . Notice also that for  $\beta > 2$  (see Fig. 4.1 for  $\beta = 2.5$ ), the maximum of the interval density is not placed exactly at zero, showing that a possible relation might exist between the multimodality of the threshold distribution and the asymmetric behavior of optimal binary quantization. It shows also that the Gaussian distribution is exactly between two subclasses of the GGD family, one subclass for which quantizing around the true parameter is very informative ( $1 < \beta < 2$ ) and another subclass for which quantizing symmetrically around the parameter, but not at the parameter, is informative ( $\beta > 2$ ).
- Observing the symmetry of the interval density we can see that, asymptotically, the best quantizer is symmetric around the parameter. Thus if we choose  $N_I$  to be a large even number, the optimal central threshold might be placed at  $x$ . For  $\beta > 2$ , if we have a moderate odd number of quantization intervals, the interval density indicates that the optimal quantizer will be probably asymmetric, as we will have to place more quantization intervals around one of the modes of the interval density.

### Results for the estimation of a GGD scale parameter

We evaluate now the derivative of the logarithm of  $f(y; \delta)$  function w.r.t.  $\delta$  and  $y$ . This gives

$$\frac{\partial S_c^\delta(y; \delta)}{\partial y} = \frac{\beta^2}{\delta^2} \left| \frac{y-x}{\delta} \right|^{\beta-1}.$$

Differently from the location problem, the derivative above exists for all positive  $\beta$ . Using this derivative and the expression for  $f(y; \delta)$  in (4.18), we have

$$\lambda_{GGD}^\delta(y) = \frac{\left| \frac{y-x}{\delta} \right|^{\frac{2\beta-2}{3}} \exp\left(-\frac{1}{3} \left| \frac{y-x}{\delta} \right|^\beta\right)}{C}, \quad (4.29)$$

where the normalizing constant can be obtained using the symmetry of the numerator

$$C = 2 \int_x^{+\infty} \left( \frac{y-x}{\delta} \right)^{\frac{2\beta-2}{3}} \exp\left[-\frac{1}{3} \left( \frac{y-x}{\delta} \right)^\beta\right] dy.$$

Changing the variables  $\varepsilon = \frac{1}{3} \left( \frac{y-x}{\delta} \right)^\beta$  and using the gamma function for rewriting the result, we get

$$C = \frac{2\delta}{\beta} 3^{\frac{1}{3}(2+\frac{1}{\beta})} \Gamma\left[\frac{1}{3}\left(2+\frac{1}{\beta}\right)\right].$$

Using again the symmetry and a similar development as it was done for obtaining the CDF for the interval density of the location problem, we have

$$F_{\lambda,GGD}^\delta(y) = \frac{1}{2} + \frac{\text{sign}(y-x)}{2} \frac{\gamma\left[\frac{1}{3}\left(2+\frac{1}{\beta}\right), \frac{1}{3} \left| \frac{y-x}{\delta} \right|^\beta\right]}{\Gamma\left[\frac{1}{3}\left(2+\frac{1}{\beta}\right)\right]}.$$

Its inverse gives the threshold approximation. For  $i \in \{1, \dots, N_I\}$

$$\tau_{i,GGD}^{*,\delta} = x + \delta \text{sign}\left(\frac{2i}{N_I} - 1\right) \left\{ 3\gamma^{-1}\left\{\frac{1}{3}\left(2+\frac{1}{\beta}\right), \left|\frac{2i}{N_I} - 1\right| \Gamma\left[\frac{1}{3}\left(2+\frac{1}{\beta}\right)\right]\right\}\right\}^{\frac{1}{\beta}}. \quad (4.30)$$

The main differences w.r.t. the location parameter case is that, now, it is the Laplacian distribution that is in the border between the distributions for which  $x$  is a very informative point and the distributions for which most of the information is around  $x$  but not at  $x$ . Note also that the interval density is still dependent on  $\delta$ , thus, as it was said before, for placing optimally the thresholds, an adaptive scheme is necessary.

### Results for the estimation of a STD location parameter

Using the STD PDF (3.72)

$$f(y; x) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\delta\sqrt{\beta\pi}\Gamma\left(\frac{\beta}{2}\right)} \left[1 + \frac{1}{\beta} \left(\frac{y-x}{\delta}\right)^2\right]^{-\frac{\beta+1}{2}},$$

the derivative of the score function is

$$\frac{\partial S_c^x(y; x)}{\partial y} = \frac{\beta + 1}{\beta \delta^2} \frac{\left[1 - \frac{1}{\beta} \left(\frac{y-x}{\delta}\right)^2\right]}{\left[1 + \frac{1}{\beta} \left(\frac{y-x}{\delta}\right)^2\right]^2}.$$

Replacing this expression and the PDF expression above in the interval density (4.18), we obtain

$$\lambda_{STD}^x(y) = \frac{1}{C} \frac{\left[1 - \frac{1}{\beta} \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{2}{3}}}{\left[1 + \frac{1}{\beta} \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{9+\beta}{6}}}. \quad (4.31)$$

The constant  $C$  and the corresponding CDF cannot be expressed analytically, with known special functions. For obtaining a general expression for the thresholds, it might be necessary to use numerical integration of the density for each  $y$  and then invert an interpolation of the numerical integration.

In the special case of a Cauchy distribution ( $\beta = 1$ ), we can evaluate analytically the constant in the density and the CDF. For this distribution the interval density is

$$\lambda_C^x(y) = \frac{1}{C} \frac{\left[1 - \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{2}{3}}}{\left[1 + \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{5}{3}}}. \quad (4.32)$$

From the symmetry, the constant is

$$C = 2 \int_x^{+\infty} \frac{\left[1 - \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{2}{3}}}{\left[1 + \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{5}{3}}} dy.$$

Using the change of variables  $\tan\left(\frac{\theta}{2}\right) = \frac{y-x}{\delta}$ , we obtain

$$C = \delta \int_0^{\pi} \left[ \cos^2\left(\frac{\theta}{2}\right) - \sin^2\left(\frac{\theta}{2}\right) \right]^{\frac{2}{3}} d\theta = 2\delta \int_0^{\frac{\pi}{2}} [\cos^2(\theta)]^{\frac{1}{3}} d\theta,$$

where the second equality was obtained using a relation between trigonometric functions and the periodic pattern of the resulting function in the interval  $[0, \pi)$ . Using another change of variables  $u = \cos^2(\theta)$  and identifying the resulting integral factor with the beta function, we have

$$C = \delta B\left(\frac{1}{2}, \frac{5}{6}\right).$$

Exploiting the symmetry of the interval density and using a similar development, we can obtain the CDF related to the interval density

$$F_{\lambda, C}^x(y) = \frac{1}{2} + \frac{\text{sign}(y-x)}{2} \frac{\int_0^{\phi} [\cos^2(\theta)]^{\frac{1}{3}} d\theta}{B\left(\frac{1}{2}, \frac{5}{6}\right)},$$

with  $\phi = 2 \arctan \left( \left| \frac{y-x}{\delta} \right| \right)$ . Using again the change of variables  $u = \cos^2(\theta)$ , we can rewrite the integral above using the incomplete beta function

$$\int_0^\phi [\cos^2(\theta)]^{\frac{1}{3}} d\theta = \begin{cases} \frac{1}{2} \left[ B\left(\frac{1}{2}, \frac{5}{6}\right) - I_{\cos^2(\phi)}\left(\frac{1}{2}, \frac{5}{6}\right) \right], & \text{for } \phi \in \left[0, \frac{\pi}{2}\right), \\ \frac{1}{2} \left[ 2B\left(\frac{1}{2}, \frac{5}{6}\right) - I_{\cos^2\left(\phi - \frac{\pi}{2}\right)}\left(\frac{1}{2}, \frac{5}{6}\right) \right], & \text{for } \phi \in \left[\frac{\pi}{2}, \pi\right). \end{cases}$$

Transforming  $\phi$  into the initial variable, we have the following CDF

$$F_{\lambda, C}^x(y) = \frac{1}{2} + \frac{\text{sign}(y-x)}{4B\left(\frac{1}{2}, \frac{5}{6}\right)} \left\{ \frac{[1 + \text{sign}(1 - \left| \frac{y-x}{\delta} \right|)]}{2} \left\{ B\left(\frac{1}{2}, \frac{5}{6}\right) - I_{\left[\frac{1 - \left(\frac{y-x}{\delta}\right)^2}{1 + \left(\frac{y-x}{\delta}\right)^2}\right]^2}\left(\frac{1}{2}, \frac{5}{6}\right) \right\} \right. \\ \left. + \frac{[1 + \text{sign}\left(\left| \frac{y-x}{\delta} \right| - 1\right)]}{2} \left\{ B\left(\frac{1}{2}, \frac{5}{6}\right) + I_{\left[\frac{1 - \left(\frac{y-x}{\delta}\right)^2}{1 + \left(\frac{y-x}{\delta}\right)^2}\right]^2}\left(\frac{1}{2}, \frac{5}{6}\right) \right\} \right\}.$$

Inverting the CDF we can obtain the approximate expression for the thresholds. For  $i \in \{1, \dots, N_I\}$  and  $i' = i - \frac{N_I}{2}$

$$\tau_{i, C}^{*, x} = \begin{cases} x + \delta \text{sign}(i') \sqrt{\frac{1 - \sqrt{I^{-1}\left(B\left(\frac{1}{2}, \frac{5}{6}\right)\left(1 - \frac{4|i'|}{N_I}\right)\left(\frac{1}{2}, \frac{5}{6}\right)\right)}}{1 + \sqrt{I^{-1}\left(B\left(\frac{1}{2}, \frac{5}{6}\right)\left(1 - \frac{4|i'|}{N_I}\right)\left(\frac{1}{2}, \frac{5}{6}\right)\right)}}}, & \text{when } |i'| \leq \frac{1}{4}, \\ x + \delta \text{sign}(i') \sqrt{\frac{1 + \sqrt{I^{-1}\left(B\left(\frac{1}{2}, \frac{5}{6}\right)\left(\frac{4|i'|}{N_I} - 1\right)\left(\frac{1}{2}, \frac{5}{6}\right)\right)}}{1 - \sqrt{I^{-1}\left(B\left(\frac{1}{2}, \frac{5}{6}\right)\left(\frac{4|i'|}{N_I} - 1\right)\left(\frac{1}{2}, \frac{5}{6}\right)\right)}}}, & \text{when } |i'| > \frac{1}{4}, \end{cases} \quad (4.33)$$

where  $I_{(\cdot)}^{-1}(\cdot, \cdot)$  is the inverse incomplete beta function.

An interesting point on the optimal interval density for the estimation of a location parameter of the STD is that it equals zero exactly at  $x \pm \sqrt{\beta}\delta$  indicating that around this point not much statistical information can be obtained about the location parameter. If we observe the score function we will see that it is a function with " $\sim$ " shape, the zero derivative point is then related to the maximum and minimum of the score, in a practical sense, the points larger than the maximum and smaller than the minimum can be seen as outliers, so for estimation purposes we might not be interested in quantizing around the transition point. Note however that from a threshold placement point of view, the only practical way of having a zero interval density at this point is by placing a threshold at it, therefore in practice, we are interested in knowing if the measurement is an outlier or not.

### Results for the estimation of a STD scale parameter

For estimating the scale parameter of the STD, we have the following derivative of the score function

$$\frac{\partial S_c^\delta(y; \delta)}{\partial y} = \frac{2}{\delta^2} (\beta + 1) \left\{ \frac{\frac{1}{\beta} \left( \frac{y-x}{\delta} \right)}{\left[ 1 + \frac{1}{\beta} \left( \frac{y-x}{\delta} \right)^2 \right]^2} \right\}.$$

This leads to the following interval density

$$\lambda_{STD}^\delta(y) = \frac{1}{C} \left\{ \frac{\frac{1}{\beta} \left( \frac{y-x}{\delta} \right)}{\left[ 1 + \frac{1}{\beta} \left( \frac{y-x}{\delta} \right)^2 \right]^2} \right\}, \quad (4.34)$$

with  $C$  given by

$$C = 2 \int_x^{+\infty} \left\{ \frac{\frac{1}{\beta} \left( \frac{y-x}{\delta} \right)}{\left[ 1 + \frac{1}{\beta} \left( \frac{y-x}{\delta} \right)^2 \right]^2} dy \right\}.$$

Using a change of variables  $\varepsilon = \frac{1}{\left[ 1 + \frac{1}{\beta} \left( \frac{y-x}{\delta} \right)^2 \right]}$  and identifying the resulting integral with the beta function, we obtain

$$C = \sqrt{\beta} \delta B \left( \frac{5}{6}, \frac{\beta+4}{6} \right).$$

Exploiting the symmetry of the interval density and using the previous change of variables, we have the following CDF related to the interval density

$$F_{\lambda, STD}^\delta(y) = \frac{1}{2} + \frac{\text{sign}(y-x)}{2} \frac{\left[ B \left( \frac{5}{6}, \frac{\beta+4}{6} \right) - I_{\frac{1}{1 + \frac{1}{\beta} \left( \frac{y-x}{\delta} \right)^2}} \left( \frac{5}{6}, \frac{\beta+4}{6} \right) \right]}{B \left( \frac{5}{6}, \frac{\beta+4}{6} \right)}.$$

For  $i \in \{1, \dots, N_I\}$ , the approximately optimal thresholds are then given by

$$\tau_{i, STD}^{\star, \delta} = x + \delta \text{sign} \left( \frac{2i}{N_I} - 1 \right) \left\{ \beta \left\{ I_{B \left( \frac{5}{6}, \frac{\beta+4}{6} \right) \left( 1 - \left| \frac{2i}{N_I} - 1 \right| \right)} \left( \frac{5}{6}, \frac{\beta+4}{6} \right) \right\}^{-1} - \beta \right\}^{\frac{1}{2}}. \quad (4.35)$$

Note that similarly to the GGD scale parameter estimation case, the point  $x$  is not very informative. Most of the quantizer intervals must be placed around  $x$  but not very close to  $x$ .

#### 4.1.7 Location parameter estimation

To check the results, we will now focus on location parameter estimation.

First observe that using the normalized form for the PDF  $f(y; x) = \frac{1}{\delta} f_n \left( \frac{y-x}{\delta} \right)$ , we can rewrite the interval density given by (4.18)

$$\lambda^\star(y) \propto \left[ \frac{\partial^2 \log \left[ \frac{1}{\delta} f_n \left( \frac{y-x}{\delta} \right) \right]}{\partial y \partial x} \right]^{\frac{2}{3}} \left[ \frac{1}{\delta} f_n \left( \frac{y-x}{\delta} \right) \right]^{\frac{1}{3}} \propto \frac{\left[ f_n^{(1)2} \left( \frac{y-x}{\delta} \right) - f_n \left( \frac{y-x}{\delta} \right) f_n^{(2)} \left( \frac{y-x}{\delta} \right) \right]^{\frac{2}{3}}}{f_n \left( \frac{y-x}{\delta} \right)},$$

where  $f_n^{(1)}$  and  $f_n^{(2)}$  are the first and second derivatives of  $f_n$  w.r.t. its argument. For a  $f_n$  with even symmetry,  $f_n^{(1)2}$  is even,  $f_n^{(2)}$  is even and consequently  $\lambda^*(y)$  is symmetric around  $x$ . This means that for large  $N_I$ , the optimal quantizer is symmetric around  $x$ , indicating that, asymptotically, the asymmetry of the optimal quantizer for binary quantization under some distributions (Subsec. 1.3.4, p. 48) might disappear.

The asymptotic approximation of  $I_q$  given by (4.19) can also be rewritten using the normalized PDF

$$I_q^* \approx \frac{1}{\delta^2} \left\{ I_{c,n}^x - \frac{2^{-2N_B}}{12} \left[ \int \frac{[f_n^{(1)2}(\varepsilon) - f_n(\varepsilon)f_n^{(2)}(\varepsilon)]^{\frac{2}{3}}}{f_n(\varepsilon)} d\varepsilon \right]^3 \right\}, \quad (4.36)$$

where  $I_{c,n}^x$  is the FI for estimating a location parameter when  $\delta = 1$ . Note that the FI approximation can be written as  $\frac{\kappa(f_n)}{\delta^2}$ , where  $\kappa$  is a functional depending only on the normalized PDF and independent of  $x$  and  $\delta$ . Therefore, we can have a characterization of the optimal estimation performance based on quantized measurements for a family of distributions with different  $\delta$  and  $x$  only by evaluating  $\kappa(f_n)$ .

#### FI for the Gaussian and Cauchy cases

We will check the results using the Gaussian (GGD with  $\beta = 2$ ) and Cauchy (STD with  $\beta = 1$ ) distributions.

For the Gaussian distribution, the interval density (4.27) and the asymptotic approximation of the FI (4.18) are given by

$$\lambda_G^x(y) = \frac{1}{\delta\sqrt{3\pi}} \exp \left[ - \left( \frac{y-x}{\sqrt{3}\delta} \right)^2 \right], \quad I_{q,G}^x \approx \frac{2}{\delta^2} \left[ 1 - \pi\sqrt{3} 2^{-(2N_B-1)} \right]. \quad (4.37)$$

We can note that the interval density in this case is exactly the same as for standard quantization (proportional to  $f^{\frac{1}{3}}$ ). Thus, in the Gaussian case when  $N_I$  is large, the optimal quantizer for estimating the location parameter and for recovering the continuous measurement is the same. This coincidence between the optimal quantizer for estimation and for reconstruction happens whenever the score function is constant. In the location parameter estimation case, this happens only for the Gaussian distribution. If we look to the scale parameter case, this will happen for the Laplacian distribution.

Observe also that if it was possible to implement the variable rate encoder in the Gaussian case, then the optimal quantizer would be a uniform quantizer and it would coincide with the optimal variable rate quantizer for reconstruction which is uniform [Gersho 1992, p. 299].

For the Cauchy distribution, the interval density (4.32) and the asymptotic FI approximation are the following:

$$\lambda_C^x(y) = \frac{1}{\delta B\left(\frac{1}{2}, \frac{5}{6}\right)} \frac{\left[1 - \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{2}{3}}}{\left[1 + \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{5}{3}}}, \quad I_{q,C}^x \approx \frac{1}{2\delta^2} \left[ 1 - \frac{B\left(\frac{1}{2}, \frac{5}{6}\right)^3}{3\pi} 2^{-2N_B+1} \right]. \quad (4.38)$$

To evaluate the validity of the results, the FI (4.3) under both distributions for  $\delta = 1$  was evaluated for

- the optimal set of thresholds for  $N_B \in \{1, 2, 3\}$ . The optimal thresholds were obtained through exhaustive search. For  $N_B \in \{4, 5, 6, 7, 8\}$  the theoretical results (4.37) and (4.38) were used as an approximation.
- uniform quantization considering  $N_B \in \{1, \dots, 8\}$ . After setting the central threshold to  $x$ , the optimal quantization interval step-length  $\Delta^*$  was found by maximizing the FI also using exhaustive search.
- the approximate optimal set of thresholds given by (4.28) and by (4.33), for  $N_B \in \{1, \dots, 8\}$ .

The results are given in Tab. 4.1.

$N_B$	Gaussian ( $I_{c,n}^x = 2$ )			Cauchy ( $I_{c,n}^x = 0.5$ )		
	Optimal	Uniform	Practical approx.	Optimal	Uniform	Practical approx.
1	1.27323954 <sup>†</sup>	–	1.27323954	0.40528473 <sup>†</sup>	–	0.40528473
2	1.76503630 <sup>†</sup>	1.76503630	1.75128300	0.43433896 <sup>†</sup>	0.43433896	0.40528473
3	1.93090199 <sup>†</sup>	1.92837814	1.92740111	0.48474865 <sup>†</sup>	0.45600797	0.47893785
4	1.97874454 <sup>*</sup>	1.97841622	1.98038526	0.49533850 <sup>*</sup>	0.48136612	0.49504170
5	1.99468613 <sup>*</sup>	1.99353005	1.99489906	0.49883463 <sup>*</sup>	0.49204506	0.49879785
6	1.99867153 <sup>*</sup>	1.99807736	1.99869886	0.49970866 <sup>*</sup>	0.49656712	0.49970408
7	1.99966788 <sup>*</sup>	1.99943563	1.99967136	0.49992716 <sup>*</sup>	0.49851056	0.49992659
8	1.99991697 <sup>*</sup>	1.99983649	1.99991741	0.49998179 <sup>*</sup>	0.49935225	0.49998172

Table 4.1: FI for the estimation of Gaussian and Cauchy location parameters based on quantized measurements.  $N_B$  is the number of quantization bits. In **Optimal**<sup>†</sup> the maximum FI obtained by exhaustive search of the thresholds is shown. **Optimal**<sup>\*</sup> is the theoretical asymptotic approximation of the FI. **Uniform** shows the value of the FI for optimal uniform quantization and **Practical approx.** gives the FI for the practical approximation of the asymptotically optimal thresholds.

In all cases the fast convergence to the continuous FI with increasing  $N_B$  is verified. Again here, 4 or 5 bits are enough for obtaining an estimation performance close to the continuous measurement performance. The difference of performance between uniform and nonuniform quantization seems to be higher for the Cauchy distribution. In the Gaussian case, this difference is negligible, indicating that in practice uniform quantization should be used (as it is easier to implement). It can also be observed that the asymptotic approximation of the FI and its true value for the practical approximation of the optimal threshold set are very close, even for small values of  $N_B$  ( $N_B = 4$ ).

#### Verification with the adaptive algorithm

As it was pointed out before, an important issue for evaluating the practical approximation of the optimal thresholds  $\tau_i^*$  is that they depend explicitly on  $x$ . Thus a possible solution to,

at the same time, obtain an estimate of the parameter and set the quantizer thresholds is to use the adaptive algorithm proposed in Ch. 3

$$\hat{X}_k = \hat{X}_{k-1} + \frac{1}{kI_q} \eta(i_k),$$

with the threshold variations set  $\tau'$  given by the practical approximation  $\tau^*$  with  $x$  in (4.20) set to zero and  $\eta(i_k)$  given by  $\eta(i) = -\frac{f(\tau_{i-1}^*;x) - f(\tau_i^*;x)}{F(\tau_i^*;x) - F(\tau_{i-1}^*;x)}$ . If  $N_B \geq 4$ , for a large  $k$ , the asymptotic variance of the algorithm will be close to optimal and it will be given approximately by

$$\text{Var} [\hat{X}_k] \approx \text{CRB}_q = \frac{1}{kI_q}, \quad (4.39)$$

where  $I_q$  is the asymptotic approximation given by (4.19).

This algorithm was tested under both distributions for  $N_B = 4$  and 5. The MSE for the algorithm was evaluated using Monte Carlo simulation,  $4 \times 10^6$  realizations of blocks with  $5 \times 10^4$  samples were used. The initial error  $x - \hat{X}_0$  and  $\delta$  were both set to be 1 in all simulations. The MSE for the algorithm and the approximation given by (4.39) are both given in Fig. 4.2, where they are multiplied by  $k$  for better visualization.

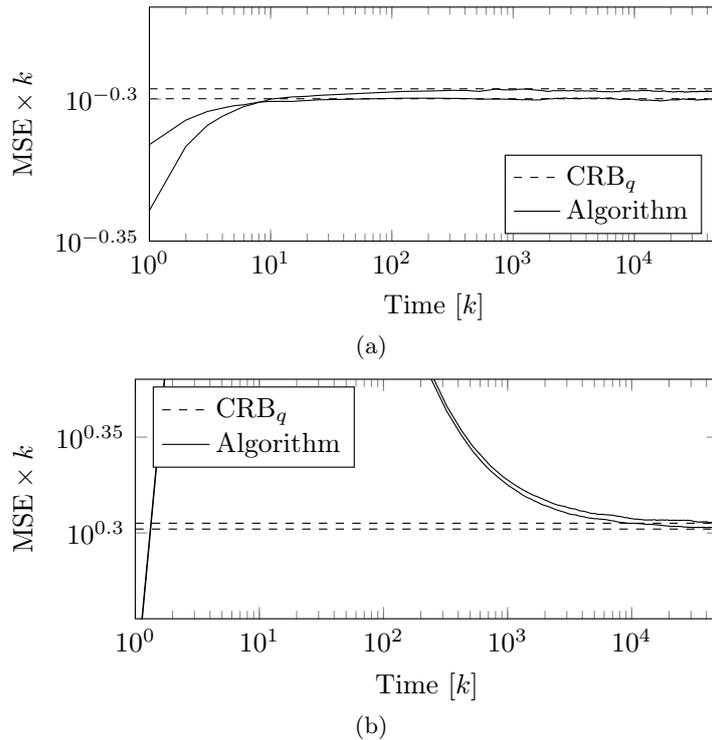


Figure 4.2: Simulated MSE for the adaptive algorithm considering Gaussian (a) and Cauchy (b) measurement distributions. The numbers of quantization bits are  $N_B = 4$  and 5. The initial estimation error and  $\delta$  were set to 1 in all the cases. The simulated MSE was obtained through Monte-Carlo simulation,  $4 \times 10^6$  realizations of blocks with  $5 \times 10^4$  error samples were used. The curves that have asymptotically higher values correspond to  $N_B = 4$ .

We observe that the asymptotic algorithm performance is very close to the approximation. For small  $k$  the CRB is not tight and that seems to be the reason for the algorithm to perform better than the bound. In other simulations, it was also observed that using uniform thresholds leads to faster convergence to the asymptotic performance. This indicates that in practice an algorithm with changing thresholds can be used for obtaining better results. In the convergence phase, a uniform set of thresholds is used, then after a given number of samples, the thresholds change to the approximately optimal set.

## 4.2 Bit allocation for scalar location parameter estimation

The objective now is to solve problem (d) (p. 175). We have  $N_s$  sensors measuring independently the same location parameter  $x$  and the continuous measurements from one sensor to another have all the same noise type with normalized PDF  $f_n$ . The only difference between the noise distribution from one sensor to another is the scale factor. For the  $N_s$  sensors, the scale factors are denoted  $\{\delta_1, \dots, \delta_{N_s}\}$ . Each sensor  $i$  quantizes its measurements with a number of bits  $N_{B,i}$  such that the total number of bits among the sensors is constrained to be  $N_B$ . The objective then is to find the allocation of bits  $\{N_{B,1}, \dots, N_{B,N_s}\}$  that maximizes the estimation performance.

The estimation performance for unbiased estimators in terms of variance can be characterized asymptotically by the CRB, which is related to the inverse of the FI. Thus, by maximizing the FI, the asymptotic estimation performance is maximized. As the sensors measurements are independent, the FI for the measurements from all the sensors  $I_q$  is the sum of the FI  $I_{q,i}(N_{B,i})$  for each sensor

$$I_q = \sum_{i=1}^{N_s} I_{q,i}(N_{B,i}), \quad (4.40)$$

where we made explicit the dependence of the FI for each sensor on the allocated number of bits.

We will assume that the thresholds can be chosen so that  $I_{q,i}(N_{B,i})$  is maximum. This can be done for example by using the adaptive algorithm with decreasing gain to set optimally the central threshold and then by choosing optimally the threshold variations. Thus, we want to solve the following optimization problem:

$$\begin{aligned} & \text{maximize} && I_q = \sum_{i=1}^{N_s} I_{q,i}(N_{B,i}), \\ & \text{w.r.t. } N_{B,i} && \\ & \text{subject to} && \sum_{i=1}^{N_s} N_{B,i} = N_B, \\ & && N_{B,i} \in \mathbb{N}, \end{aligned}$$

where  $I_{q,i}(N_{B,i})$  is the maximum FI for  $N_{B,i}$ .

This problem can be solved exactly by evaluation of  $I_q$  for all possible combinations of the  $N_{B,i}$ . The numbers of allocated bits  $N_{B,i}$  can assume values from 0 to  $N_B$  but their sum

must be  $N_B$ . Therefore, the  $N_{B,i}$  form a weak composition of  $N_B$  into  $N_s$  parts. The number of possible allocations is  $\binom{N_B + N_s - 1}{N_s - 1} = \frac{(N_B + N_s - 1)!}{(N_s - 1)! N_B!}$ . If we have to solve the allocation problem for  $N_s = 20$  and  $N_B = 100$ , then the number of possible allocations that we have to compare is approximately  $4.9 \times 10^{21}$ , which indicates that in practice the exact solution for this problem is difficult to be obtained by exhaustive search.

### 4.2.1 Unconstrained numbers of bits

If we neglect the constraint that  $N_{B,i}$  must be a non-negative integer and we suppose that the asymptotic approximation of  $I_q$  (4.36) is valid for all real  $N_{B,i}$ , then we can define a maximization problem that can be solved analytically. Using the approximation (4.36), we have that the total FI can be approximated by

$$I_q \approx \sum_{i=1}^{N_s} \frac{1}{\delta_i^2} \left\{ I_{c,n}^x - \frac{2^{-2N_{B,i}}}{12} \left[ \int \frac{[f_n^{(1)2}(\varepsilon) - f_n(\varepsilon) f_n^{(2)}(\varepsilon)]^{\frac{2}{3}}}{f_n(\varepsilon)} d\varepsilon \right]^3 \right\}. \quad (4.41)$$

Maximizing the approximation in the RHS of (4.41) is equivalent to minimizing  $\sum_{i=1}^{N_s} \frac{2^{-2N_{B,i}}}{\delta_i^2}$  as  $I_{c,n}^x$  and the integral are constants if all the sensor noise types are equal. Thus the relaxed form (without the integer constraints) of the bit allocation problem is the following:

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^{N_s} \frac{2^{-2N_{B,i}}}{\delta_i^2}, \\ & \text{w.r.t. } N_{B,i} && \\ & \text{subject to} && \sum_{i=1}^{N_s} N_{B,i} = N_B. \end{aligned}$$

We can solve this optimization problem by integrating the constraint in the function to be minimized using a Lagrange multiplier. The Lagrangian (the function to be minimized) for this minimization problem is

$$\mathcal{L} = \left[ \sum_{i=1}^{N_s} \frac{2^{-2N_{B,i}}}{\delta_i^2} \right] + \lambda \left[ \left( \sum_{i=1}^{N_s} N_{B,i} \right) - N_B \right],$$

where  $\lambda$  is the Lagrange multiplier. As the function is convex (it is a sum of negative exponentials plus a sum of linear terms), the zero gradient point of the Lagrangian w.r.t. the  $N_{B,i}$  gives a global minimum. The derivative of the Lagrangian w.r.t.  $N_{B,i}$  is

$$\frac{\partial \mathcal{L}}{\partial N_{B,i}} = \frac{-2 \ln(2) 2^{-2N_{B,i}}}{\delta_i^2} + \lambda,$$

which is zero for

$$N_{B,i} = \frac{\log_2 \left[ \frac{\lambda \delta_i^2}{2 \ln(2)} \right]}{-2}. \quad (4.42)$$

To find  $\lambda$  it is necessary to use (4.42) in the sum constraint

$$\sum_{i=1}^{N_s} N_{B,i} = \frac{N_s \log_2(\lambda) - N_s \log_2[2 \ln(2)] + 2 \left[ \sum_{i=1}^{N_s} \log_2(\delta_i) \right]}{-2} = N_B,$$

thus,

$$\log_2(\lambda) = -2 \frac{N_B}{N_s} + \log_2[2 \ln(2)] - \frac{2}{N_s} \left[ \sum_{i=1}^{N_s} \log_2(\delta_i) \right]. \quad (4.43)$$

Using (4.43) in (4.42) gives

$$\begin{aligned} N_{B,i} &= \frac{N_B}{N_s} - \frac{\log_2[2 \ln(2)]}{2} + \frac{1}{N_s} \left[ \sum_{i=1}^{N_s} \log_2(\delta_i) \right] - \log_2(\delta_i) + \frac{\log_2[2 \ln(2)]}{2} \\ &= \frac{N_B}{N_s} + \frac{1}{N_s} \left[ \sum_{i=1}^{N_s} \log_2(\delta_i) \right] - \log_2(\delta_i), \end{aligned}$$

which can be rewritten as

$$N_{B,i} = \frac{N_B}{N_s} - \log_2 \left( \frac{\delta_i}{\sqrt[N_s]{\prod_{j=1}^{N_s} \delta_j}} \right). \quad (4.44)$$

This is a correction on the uniform bit allocation that depends on the weight of the distribution scale parameter in the geometric mean of the scale factors.

Note that the approximate allocation depends only on  $\delta_i$  and no other information about the distribution is required. In practice we can estimate  $\delta_i$  for each sensor with an arbitrary allocation and then we can use the estimates in (4.44) and round the results in a proper way for obtaining for obtaining integer  $N_{B,i}$ .

If we use the approximate solution from (4.44), we obtain

$$\begin{aligned} I_q &\approx \left[ I_{c,n}^x \left( \sum_{i=1}^{N_s} \frac{1}{\delta_i^2} \right) \right] - \left\{ \frac{\kappa'(f_n)}{12} \left[ \sum_{i=1}^{N_s} \frac{2^{-2N_{B,i}}}{\delta_i^2} \right] \right\} = \\ &= \left[ I_{c,n}^x \left( \sum_{i=1}^{N_s} \frac{1}{\delta_i^2} \right) \right] - \left\{ \frac{\kappa'(f_n)}{12} \left\{ \sum_{i=1}^{N_s} \frac{2^{-2 \left[ \frac{N_B}{N_s} - \log_2 \left( \frac{\delta_i}{\sqrt[N_s]{\prod_{j=1}^{N_s} \delta_j}} \right) \right]}}{\delta_i^2} \right\} \right\} = \\ &= N_s \left[ \frac{I_{c,n}^x}{HM(\delta_1^2, \dots, \delta_{N_s}^2)} - \frac{\kappa'(f_n)}{12} \frac{2^{-2\bar{N}_B}}{GM(\delta_1^2, \dots, \delta_{N_s}^2)} \right], \quad (4.45) \end{aligned}$$

where  $HM(\delta_1^2, \dots, \delta_{N_s}^2) = \frac{N_s}{\sum_{i=1}^{N_s} \frac{1}{\delta_i^2}}$  and  $GM(\delta_1^2, \dots, \delta_{N_s}^2) = \sqrt[N_s]{\prod_{j=1}^{N_s} \delta_j^2}$  are the harmonic and geometric means of the squared scale factors,  $\kappa'(f_n)$  is the integral factor in (4.36) and  $\bar{N}_B = \frac{N_B}{N_s}$  is the number of allocated bits per sensor that would be obtained if we had used a uniform bit allocation.

If we compare this result to uniform bit allocation

$$I_q \approx \frac{N_s}{HM(\delta_1^2, \dots, \delta_{N_s}^2)} \left[ I_{c,n}^x - \frac{\kappa'(f_n)}{12} 2^{-2\bar{N}_B} \right],$$

we can verify that, as the geometric mean is larger than the harmonic mean, the approximate optimal bit allocation performs better than or equal to the uniform bit allocation.

If it was possible to implement this allocation scheme, an interesting point for future study would be to study the influence of the variability of the precision of the sensors  $\frac{1}{\delta_i^2}$  on the estimation performance. This might be done for example by considering that the  $\frac{1}{\delta_i^2}$  are i.i.d. r.v. with a given distribution (a gamma distribution for example) with known parameters, then by assuming large  $N_s$  for a fixed  $\bar{N}_B$ , we can apply the law of large numbers to the harmonic and the geometric means in the approximation of  $I_q$  (4.45) to obtain a characterization of the approximately optimal FI as a function of the parameters of the precision distribution. This approach, even if approximate, might give some insight on the performance of estimation of asymptotically large heterogeneous sensor arrays under communication rate constraints.

We have the following solution to problem (d) (p. 175):

**Solution to (d) - Unconstrained approximate optimal bit allocation for location parameter estimation**

(d1) For  $i \in \{1, \dots, N_s\}$ , the approximate optimal bit allocation is given by (4.44)

$$N_{B,i} = \frac{N_B}{N_s} - \log_2 \left( \frac{\delta_i}{\sqrt[N_s]{\prod_{j=1}^{N_s} \delta_j}} \right).$$

Appropriate rounding can be used to obtain  $N_{B,i} \in \mathbb{N}$ .

- For the approximate optimal bit allocation, the FI is given by (4.45)

$$I_q \approx N_s \left[ \frac{I_{c,n}^x}{HM(\delta_1^2, \dots, \delta_{N_s}^2)} - \frac{\kappa'(f_n)}{12} \frac{2^{-2\bar{N}_B}}{GM(\delta_1^2, \dots, \delta_{N_s}^2)} \right],$$

where  $I_{c,n}^x$  is the continuous FI for  $\delta = 1$ ,  $\kappa'(f_n) = \frac{1}{12} \left[ \int \frac{[f_n^{(1)2}(\varepsilon) - f_n(\varepsilon)f_n^{(2)}(\varepsilon)]^{\frac{2}{3}}}{f_n(\varepsilon)} d\varepsilon \right]^3$ ,  $\bar{N}_B = \frac{N_B}{N_s}$  is the average number of bits per sensor and  $HM$  and  $GM$  are the harmonic and geometric means of the scale parameters.

### 4.2.2 Positive numbers of bits

For obtaining a more realistic solution, we can constrain the numbers of bits to be nonnegative reals. This gives the following optimization problem:

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^{N_s} \frac{2^{-2N_{B,i}}}{\delta_i^2}, \\ & \text{w.r.t. } N_{B,i} && \\ & \text{subject to} && \sum_{i=1}^{N_s} N_{B,i} = N_B, \\ & && N_{B,i} \geq 0. \end{aligned}$$

The Lagrangian is the same as for the unconstrained problem. Using the zero gradient condition, we have

$$N_{B,i} = \frac{\log_2 \left[ \frac{\lambda \delta_i^2}{2 \ln(2)} \right]}{-2} = \nu - \log_2(\delta_i),$$

where  $\nu$  is a constant to be chosen. Note that the positivity constraint imposes the following form for  $N_{B,i}$

$$N_{B,i} = [\nu - \log_2(\delta_i)]_+, \quad (4.46)$$

with  $[x]_+ = \max(x, 0)$ . The sum constraint gives

$$\sum_{i=1}^{N_s} N_{B,i} = \sum_{i=1}^{N_s} [\nu - \log_2(\delta_i)]_+ = N_B. \quad (4.47)$$

Thus, the constant  $\nu$  is chosen so that (4.47) is satisfied and then the number of bits can be chosen according to (4.46). Again here, appropriate rounding might be used to obtain integer numbers of bits.

Observe that this approximate bit allocation is equivalent to water-filling, a common solution to allocate power to carriers in multicarrier modulation. The main difference is that in this case the channel noise is replaced by  $\log_2(\delta_i)$  and the "water depths" are the number of bits instead of the power levels.

In Fig. 4.3, both water-filling solutions are shown, for power allocation in multicarrier systems and for approximate bit allocation in constrained rate sensing systems.

When the  $\delta_i$  are also unknown, we can mix the two extensions of the adaptive algorithm with decreasing gain presented in Ch. 3 (fusion center + joint estimation of the scale) to have estimates of the scale parameters. Then, we can use the estimates to obtain the approximate allocation. In practice, the value of  $\nu$  can be evaluated at the fusion center and broadcasted to the sensors with the location parameter. The sensors can use the broadcasted  $\nu$  with a local estimate of the location parameter for obtaining the optimal  $N_{B,i}$ . The critical point with this approach will be the final rounding step, which will require an agreement (and consequently communication) between the sensors to respect the total bandwidth constraint.

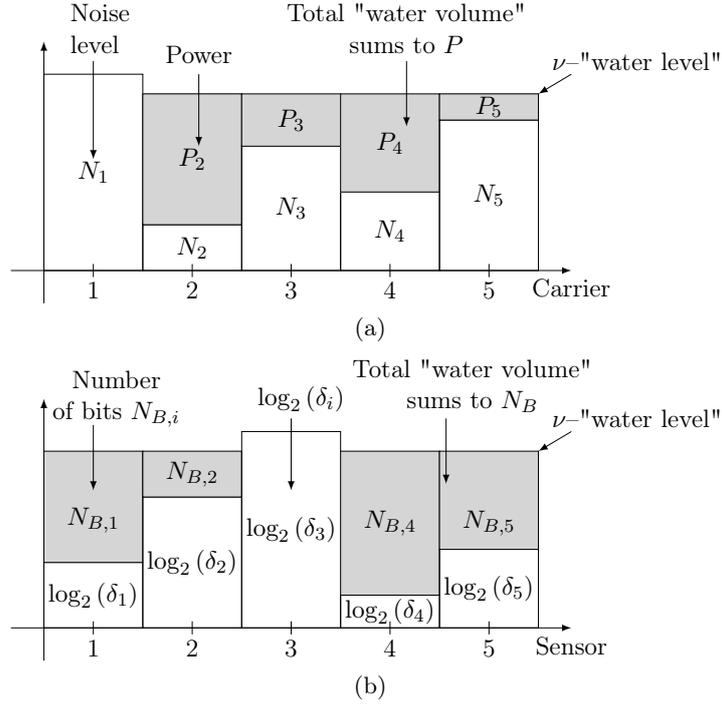


Figure 4.3: Both water-filling solutions for multicarrier modulation power allocation (a) and for rate constrained sensing system bit allocation (b).

This gives the following solution to problem (d) (p. 175):

**Solution to (d) - Constrained approximate optimal bit allocation  
for location parameter estimation**

(d2) For  $i \in \{1, \dots, N_s\}$ , the approximate optimal bit allocation is obtained by choosing  $\nu$  so that (4.47) is satisfied

$$\sum_{i=1}^{N_s} N_{B,i} = \sum_{i=1}^{N_s} [\nu - \log_2(\delta_i)]_+ = N_B.$$

With the value of  $\nu$  satisfying (4.47), the numbers of bits can be obtained using (4.46)

$$N_{B,i} = [\nu - \log_2(\delta_i)]_+.$$

Integer  $N_{B,i}$  can be obtained with appropriate rounding. The corresponding FI can be approximated by substituting the optimal  $N_{B,i}$  in (4.41).

### 4.3 Generalization with the $f$ -divergence

In this section, we will discuss a generalization of the asymptotic results to different inference problems. The generalization that we will study is based on the **generalized  $f$ -divergence (GFD)**, which is presented in [Poor 1988]. The objective of this section is to show the main differences between the asymptotically optimal quantizers for different inference problems.

#### 4.3.1 Definition of the generalized $f$ -divergence

The GFD is a generalization of the  $f$ -divergence (also known as the Ali-Silvey distance) studied in [Ali 1966] (cited in [Poor 1988]). For a continuous r.v.  $Y$ , the GFD  $D_f$  is defined as

$$D_{f,c} = \mathbb{E} \{f[l(Y)]\}, \quad (4.48)$$

where  $l$  is a measurable function and  $f$  is a continuous convex function. For a quantized measurement  $i$  from  $Y$  the GFD is defined as

$$D_{f,q} = \mathbb{E} \{f\{\mathbb{E}_{Y|i}[l(Y)]\}\} = \mathbb{E}[f(l_{q_i})]. \quad (4.49)$$

Developing the conditional expectation and supposing that  $Y$  accepts a PDF  $p(y)$ , we can rewrite (4.49) as

$$D_{f,q} = \sum_{i=1}^{N_I} f(l_{q_i}) \mathbb{P}(i), \quad (4.50)$$

where

$$\mathbb{P}(i) = \int_{q_i} p(y) \, dy, \quad (4.51)$$

and

$$l_{q_i} = \frac{\int_{q_i} l(y) p(y) \, dy}{\int_{q_i} p(y) \, dy}. \quad (4.52)$$

#### 4.3.2 Generalized $f$ -divergence in inference problems

The performance of some important inference problems can be written as a function of the  $f$ -divergence. Three examples are given below.

##### Classical estimation

For classical estimation, we want to estimate a deterministic parameter  $x$  embedded in noisy independent measurements  $Y_{1:N}$ . The quantized version of this problem is the main problem treated in thesis.

Under some regularity conditions, we know that the asymptotic MSE of the optimal unbiased estimator of  $x$  attains the CRB which is given by the inverse of the FI. The FI for  $N$  independent measurements is given by  $N$  times the FI for one measurement.

If we look to the forms of  $I_c$  and  $I_q$  we can see that the FI for one measurement is a GFD with  $l(y) = S_c(y; x)$  and  $f(l) = l^2$ . Therefore, the GFD is directly linked to the asymptotic performance of classical estimation.

### Bayesian estimation

Consider that instead of estimating a deterministic parameter, we want to estimate a random parameter  $X$  based on a noisy measurement  $Y$ .

From Ch. 2 (2.7) (p. 78), we know that  $\text{MSE} = \mathbb{E}_Y [\text{Var}_{X|Y}(X)]$ , which can be rewritten as  $\mathbb{E}_Y \left\{ \mathbb{E}_{X|Y}(X^2) - \mathbb{E}_{X|Y}^2(X) \right\}$ . This gives

$$\text{MSE} = \mathbb{E}(X^2) - \mathbb{E}_Y \left[ \mathbb{E}_{X|Y}^2(X) \right].$$

This function is decreasing w.r.t. the second term, which is a GFD. Proceeding similarly for the quantized measurement version of the problem, we can conclude that the performance depends on a GFD with  $l(y) = \mathbb{E}_{X|Y=y}(X)$  and  $f(l) = l^2$ .

For  $N$  identically distributed measurements, the MSE for Bayesian estimation can also be rewritten as a GFD, but in this case a generalization to the non-scalar case is needed. In [Marano 2007], we can find details for this case with a variable rate approach for quantization. We can also approach Bayesian estimation for  $N$  measurements as a sequential single measurement problem, where at each new observation we can use the last posterior as the new prior. Using this approach for each measurement, the MSE will be given by the scalar version of the GFD explained above.

### Neyman–Pearson detection

We consider now the detection problem. We have  $N$  i.i.d. measurements  $Y_{1:N}$ . The measurements are all obtained from one of two distributions with PDF  $p_0(y)$  or  $p_1(y)$ . Based on the  $N$  measurements we want to decide from which of the two distributions the measurements are obtained. The index of the true measurements distribution will be denoted  $H \in \{0, 1\}$  and the decision that we make based on the  $N$  measurements will be denoted  $\hat{H}$ .

For specifying the performance of the decision procedure we can consider a Neyman–Pearson strategy [Van Trees 1968, p. 33]. In the Neyman–Pearson strategy, we set an upper bound on the probability of deciding  $\hat{H} = 1$  when  $H = 0$  to a fixed constant  $\alpha$  and the performance of the decision procedure is given by the minimum probability  $\beta$  of deciding  $\hat{H} = 0$  when  $H = 1$ . When  $N$  is asymptotically large, the limit of  $\beta$  can be characterized using Stein’s lemma [Blahut 1987] (cited in [Gupta 2003])

$$\lim_{N \rightarrow +\infty} \beta^{\frac{1}{N}} = \exp \{ -D_{KL} [p_0(y) || p_1(y)] \},$$

where  $D_{KL} [p_0(y) || p_1(y)]$  is the **Kullback–Leibler divergence (KLD)**

$$D_{KL} [p_0(y) || p_1(y)] = \int p_0(y) \log \left[ \frac{p_0(y)}{p_1(y)} \right] dy.$$

For quantized measurements, a similar theorem can be stated by replacing  $p_0(y)$  and  $p_1(y)$  by the corresponding probabilities of the quantizer outputs. Therefore, for this problem, the KLD is the criterion to be maximized to increase detection performance. If we take the opposite of the KLD, we can see that it is a GFD with  $l$  the likelihood ratio  $l(y) = \frac{p_0(y)}{p_1(y)}$  and  $f(l) = -\log(l)$ . The expectation in the GFD in this case is evaluated w.r.t. the probability measure for  $H = 0$ .

### Detection of weak signals

We can also consider the detection of a low amplitude signal. We follow a similar presentation as in [Poor 1988]. For this problem, the  $Y_k$  for  $k \in \{1, \dots, N\}$  are i.i.d. and distributed according to  $p(y_i)$  or  $p(y_i - \theta x_i)$ , where  $p$  is the noise marginal PDF and  $x_i$  is a known signal with finite average power  $\bar{x}^2 = \frac{1}{N} \sum_{k=1}^N x_k^2$ . If we consider a large number of measurements  $N \rightarrow \infty$  and small signal amplitude  $\theta \rightarrow 0$ , then the performance of the optimal detector in terms of  $\beta$  in the Neyman–Pearson strategy is related to the efficacy

$$\rho = \bar{x}^2 \int \left[ \frac{\frac{dp(y)}{dy}}{p(y)} \right]^2 p(y) dy = \bar{x}^2 \int \left\{ \frac{d \log [p(y)]}{dy} \right\}^2 p(y) dy.$$

When this quantity is maximized, we maximize asymptotic detection performance. Note that the integral factor is exactly the FI for estimating a location parameter of the PDF  $p$ . Thus in this case, the inference performance can also be written as a GFD with  $l(y) = \frac{dp(y)}{p(y)} = \frac{d \log [p(y)]}{dy}$  and  $f(l) = l^2$ .

### 4.3.3 Asymptotic results

Similarly to the asymptotic development for the FI, we will write asymptotic approximations for the loss of GFD incurred by quantization. After obtaining the asymptotic loss for the GFD, we will obtain the optimal interval densities for the fixed rate and variable rate encoding cases.

#### Asymptotic GFD loss

The loss of GFD due to quantization can be defined as

$$L_f = D_{f,c} - D_{f,q} = \sum_{i=1}^{N_I} L_{f,i}, \quad (4.53)$$

where  $L_{f,i}$  is the loss for each quantization interval

$$L_{f,i} = \int_{q_i} \{f[l(y)] - f(l_{q_i})\} p(y) dy. \quad (4.54)$$

For obtaining the asymptotic approximation, we write the Taylor series expansions of  $l$  and  $p$  around the central point  $y_i$  and of  $f$  around a point  $l_i$

$$l(y) = l_i + l_i^{(y)}(y - y_i) + \frac{l_i^{(yy)}}{2}(y - y_i)^2 + o(y - y_i)^2, \quad (4.55)$$

$$p(y) = p_i + p_i^{(y)}(y - y_i) + \frac{p_i^{(yy)}}{2}(y - y_i)^2 + o(y - y_i)^2, \quad (4.56)$$

$$f(l) = f_i + f_i^{(l)}(l - l_i) + \frac{f_i^{(ll)}}{2}(l - l_i)^2 + o(l - l_i)^2. \quad (4.57)$$

Using (4.57) and (4.55), the function  $f[l(y)]$  on the interval  $q_i$  can be written as

$$f[l(y)] = f_i + f_i^{(l)} \left[ l_i^{(y)}(y - y_i) + \frac{l_i^{(yy)}}{2}(y - y_i)^2 \right] + \frac{f_i^{(ll)}}{2} \left[ \left( l_i^{(y)} \right)^2 (y - y_i)^2 \right] + o(y - y_i)^2. \quad (4.58)$$

We use (4.55) and (4.56) in (4.52) to evaluate  $l_{q_i}$

$$\begin{aligned} l_{q_i} &= \\ &= \frac{\int_{q_i} \left[ l_i + l_i^{(y)}(y - y_i) + \frac{l_i^{(yy)}}{2}(y - y_i)^2 + o(y - y_i)^2 \right] \left[ p_i + p_i^{(y)}(y - y_i) + \frac{p_i^{(yy)}}{2}(y - y_i)^2 + o(y - y_i)^2 \right] dy}{\int_{q_i} \left[ p_i + p_i^{(y)}(y - y_i) + \frac{p_i^{(yy)}}{2}(y - y_i)^2 + o(y - y_i)^2 \right] dy} \\ &= \frac{l_i p_i \Delta_i + l_i \frac{p_i^{(yy)}}{2} \frac{\Delta_i^2}{12} + l_i^{(y)} p_i^{(y)} \frac{\Delta_i^3}{12} + \frac{l_i^{(yy)}}{2} p_i \frac{\Delta_i^3}{12} + o(\Delta_i^3)}{p_i \Delta_i + \frac{p_i^{(yy)}}{2} \frac{\Delta_i^3}{12} + o(\Delta_i^3)}. \end{aligned} \quad (4.59)$$

To evaluate  $f[l_{q_i}]$  we will replace (4.59) in (4.58). We proceed first by evaluating  $l_{q_i} - l_i$

$$l_{q_i} - l_i = \frac{l_i^{(y)} p_i^{(y)} \frac{\Delta_i^3}{12} + \frac{l_i^{(yy)}}{2} p_i \frac{\Delta_i^3}{12} + o(\Delta_i^3)}{p_i \Delta_i + \frac{p_i^{(yy)}}{2} \frac{\Delta_i^3}{12} + o(\Delta_i^3)}.$$

Note that  $l_{q_i} - l_i$  has a factor  $\Delta_i^2$ , thus  $(l_{q_i} - l_i)^2 = o(\Delta_i^3)$ . This leads to

$$f[l_{q_i}] = f_i + f_i^{(l)} \left[ \frac{l_i^{(y)} p_i^{(y)} \frac{\Delta_i^3}{12} + \frac{l_i^{(yy)}}{2} p_i \frac{\Delta_i^3}{12} + o(\Delta_i^3)}{p_i \Delta_i + \frac{p_i^{(yy)}}{2} \frac{\Delta_i^3}{12} + o(\Delta_i^3)} \right] + o(\Delta_i^3). \quad (4.60)$$

Now, we will evaluate the two terms in  $L_{f,i}$  (4.54). Multiplying the expansion of  $f[l(y)]$  (4.58) by the expansion of  $p(y)$  (4.56) and integrating, we obtain

$$\begin{aligned} \int_{q_i} f[l(y)] p(y) dy &= f_i p_i \Delta_i + f_i \frac{p_i^{(yy)}}{2} \frac{\Delta_i^3}{12} + f_i^{(l)} \left[ l_i^{(y)} p_i^{(y)} \frac{\Delta_i^3}{12} + \frac{l_i^{(yy)}}{2} p_i \frac{\Delta_i^3}{12} \right] \\ &\quad + \frac{f_i^{(ll)}}{2} \left( l_i^{(y)} \right)^2 \frac{\Delta_i^3}{12} p_i + o(\Delta_i^3). \end{aligned} \quad (4.61)$$

Using (4.60) and integrating the expansion for  $p(y)$ , we get

$$\int_{q_i} f[l_{q_i}] p(y) dy = f_i p_i \Delta_i + f_i \frac{p_i^{(yy)}}{2} \frac{\Delta_i^3}{12} + f_i^{(l)} \left[ l_i^{(y)} p_i \frac{\Delta_i^3}{12} + \frac{l_i^{(yy)}}{2} p_i \frac{\Delta_i^3}{12} \right] + o(\Delta_i^3). \quad (4.62)$$

Subtracting (4.62) from (4.61), we get the loss in the interval  $q_i$

$$L_{f,i} = \frac{f_i^{(ll)}}{2} \left( l_i^{(y)} \right)^2 \frac{\Delta_i^3}{12} p_i + o(\Delta_i^3).$$

Therefore, the total loss is

$$L_f = \sum_{i=1}^{N_I} \left[ \frac{f_i^{(ll)}}{2} \left( l_i^{(y)} \right)^2 \frac{\Delta_i^3}{12} p_i + o(\Delta_i^3) \right]. \quad (4.63)$$

Similarly to the asymptotic development for the FI, we have

$$\lim_{N_I \rightarrow \infty} N_I^2 L_f = \frac{1}{24} \int \frac{f^{(ll)} [l(y)] [l^{(y)}(y)]^2 p(y)}{\lambda^2(y)} dy. \quad (4.64)$$

The optimal interval density for fixed rate encoding is then given by

$$\lambda^*(y) = \frac{f^{(ll) \frac{1}{3}} [l(y)] [l^{(y)}(y)]^{\frac{2}{3}} p^{\frac{1}{3}}(y)}{\int f^{(ll) \frac{1}{3}} [l(y)] [l^{(y)}(y)]^{\frac{2}{3}} p^{\frac{1}{3}}(y) dy}. \quad (4.65)$$

If the PDF of the measurements is completely known and given by  $p(y)$ , then a similar development as it was done for the FI leads to the following optimal variable rate encoding interval density:

$$\lambda_{vr}^*(y) = \frac{\sqrt{f^{(ll)} [l(y)]} |l^{(y)}(y)|}{\int \sqrt{f^{(ll)} [l(y)]} |l^{(y)}(y)| dy}.$$

#### 4.3.4 Interval densities for inference problems

We will compare now the different interval densities for the inference problems described above. In Tab. 2 we give the different functions defining the GFD for each problem and the corresponding optimal interval density. We also give the optimal interval density for variable rate encoding, whenever variable rate encoding is possible.

Inference problem	$l(y)$	$f(l)$	$\lambda^*(y) \propto$	$\lambda_{vr}^*(y) \propto$
<b>Classical estimation</b>	$S_c(y; x)$	$l^2$	$\left(\frac{\partial S_c(y; x)}{\partial y}\right)^{\frac{2}{3}} p^{\frac{1}{3}}(y; x)$	–
<b>Bayesian estimation</b>	$\mathbb{E}_{X Y=y}(X)$	$l^2$	$\left(\frac{d\mathbb{E}_{X Y=y}(X)}{dy}\right)^{\frac{2}{3}} p^{\frac{1}{3}}(y)$	$\left \frac{d\mathbb{E}_{X Y=y}(X)}{dy}\right $
<b>N–P detection</b>	$\frac{p_0(y)}{p_1(y)}$	$-\log(l)$	$\left\{\frac{d \log\left[\frac{p_0(y)}{p_1(y)}\right]}{dy}\right\}^{\frac{2}{3}} p_0^{\frac{1}{3}}(y)$	–
<b>Weak signal detection</b>	$\frac{dp(y)}{dy}$	$l^2$	$\left\{\frac{d^2 \log[p(y)]}{dy^2}\right\}^{\frac{2}{3}} p^{\frac{1}{3}}(y)$	$\left \frac{d^2 \log[p(y)]}{dy^2}\right $

Table 4.2: Functions characterizing the GFD for different inference problems and interval densities maximizing the inference performance based on quantized measurements. The interval density  $\lambda^*(y)$  is the density optimizing the performance when encoding is done with fixed rate,  $\lambda_{vr}^*(y)$  is the density for variable rate encoding.

Notice that for Bayesian estimation and weak signal detection, we give expressions for the variable rate optimal density. In Bayesian estimation, as we have a prior on the true parameter, we know the probabilities of the quantizer outputs, thus we can define correct lengths for the outputs. In weak signal detection, as the amplitude of the signal is small, we can consider that the encoding can be done approximately by using the noise distribution.

While in classical estimation, we have the effect of the score function derivative, in Bayesian estimation the optimal interval density is affected by the optimal estimator  $\hat{x} = \mathbb{E}_{X|Y=y}(X)$  function. Note also that, differently from the classical estimation case, where the interval density is affected directly by the true parameter value, in Bayesian estimation the influence of the parameter appears only through its prior. Thus even if  $x$  is unknown in the Bayesian case, an optimal quantizer can be implemented in practice<sup>3</sup>.

Observe that classical estimation for a location parameter with value  $x = 0$  and weak signal detection have exactly the same interval density. Actually, the performance of weak signal detection can be seen equivalently as the performance of estimating a small constant with i.i.d noise and marginal PDF  $p(y)$ . Thus it is not surprising that the optimal interval densities are the same.

The optimal density for Neyman–Pearson detection that we have obtained here is exactly the same as the one obtained in [Gupta 2003] in the scalar case. Note that similarly to Bayesian estimation, where the sensibility of the key element for inference, the optimal estimator, has a direct impact on the interval density, in detection, the sensibility of the logarithm of the continuous measurement likelihood ratio plays an important role. Note also that the log-likelihood ratio  $\log\left[\frac{p_0(y)}{p_1(y)}\right] = \log[p_0(y)] - \log[p_1(y)]$  for two distributions parametrized by  $x$  and  $x + \varepsilon$  with small  $\varepsilon$  can be rewritten using an expansion around  $x$

$$\log[p_0(y)] - \log[p_1(y)] = \log[p(y; x)] - \log[p(y; x + \varepsilon)] = \varepsilon \frac{\partial \log[p(y; x)]}{\partial x} + o(\varepsilon).$$

The optimal interval density is then approximately given by the optimal density for classical estimation. This makes explicit the link between the density for weak signal detection and for

<sup>3</sup>Optimal in this case for a given prior, if the prior does not represent well the reality, then the Bayesian setting is not useful and optimality is meaningless.

classical estimation.

## 4.4 Chapter summary and directions

We summarize the main points from this chapter and possible directions for future work:

- We developed an asymptotic high-rate approximation for the FI for quantized measurements. The approximation shows that the FI for quantized measurements tends to the FI for continuous measurements when the number of quantization intervals tends to infinity. When the quantizer outputs are all coded with binary words with the same length (fixed rate encoding), the approximation of the FI tends exponentially to the FI for continuous measurements as a function of the number of quantization bits.
- The asymptotic performance approximation obtained depends on the specific choice of the quantizer intervals through the quantizer interval density. For fixed rate encoding, the optimal interval density is shown to depend not only on the PDF through  $f^{\frac{1}{3}}$ , as it is common in standard quantization, but also on the derivative of the score function.

In practice for finite number of bits, the optimal interval density can be approximated by setting the quantization thresholds using the inverse of the CDF related to the interval density. As the CDF depends on the parameter that we want to estimate, a recursive procedure for joint estimating and resetting the thresholds is necessary for obtaining asymptotically optimal performance (asymptotic both in  $N$  and  $N_I$ ). For example, we can use the adaptive algorithms presented in Ch. 3 when we want to solve a location estimation problem. In general we can use the adaptive MLE approach.

When the length of the binary words are chosen to minimize the mean length of the quantizer output, the optimal density is shown to depend directly on the derivative of the score function. The problem with this approach is that not only setting the quantizer thresholds depends on the measurement distribution, but also the encoding method depends on it. Even if we can attain the best asymptotic performance by using an adaptive technique for setting the thresholds, we will not respect the rate constraint during all the initial time of the estimation procedure, when the parameter estimate is far from the true parameter value.

- The practical approximation of the asymptotically optimal quantization thresholds was obtained for the estimation of location and scale parameters of the GGD. For the STD, we obtained the practical approximation in the Cauchy case for location and in general for scale.

The asymptotic results were tested in the location problem with the Gaussian and Cauchy distributions. We compared the asymptotic approximation of the FI with the FI for optimal uniform quantization and with the FI for the practical approximation of the optimal thresholds. We observed that, with only 4 bits, the FI obtained with the practical approximation is very close to the asymptotic approximation. We also observed that, in the Gaussian case, the gain of performance obtained with nonuniform quantization is

negligible, while in the Cauchy case it is small. This indicates that in practice uniform quantization might be a better solution, as it requires a lower complexity.

By using the adaptive algorithm, we have shown that the asymptotically optimal results can be obtained in practice. During the simulation of the adaptive algorithm, it was observed that uniform quantization leads to faster convergence when compared with nonuniform quantization. An interesting point for future research is then to study adaptive algorithms that start with a threshold set optimized for faster convergence and then change the threshold set, so that asymptotically the performance is also optimal.

- By using the asymptotic results, we have obtained approximations for the optimal bit allocation when we estimate a location parameter using multiple sensors and the total number of quantization bits is constrained.

The first approximate solution was given by considering unconstrained numbers of bits (positive and negative reals), the approximate optimal bit allocation is shown to be a correction on the uniform bit allocation (equal number of bits for each sensor) that depends on the weight of the noise scale parameter on the geometric mean of all the scale parameters. The FI given by this approximate optimal bit allocation is shown to depend on the harmonic and geometric means of the noise scale parameters. An interesting point for future work is the analysis of the approximate FI for the optimal allocation when the number of sensors is very large and the sensors scale parameters are random with a given known distribution. As the approximate FI depends on the geometric and harmonic means of the scale parameters, by using the law of large numbers, we expect to obtain an approximation of the FI depending on the parameters of the scale distribution.

The second approximation was given by considering a more realistic scenario, with the numbers of bits constrained to be positive. The approximate optimal bit allocation is given by a water-filling solution, which is a well known solution for the problem of power allocation in multicarrier modulations. For the bit allocation problem, the logarithm of the scale parameter plays the equivalent role of the noise power in multicarrier power allocation and the number of bits plays the role of the power to be allocated.

The water-filling solution depends on the scale parameters, in a fusion center approach, the fusion center can use the estimates of the scale parameters to obtain an approximate solution. The solution of the problem is mainly determined by only one parameter, the "water level", after obtaining the approximate "water level", the fusion center can broadcast it to the sensors so that they can set their quantizer resolution. A problem that still need to be solved in this case is how the sensors will coordinate their final choice on the numbers of bits (which are constrained to be integers) so that the total rate constraint is respected.

- As a final point of this chapter, we revisited the asymptotic approximation for the  $f$ -divergence loss due to quantization presented in [Poor 1988]. The objective of this part was to show that the asymptotic approximation of the FI presented in this chapter can be seen as a special case of the asymptotic approximation of a general performance measure for inference problems and to show the links between the asymptotic characterization of the quantizers for different inference problems.

We saw that there is a close link between quantization for weak signal detection and for classical estimation of a location parameter. In practice, as we will use an adaptive algorithm with static quantizer centered at zero for classical estimation, the quantizer thresholds for these two problems are exactly the same. The link between Neyman–Pearson detection and Bayesian estimation is that, for both, the quantizer depends on the sensibility of their key quantities: the Neyman–Pearson detection optimal interval density depends on the sensibility of the log-likelihood ratio and the Bayesian estimation optimal density depends on the sensibility of the estimator.

- Additional to the points for further study presented above many other points can also be investigated:
  - The *vector quantizer extension* of the asymptotic approximation of the FI can be considered: vector quantization is the most natural extension of the results presented here.
  - *Further study of the Bayesian case*: for one sample asymptotic characterization, we saw that a recursive approach can be used. In practice, this solution may be too complex to be implemented as we need to evaluate completely the continuous measurement estimator and consequently the posterior for obtaining the optimal quantizer at each sample. For obtaining a simple solution, we can consider high resolution quantizers that are designed to optimize the asymptotic (large number of samples) performance of Bayesian estimation.
  - *Dealing with the overload region*: a main point that was neglected in the analysis is that, in practice, most noise PDF that are used for modeling have infinite support. In this chapter we considered explicitly that the noise PDF have bounded support, so that it would not be necessary to deal with the overload region. In future work, we can try to deal with the overload region.
  - *Asymptotic approximation of the optimal uniform quantizer for estimation*: during all this thesis we considered the explicit optimization through a grid search of the optimal quantization step in uniform quantization. We can try to obtain an analytic characterization of the optimal step by considering an asymptotic approach.



# Conclusions of Part II

In Part II, we studied the asymptotic performance of estimation of a scalar deterministic parameter based on quantized measurements. Asymptotically in this case means:

- that the number of samples tends to infinity  $N \rightarrow \infty$ , so that we can use the FI to characterize the estimation performance.
- That the number of quantization bits tends to infinity  $N_B \rightarrow \infty$ , so that we can use high-rate approximations of the FI to determine analytically the loss of performance induced by quantization.

We obtained the following conclusions:

- **The asymptotic loss of performance due to quantization decreases exponentially as a function of the number of bits.** The loss of FI due to quantization is shown to decrease exponentially with increasing numbers of bits. Even if the results are asymptotic, they indicate that it is probably not useful to increase the sensor quantizer resolution when a target performance is not met. Probably, it is more reasonable, as we saw in Part I, to increase the number of sensors, or if it is possible, to increase sampling frequency and to use sensors with smaller noise amplitude (smaller noise scale factor).
- **Asymptotic may be low to medium resolution in practice.** Using a practical approximation of the asymptotically optimal thresholds for finite number of quantization intervals in the location estimation problem (Gaussian and Cauchy cases), we have shown that the corresponding FI is very close to the asymptotic approximation of the FI for numbers of bits as low as 4. For 1,2 and 3 bits the optimal threshold variations can be found easily by grid search and the central threshold can be adjusted in all cases with an adaptive algorithm. This means that in practice, for all numbers of bits, we can set, at least approximately, the quantizer thresholds to have asymptotically optimal quantization for location parameter estimation under Gaussian and Cauchy distributions.

A question that still remains unanswered is if this is true in general, for different types of measurement distribution and for the estimation of other types of parameters.

- **Uniform is not bad at all.** Although we can use, in practice, nonuniform quantization of the measurements to have asymptotically optimal performance. The gap between the performance for optimal uniform quantization and the performance for nonuniform quantization in location parameter estimation is small. As uniform quantization is easier to be implemented, it seems that, in practice, uniform quantization may be a better solution.



# Conclusions

## Main conclusions

In this thesis, we have studied the problem of estimation based on quantized measurements, a problem that has attracted increasing attention of the signal processing community due to the emergence of sensor networks. More specifically, we treated the problem of estimating a scalar parameter, either constant or varying with a Wiener model, based on quantized noisy measurements of the parameter.

We observed that for most commonly used noise models, the estimation performance degrades when the quantizer dynamic range is far from the true parameter value, indicating that a good solution can be obtained by adaptively setting the quantizer range using the most recent estimate of the parameter.

Using the adaptive scheme, the loss of estimation performance due to quantization seems to be small. For all the tested cases (different noise PDF, constant or slowly varying parameter), a small loss is observed when we use 1–3 quantization bits and a negligible loss is observed for 4 or 5 quantization bits. This indicates that the solution of the remote sensing problem under constrained communication rate is linked to low resolution sensor networks:

- If we consider that the problem is constrained to be solved with a sensor network approach, then from the results above, we can see that quantization with low resolution is a solution to this problem.
- If we constrain the problem to be a remote sensing problem based on quantized measurements, then a low resolution sensor network approach seems to be an appropriate solution.

As the standard estimation algorithms for attaining the small loss of performance have high complexity, we proposed a low complexity adaptive algorithm that achieves asymptotically the same performance. Extensions of the algorithm were proposed for the cases when the noise scale factor is unknown and when multiple sensors are available.

We also studied the problem of how to set the quantization thresholds for obtaining optimal estimation performance when a large number of quantization intervals is available. We used the asymptotic approach (the quantizer intervals tend to zero) to obtain an approximation of the optimal thresholds, this approach also allowed to obtain an approximate analytical expression for the estimation performance (the FI) as a function of the number of quantization bits. The approximation of the FI for quantized measurements is shown to converge exponentially to the FI for continuous measurements. The approximate analytic expression was shown to be valid in the location estimation problem even for small numbers of bits (4 in this case), indicating that the result, which is expected to be exact only when the number of bits tends to infinity, can be useful in practice, if we consider non uniform quantization.

From the asymptotic approach, we show that the optimal thresholds may depend on the parameter, which is unknown. This reinforces the importance of the adaptive approach, which allows to set the thresholds asymptotically according to their optimal values, leading to asymptotically optimal estimation performance.

We also want to point out that the difference between using the optimal general threshold scheme (non uniform) and the optimal uniform scheme for the location problem is small. In practice, if low complexity is needed, then uniform quantization may be a better solution.

## Perspectives

We finish this "conversation" between quantization and estimation, highlighting some subjects for future discussion. Some details of these subjects were already discussed at the end of the chapters, therefore here we give only the main lines.

- *Vector parameter and vector quantization:* this is the direct extension of the problem, while the vector quantization extension might be straightforward to study, both in terms of proposing algorithms and studying their asymptotic (in terms of numbers of samples and quantization intervals) behavior, the vector parameter extension seems to be less straightforward, specially because it would require a redefinition of the estimation performance and it would require a full extension of the algorithms to vectors, for exploiting correctly the correlation between the components.
- *Noisy channels:* in the "DSP party", most of the time, communication is not invited, we can propose to invite it to the next party by adding the communication channel in the problem. A noisy communication channel can be considered in multiple ways. The simplest way for introducing it in the problem is by indexing the quantized measurements with binary words and then considering the channel as an extension of the binary symmetric channel. While for a fixed indexing the extensions of the algorithms, specially the low complexity one proposed here, might be simple, the problem of optimal estimation/indexing can be difficult.

Different extensions can be considered by introducing a continuous channel, for example additive and fading channels. In this case we might consider the problem of indexing, by assigning real values to the quantized measurements, this will generate again a joint problem of estimation/codebook design.

- *Estimation under unknown noise distribution:* we supposed that the noise distribution is known, at least up to a scale factor, however, in practice, this assumption cannot be always satisfied and we will need to look for different approaches to estimate the location parameter based on quantized measurements.

There are other topics that were not discussed explicitly in this thesis, but they are interesting subjects for future research. They are the following:

- 
- *Fast variations:* to develop some parts of this thesis, we considered that the parameter to be estimated was a slowly varying Wiener process, under this hypothesis we have shown that the loss of performance due to quantization is small. The unanswered question here is whether this conclusion is true or false for the estimation of fast varying processes.
  - *Distributed problem:* in this thesis, we treated the simplified remote sensing problem, where we have only one sensor. In the only case where a multiple sensor approach was treated, we used the fusion center approach. Thus, we still need to generalize the concepts and algorithms developed here to a partially or completely distributed setting, where a cluster head or each sensor wants to obtain estimates based on the information from all the sensors.
  - *Continuous time:* for a varying parameter, we considered that the parameter model was inherently discrete and we did not discuss sampling issues. Thus a subject to be studied is the estimation of a continuous process based on sampled and quantized measurements.



# Appendices

---

## A.1 Why? - Proofs

### A.1.1 Proof that $\mathbb{E}[S_c S_q] = \mathbb{E}[S_q^2]$

We will consider a general parameter estimation problem in the proof. The density of the measurement will be  $f(y; x)$  instead of  $f(y - x)$ . Adding the dependence of  $S_q$  on the quantizer output index  $i$ ,  $y$  and  $x$ , the expectation of the product is

$$\mathbb{E}[S_c S_q] = \int_{\mathbb{R}} \frac{\partial \log f(y; x)}{\partial x} S_q(i(y); x) f(y; x) dy. \quad (\text{A.1})$$

Separating the integral in (A.1) in a sum of integrals on the different quantization intervals  $q_i$ :

$$\mathbb{E}[S_c S_q] = \sum_{i \in \mathcal{I}} \int_{q_i} \frac{\partial \log f(y; x)}{\partial x} S_q(i(y); x) f(y; x) dy.$$

$S_q$  is a constant function inside an interval  $q_i$ , thus, in an interval, it does not depend on  $y$  and it can leave the integral

$$\mathbb{E}[S_c S_q] = \sum_{i \in \mathcal{I}} S_q(i; x) \int_{q_i} \frac{\partial \log f(y; x)}{\partial x} f(y; x) dy.$$

Rewriting the continuous measurement score function in ratio form gives

$$\mathbb{E}[S_c S_q] = \sum_{i \in \mathcal{I}} S_q(i; x) \int_{q_i} \frac{\frac{\partial f(y; x)}{\partial x}}{f(y; x)} f(y; x) dy,$$

supposing that we can change the order of integral and the partial derivative leads to

$$\mathbb{E}[S_c S_q] = \sum_{i \in \mathcal{I}} S_q(i; x) \frac{\partial \mathbb{P}(i; x)}{\partial x}.$$

Multiplying and dividing each term of the sum by its corresponding  $\mathbb{P}(i; x)$ , we have

$$\mathbb{E}[S_c S_q] = \sum_{i \in \mathcal{I}} S_q(i; x) \frac{\frac{\partial \mathbb{P}(i; x)}{\partial x}}{\mathbb{P}(i; x)} \mathbb{P}(i; x).$$

We can identify the score function as the second factor, leading to

$$\mathbb{E}[S_c S_q] = \sum_{i \in \mathcal{I}} S_q^2(i; x) \mathbb{P}(i; x) = \mathbb{E}[S_q^2]. \quad (\text{A.2})$$

### A.1.2 Proof of the upper bound on $F(\varepsilon)[1 - F(\varepsilon)]$ for the Gaussian distribution

We can write  $F(\varepsilon)[1 - F(\varepsilon)]$  as the probability of two i.i.d. Gaussian r.v.  $X_1$  and  $X_2$  to be in the respective intervals  $[-\infty, x]$  and  $[x, \infty]$ . Thus, this probability can be written as the integral of their joint PDF (see (1.24) for the marginal Gaussian PDF form)

$$f_{1,2}(x_1, x_2) = f(x_1)f(x_2) = \frac{1}{\pi\delta^2} \exp\left[-\frac{(x_1^2 + x_2^2)}{\delta^2}\right]$$

on the area  $A_0 + A_1$  of Fig. A.1. From the i.i.d. assumption, the integral on the area  $A_1$  is equal to the integral on the area  $A'_1$ . Therefore,  $F(\varepsilon)[1 - F(\varepsilon)]$  is equal to the integral of  $f_{1,2}(x_1, x_2)$  on  $A_0 + A'_1$ . It is easy to see that the area outside the quarter circle  $C_1$  in the fourth quadrant is not smaller than the area of  $A_0 + A'_1$ . Denoting the area outside the quarter circle in the fourth quadrant by  $\bar{C}_1$ , we can say that  $\mathbb{P}(X_1, X_2 \in A_0 + A_1) \leq \mathbb{P}(X_1, X_2 \in \bar{C}_1)$ . Changing the coordinates from rectangular  $(x_1, x_2)$  to polar  $(r, \theta)$ , where  $r = \sqrt{x_1^2 + x_2^2}$  is the radius and  $\theta = \arctan\left(\frac{x_1}{x_2}\right)$  is the angle, we have that

$$\mathbb{P}(X_1, X_2 \in \bar{C}_1) = \int_{-\frac{\pi}{2}}^0 \int_x^\infty \frac{1}{\pi\delta^2} r \exp\left[-\frac{(r^2)}{\delta^2}\right] dr d\theta = \frac{1}{2\delta^2} \int_x^\infty r \exp\left[-\frac{(r^2)}{\delta^2}\right] dr.$$

Changing variables one more time  $r' = \frac{r}{\delta}$ , we obtain

$$\mathbb{P}(X_1, X_2 \in \bar{C}_1) = \frac{1}{2} \int_{\frac{x}{\delta}}^\infty r' \exp(-r'^2) dr' = -\frac{1}{4} \int_{\frac{x}{\delta}}^\infty -2r' \exp(-r'^2) dr' = \frac{1}{4} \exp\left[-\left(\frac{x}{\delta}\right)^2\right].$$

Consequently,

$$F(\varepsilon)[1 - F(\varepsilon)] = \mathbb{P}(X_1, X_2 \in A_0 + A_1) \leq \mathbb{P}(X_1, X_2 \in \bar{C}_1) = \frac{1}{4} \exp\left[-\left(\frac{x}{\delta}\right)^2\right].$$

### A.1.3 Proof that the FI for estimating a Laplacian location parameter with noise scale $\delta$ is $\frac{1}{\delta^2}$ .

The score function (1.15) for the location parameter of the Laplacian distribution is (PDF given by (1.27)):

$$S_c = \frac{\partial \log f(y-x)}{\partial x} = \frac{\partial \left[ \log\left(\frac{1}{2\delta^2}\right) - \left|\frac{y-x}{\delta}\right| \right]}{\partial x} = \frac{1}{\delta} \text{sign}(y-x),$$

where we used the fact that the derivative of the absolute value function is the sign function. The FI is then given by

$$I_c = \mathbb{E}[S_c^2] = \int_{-\infty}^{+\infty} \frac{1}{\delta^2} \frac{1}{2\delta} \exp\left(-\left|\frac{y-x}{\delta}\right|\right) dy.$$

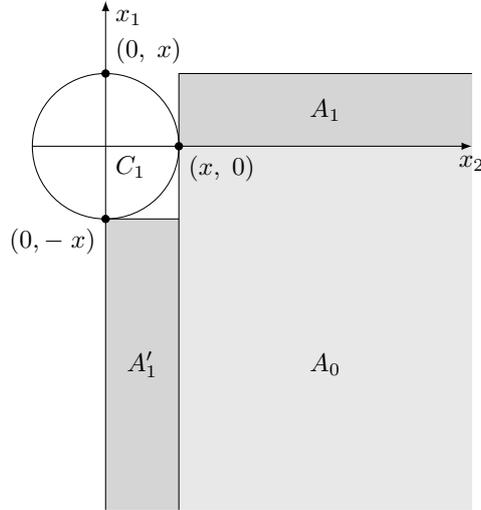


Figure A.1: Geometric scheme to show that the probability of the interval  $A_0 + A_1$  is less than the probability of the exterior region of the left quarter circle  $C_1$ .

Changing variables  $y' = \frac{y-x}{\delta}$  and using the symmetry of  $\exp(-|y'|)$ , we get

$$I_c = \frac{1}{\delta^2} \int_0^{+\infty} \exp(-y') dy' = \frac{1}{\delta^2}.$$

**A.1.4 Proof that the FI for estimating a Cauchy location parameter with noise scale  $\delta$  is  $\frac{1}{2\delta^2}$ .**

The score function (1.15) for the location parameter of a Cauchy distribution (PDF given by (1.33)) is the following:

$$S_c = \frac{\partial \log f(y-x)}{\partial x} = \frac{\partial \left[ -\log(\pi\delta) - \log \left[ 1 + \left( \frac{y-x}{\delta} \right)^2 \right] \right]}{\partial x} = \frac{\frac{2}{\delta} \left( \frac{y-x}{\delta} \right)}{\left[ 1 + \left( \frac{y-x}{\delta} \right)^2 \right]}.$$

The FI can be evaluated then with the following integral

$$\begin{aligned} I_c &= \mathbb{E} [S_c^2] = \frac{4}{\pi\delta^2} \int_{-\infty}^{+\infty} \frac{\left( \frac{y-x}{\delta} \right)^2}{\left[ 1 + \left( \frac{y-x}{\delta} \right)^2 \right]^3} dy \\ &= \frac{8}{\pi\delta^3} \int_x^{+\infty} \frac{\left( \frac{y-x}{\delta} \right)^2}{\left[ 1 + \left( \frac{y-x}{\delta} \right)^2 \right]^3} dy, \end{aligned}$$

where the second equality comes from the symmetry of the integrand. Changing variables  $\tan(\theta) = \frac{y-x}{\delta}$ . We must change  $dy = \delta \sec^2(\theta) d\theta$  and the integration limits also change to 0

and  $\frac{\pi}{2}$ . Using the trigonometric identity  $1 + \tan^2(\theta) = \sec^2(\theta)$ , we have

$$I_c = \frac{8}{\pi\delta^2} \int_0^{\frac{\pi}{2}} \frac{\tan^2(\theta)}{\sec^6(\theta)} \sec^2(\theta) d\theta = \frac{8}{\pi\delta^2} \int_0^{\frac{\pi}{2}} \sin^2(\theta) \cos^2(\theta) d\theta.$$

Using trigonometric identities, we have that  $\sin^2(\theta) \cos^2(\theta) = \frac{1}{8} [1 - \cos(4x)]$ . The integral of the term  $\cos(4x)$  is zero on the interval  $[0, \frac{\pi}{2}]$ . Therefore, we finally obtain

$$I_c = \frac{1}{\pi\delta^2} \int_0^{\frac{\pi}{2}} d\theta = \frac{1}{2\delta^2}.$$

### A.1.5 Proof that the FI for $N$ measurements quantized adaptively with $N_I$ quantization intervals is $I_q^{N_I} = \sum_{k=1}^{N_I} \mathbb{E} [I_q(\varepsilon_k)]$ .

Making more explicit the dependence of  $\mathbb{P}(i_{1:N}; x)$  on the adaptive central thresholds  $\tau_{0,0:N-1}$  by the conditional probability  $\mathbb{P}(i_{1:N} | \tau_{0,0:N-1}; x)$  and exploiting the independence between the measurements conditioned on the central threshold used to obtain them, we can write that the joint probability used in the score function evaluation factorizes as follows:

$$\mathbb{P}(i_{1:N} | \tau_{0,0:N-1}; x) = \prod_{k=1}^N \mathbb{P}(i_k | \tau_{0,k-1}; x).$$

Thus, the log-likelihood is given by

$$\log L(x; i_{1:N}) = \sum_{k=1}^N \log \mathbb{P}(i_k | \tau_{0,k-1}; x).$$

The FI is then given by

$$\begin{aligned} I_q^{N_I} &= \mathbb{E} \left\{ \left[ \frac{\partial \log L(x; i_{1:N})}{\partial x} \right]^2 \right\} = \mathbb{E} \left\{ \left\{ \frac{\partial \left[ \sum_{k=1}^N \log \mathbb{P}(i_k | \tau_{0,k-1}; x) \right]}{\partial x} \right\}^2 \right\} \\ &= \mathbb{E} \left\{ \left[ \sum_{k=1}^N \frac{\partial \log \mathbb{P}(i_k | \tau_{0,k-1}; x)}{\partial x} \right]^2 \right\}, \end{aligned}$$

where the expectation is evaluated w.r.t. the joint probability measure of the r.v.  $i_{1:N}$  and  $\tau_{0,0:N-1}$ . We can decompose the joint expectation in a composition of two expectations using conditioning. For 2 r.v.  $X$  and  $Y$  and a function  $h$ , this is

$$\mathbb{E}_{X,Y} [h(Y, X)] = \mathbb{E}_X \{ \mathbb{E}_{Y|X} [h(X, Y)] \}.$$

The subscripts indicate the corresponding probability measure used for the evaluation. For example,  $X|Y$  corresponds to the conditional probability measure of  $X$  given  $Y$ . Using this decomposition on the FI above:

$$I_q^{NI} = \mathbb{E}_{\tau_{0,0:k-1}} \left\{ \mathbb{E}_{i_{1:k}|\tau_{0,0:k-1}} \left\{ \left[ \sum_{k=1}^N \frac{\partial \log \mathbb{P}(i_k|\tau_{0,k-1};x)}{\partial x} \right]^2 \right\} \right\}.$$

By expanding the square of the inner sum, we have that the inner expectation is a sum of expectations of squared score functions  $\left[ \frac{\partial \log \mathbb{P}(i_k|\tau_{0,k-1};x)}{\partial x} \right]^2$  and products of score functions for different samples  $\frac{\partial \log \mathbb{P}(i_k|\tau_{0,k-1};x)}{\partial x} \frac{\partial \log \mathbb{P}(i_j|\tau_{0,j-1};x)}{\partial x}$  with  $j \neq k$ . As the samples are conditionally independent given their central thresholds, the conditional expectations of the squared scores are equal to the sum of conditional expectations, each conditional expectation will be evaluated with the probability measure of its corresponding  $i_k|\tau_{0,k-1}$ . For the crossed terms the same happens, but now each conditional expectation will be evaluated with respect to  $i_{k,j}|\tau_{0,k,j}$ , as the pairs of measurements are conditionally independent, the conditional expectation of the product of scores is the product of conditional expectations. Finally, as the expectation of each score function is zero [Kay 1993, pp. 67], the expectation of the sum of cross products is zero. Therefore, we have

$$I_q^{NI} = \mathbb{E}_{\tau_{0,0:k-1}} \left\{ \sum_{k=1}^N \mathbb{E}_{i_k|\tau_{0,k-1}} \left\{ \left[ \frac{\partial \log \mathbb{P}(i_k|\tau_{0,k-1};x)}{\partial x} \right]^2 \right\} \right\}.$$

The terms in the inner sum depends each on a different  $\tau_{0,k-1}$ , thus by marginalization (integration w.r.t. others  $\tau_{0,k-1}$ ), we get

$$I_q^{NI} = \sum_{k=1}^N \mathbb{E}_{\tau_{0,k-1}} \left\{ \mathbb{E}_{i_k|\tau_{0,k-1}} \left\{ \left[ \frac{\partial \log \mathbb{P}(i_k|\tau_{0,k-1};x)}{\partial x} \right]^2 \right\} \right\}.$$

Observe that the inner expectation is the FI for each observation  $i_k$  parametrized by  $\tau_{0,k-1}$  and  $x$ . We can re-parametrize it by the difference  $\varepsilon_k = \tau_{0,k-1} - x$ , writing it using the notation of (1.13). Therefore, we obtain

$$I_q^{NI} = \sum_{k=1}^N \mathbb{E}_{\varepsilon_k} \{I_q(\varepsilon_k)\}.$$

### A.1.6 Proof that the posterior PDF can be written in recursive form using prediction and update expressions

For obtaining a relation between the PDF  $p(x_k|i_{1:k-1})$ , that we will call prediction PDF, and the posterior for instant  $k-1$ ,  $p(x_{k-1}|i_{1:k-1})$ , we will use conditioning on the joint density/distribution (PDF for  $X$  and probability for  $i$ ) of the variables  $X_k$ ,  $X_{k-1}$  and  $i_{1:k-1}$

$$p(x_k, x_{k-1}, i_{1:k-1}) = p(x_k|x_{k-1}, i_{1:k-1})p(x_{k-1}|i_{1:k-1})\mathbb{P}(i_{1:k-1}).$$

Exploiting the fact that conditioned on  $X_{k-1}$  the r.v.  $X_k$  is independent of all the past measurements, we have

$$p(x_k, x_{k-1}, i_{1:k-1}) = p(x_k|x_{k-1})p(x_{k-1}|i_{1:k-1})\mathbb{P}(i_{1:k-1}).$$

On the other hand, conditioning only on the measurements, we obtain

$$p(x_k, x_{k-1}, i_{1:k-1}) = p(x_k, x_{k-1}|i_{1:k-1})\mathbb{P}(i_{1:k-1}).$$

Equating the last expressions gives

$$p(x_k, x_{k-1}|i_{1:k-1}) = p(x_k|x_{k-1})p(x_{k-1}|i_{1:k-1}).$$

Marginalization of  $X_{k-1}$  gives the prediction expression

$$p(x_k|i_{1:k-1}) = \int_{\mathbb{R}} p(x_k|x_{k-1})p(x_{k-1}|i_{1:k-1}) dx_{k-1}.$$

As it was stated before, we can notice that for obtaining the prediction PDF we must use the last posterior and the transition PDF  $p(x_k|x_{k-1})$  that characterizes the dynamical model.

For obtaining the update expression, we will start by conditioning the joint density/distribution function  $p(x_k, i_k, i_{1:k-1})$

$$p(x_k, i_k, i_{1:k-1}) = \mathbb{P}(i_k|x_k, i_{1:k-1})p(x_k|i_{1:k-1}).$$

As  $i_k$  given  $x_k$  is independent from all the other r.v., we have

$$p(x_k, i_k, i_{1:k-1}) = \mathbb{P}(i_k|x_k)p(x_k|i_{1:k-1})\mathbb{P}(i_{1:k-1}).$$

Now, conditioning on the entire set of measurements

$$p(x_k, i_k, i_{1:k-1}) = p(x_k|i_{1:k})\mathbb{P}(i_{1:k}).$$

Using both last expressions, we get

$$p(x_k|i_{1:k}) = \frac{\mathbb{P}(i_k|x_k)p(x_k|i_{1:k-1})\mathbb{P}(i_{1:k-1})}{\mathbb{P}(i_{1:k})},$$

this result can be simplified by applying conditioning on the denominator. Absorbing the factor  $\mathbb{P}(i_{1:k-1})$ , we have

$$p(x_k|i_{1:k}) = \frac{\mathbb{P}(i_k|x_k)p(x_k|i_{1:k-1})}{\mathbb{P}(i_k|i_{1:k-1})}.$$

The conditional probability on the denominator can be expressed using marginalization of

$$p(i_k, x_{k-1}|i_{1:k-1}) = \mathbb{P}(i_k|x_k)p(x_k|i_{1:k-1}),$$

which finally gives the update expression

$$p(x_k|i_{1:k}) = \frac{\mathbb{P}(i_k|x_k)p(x_k|i_{1:k-1})}{\int_{\mathbb{R}} \mathbb{P}(i_k|x'_k)p(x'_k|i_{1:k-1}) dx'_k}.$$

Note that for updating the prediction to the posterior distribution, we introduced the information from the measurement through  $\mathbb{P}(i_k|x_k)$ .

### A.1.7 Proof that $I_c$ for the GGD is $\frac{1}{\delta^2} \frac{\beta(\beta-1)\Gamma(1-\frac{1}{\beta})}{\Gamma(\frac{1}{\beta})}$ .

The continuous measurement FI for the GGD distribution is obtained using the PDF expression (1.39) in the integral (3.58)

$$I_{c,GGD} = \int_{\mathbb{R}} \frac{[f_{GGD}^{(1)}(x)]^2}{f_{GGD}(x)} dx = \frac{\beta^3}{\delta^3 \Gamma(\frac{1}{\beta})} \int_0^{+\infty} \left(\frac{x}{\delta}\right)^{2\beta-2} \exp\left[-\left(\frac{x}{\delta}\right)^\beta\right] dx,$$

where we used the fact that the function to be integrated is an even function for obtaining an integral on  $[0, +\infty)$ . We can now change the integration variable to  $z = \left(\frac{x}{\delta}\right)^\beta$ , this produces  $dx = \frac{\delta}{\beta} z^{\frac{1}{\beta}-1} dz$ , leading to the following integral

$$I_{c,GGD} = \frac{\beta^2}{\delta^2 \Gamma(\frac{1}{\beta})} \int_0^{+\infty} z^{1-\frac{1}{\beta}} \exp(-z) dz.$$

The integral is equal to  $\Gamma\left(2 - \frac{1}{\beta}\right)$ , thus using the property of the gamma function  $\Gamma(1+z) = z\Gamma(z)$ , we have finally

$$I_{c,GGD} = \frac{1}{\delta^2} \frac{\beta(\beta-1)\Gamma\left(1 - \frac{1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)}.$$

### A.1.8 Proof that $I_c$ for the STD is $\frac{1}{\delta^2} \frac{\beta+1}{\beta+3}$ .

For the STD, the continuous measurement FI is obtained using its PDF expression (3.72) in (3.58). As the function to be integrated is an even function, we can integrate it only in the positive real semi-axis. This gives

$$\begin{aligned} I_{c,STD} &= \int_{\mathbb{R}} \frac{[f_{STD}^{(1)}(x)]^2}{f_{STD}(x)} dx \\ &= \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\Gamma\left(\frac{\beta}{2}\right)} \frac{1}{\delta^3 \sqrt{\beta} \sqrt{\pi}} \frac{(\beta+1)^2}{\beta} 2 \int_0^{+\infty} \left(\frac{x}{\delta\sqrt{\beta}}\right)^2 \left[1 + \left(\frac{x}{\delta\sqrt{\beta}}\right)^2\right]^{-\frac{\beta+5}{2}} dx. \end{aligned}$$

For evaluating this integral, we can change the integration variable to  $\theta$  using  $\tan(\theta) = \frac{x}{\delta\sqrt{\beta}}$ , this produces  $dx = \sqrt{\beta}\delta \frac{1}{\cos^2(\theta)}$ ,  $1 + \left(\frac{x}{\delta\sqrt{\beta}}\right)^2 = \frac{1}{\cos^2(\theta)}$  and an integration interval  $[0, \frac{\pi}{2})$ , leading to

$$I_{c,STD} = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\Gamma\left(\frac{\beta}{2}\right)} \frac{1}{\delta^2 \sqrt{\pi}} \frac{(\beta+1)^2}{\beta} 2 \int_0^{\frac{\pi}{2}} \sin^2(\theta) \cos(\theta)^{\beta+1} dx.$$

The integral factor multiplied by 2 can be identified to the beta function  $B\left(\frac{3}{2}, \frac{\beta}{2} + 1\right)$ . The beta function can be written using the gamma function

$$B\left(\frac{3}{2}, \frac{\beta}{2} + 1\right) = \frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{\beta}{2} + 1\right)}{\Gamma\left(\frac{\beta+1}{2} + 2\right)},$$

which can be rewritten using the fact that  $\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$  and the property of the gamma function  $\Gamma(1+z) = z\Gamma(z)$ . This gives

$$B\left(\frac{3}{2}, \frac{\beta}{2} + 1\right) = \sqrt{\pi} \frac{\beta}{(\beta+3)(\beta+1)} \frac{\Gamma\left(\frac{\beta}{2}\right)}{\Gamma\left(\frac{\beta+1}{2}\right)},$$

leading finally to

$$I_{c,STD} = \frac{1}{\delta^2} \frac{\beta+1}{\beta+3}.$$

#### A.1.9 Minimization of the asymptotic variance w.r.t. $\boldsymbol{\eta}$ under the asymptotic zero mean constraint.

For simplifying the notation we will use  $\boldsymbol{\eta}$  and  $\mathbf{f}_d$ , suppressing the subscripts and superscripts. The problem we want to solve is

$$\begin{aligned} &\text{minimize} && \sigma_\infty^2 = \frac{\boldsymbol{\eta}^\top \mathbf{F}_d \boldsymbol{\eta}}{\boldsymbol{\eta}^\top \mathbf{f}_d \mathbf{f}_d^\top \boldsymbol{\eta}}, \\ &\text{w.r.t. } \boldsymbol{\eta} && \\ &\text{subject to} && \mathbf{F}_d^{vec, \top} \boldsymbol{\eta} = 0, \\ &&& \boldsymbol{\eta} \neq 0, \end{aligned} \tag{A.3}$$

where  $\mathbf{F}_d^{vec}$  is the diagonal of  $\mathbf{F}_d$  in vector form. This problem can be also cast as a maximization problem

$$\begin{aligned} &\text{maximize} && \frac{1}{\sigma_\infty^2} = \frac{\boldsymbol{\eta}^\top \mathbf{f}_d \mathbf{f}_d^\top \boldsymbol{\eta}}{\boldsymbol{\eta}^\top \mathbf{F}_d \boldsymbol{\eta}}, \\ &\text{w.r.t. } \boldsymbol{\eta} && \\ &\text{subject to} && \mathbf{F}_d^{vec, \top} \boldsymbol{\eta} = 0, \\ &&& \boldsymbol{\eta} \neq 0. \end{aligned} \tag{A.4}$$

As  $\mathbf{F}_d$  is a diagonal matrix it can be decomposed as the product of diagonal matrices formed with the square roots of the diagonal terms

$$\mathbf{F}_d = \mathbf{F}_d^{\frac{1}{2}} \mathbf{F}_d^{\frac{1}{2}}.$$

Thus using the change of variables

$$\boldsymbol{\eta} = \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}',$$

the problem (A.4) becomes

$$\begin{aligned} & \text{maximize} && \frac{\boldsymbol{\eta}'^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}'}{\boldsymbol{\eta}'^\top \boldsymbol{\eta}'} \\ & \text{w.r.t. } \boldsymbol{\eta}' && \\ & \text{subject to} && \mathbf{F}_d^{vec, \top} \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}' = 0, \\ & && \boldsymbol{\eta}' \neq 0. \end{aligned}$$

This problem can be solved by constraining  $\boldsymbol{\eta}'^\top \boldsymbol{\eta}'$  to be equal to one and then maximizing the numerator

$$\begin{aligned} & \text{maximize} && \boldsymbol{\eta}'^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}', \\ & \text{w.r.t. } \boldsymbol{\eta}' && \\ & \text{subject to} && \boldsymbol{\eta}'^\top \boldsymbol{\eta}' = 1, \\ & && \mathbf{F}_d^{vec, \top} \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}' = 0, \\ & && \boldsymbol{\eta}' \neq 0. \end{aligned} \tag{A.5}$$

Note that  $\mathbf{F}_d^{vec, \top} \mathbf{F}_d^{-\frac{1}{2}}$  is a transposed vector with the square roots of  $\mathbf{F}_d^{vec}$ . This term will be denoted  $\left(\mathbf{F}_d^{\frac{1}{2}, vec}\right)^\top$  from now on. This problem has been treated in [Golub 1973] and we will apply here the same development.

The Lagrangian of the maximization problem (A.5) is given by

$$\mathcal{L} = \boldsymbol{\eta}'^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}' - \lambda \left( \boldsymbol{\eta}'^\top \boldsymbol{\eta}' - 1 \right) + 2\mu \boldsymbol{\eta}'^\top \mathbf{F}_d^{\frac{1}{2}, vec},$$

where  $\lambda$  and  $\mu$  are Lagrange multipliers. The zero derivative point of the Lagrangian w.r.t.  $\boldsymbol{\eta}'$  is given as the solution of the following equation:

$$\mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}' - \lambda \boldsymbol{\eta}' + \mu \mathbf{F}_d^{\frac{1}{2}, vec} = 0. \tag{A.6}$$

Multiplying by  $\left(\mathbf{F}_d^{\frac{1}{2}, vec}\right)^\top$  gives

$$\left(\mathbf{F}_d^{\frac{1}{2}, vec}\right)^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}' - \lambda \left(\mathbf{F}_d^{\frac{1}{2}, vec}\right)^\top \boldsymbol{\eta}' + \mu \left(\mathbf{F}_d^{\frac{1}{2}, vec}\right)^\top \mathbf{F}_d^{\frac{1}{2}, vec} = 0.$$

As  $\mathbf{F}_d$  are quantizer output probabilities, we have

$$\left(\mathbf{F}_d^{\frac{1}{2}, vec}\right)^\top \mathbf{F}_d^{\frac{1}{2}, vec} = 1.$$

Now, using the expression above and the second equality constraint (that the asymptotic mean is zero) on the factor that multiplies  $\lambda$ , we obtain

$$\mu = - \left(\mathbf{F}_d^{\frac{1}{2}, vec}\right)^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}'.$$

Substituting this expression for  $\mu$  in (A.6), we get

$$\left[ \mathbf{I} - \mathbf{F}_d^{-\frac{1}{2}} \left( \mathbf{F}_d^{\frac{1}{2}, vec} \right)^\top \right] \left( \mathbf{F}_d^{\frac{1}{2}, vec} \right)^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \boldsymbol{\eta}' = \lambda \boldsymbol{\eta}',$$

where  $\mathbf{I}$  is the identity matrix. Clearly,  $\left[ \mathbf{I} - \mathbf{F}_d^{-\frac{1}{2}} \left( \mathbf{F}_d^{\frac{1}{2}, vec} \right)^\top \right] = \mathbf{P}'$  is a projection matrix and the optimal  $\boldsymbol{\eta}'$  is the eigenvector of  $\mathbf{P}' \left( \mathbf{F}_d^{\frac{1}{2}, vec} \right)^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}}$  that gives the maximum  $\lambda$ . For a squared matrix  $\mathbf{A}$  and projection matrix  $\mathbf{P}'$ , we know that the maximum eigenvalue function  $\lambda(\cdot)$  respects the following equality:

$$\lambda(\mathbf{P}'\mathbf{A}) = \lambda(\mathbf{P}'^2\mathbf{A}) = \lambda(\mathbf{P}'\mathbf{A}\mathbf{P}').$$

This means that the optimal  $\boldsymbol{\eta}'$  can also be found as the eigenvector of

$$\mathbf{P}' \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{P}'$$

related to the maximum eigenvalue  $\lambda$ . As the only non zero eigenvector of  $\mathbf{P}' \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \mathbf{f}_d^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{P}'$  is  $\mathbf{P}' \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d$ , this is the optimal  $\boldsymbol{\eta}'$ . Changing back to the initial vector  $\boldsymbol{\eta}$ , we have

$$\boldsymbol{\eta} \propto \mathbf{F}_d^{-\frac{1}{2}} \left[ \mathbf{I} - \mathbf{F}_d^{-\frac{1}{2}} \left( \mathbf{F}_d^{\frac{1}{2}, vec} \right)^\top \right] \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d.$$

The proportional  $\propto$  comes from the fact that the solution of (A.3) is defined up to a proportional factor. Expanding the expression gives

$$\begin{aligned} \boldsymbol{\eta} &\propto \mathbf{F}_d^{-1} \mathbf{f}_d - \mathbf{F}_d^{-\frac{1}{2}} \mathbf{F}_d^{\frac{1}{2}, vec} \left( \mathbf{F}_d^{\frac{1}{2}, vec} \right)^\top \mathbf{F}_d^{-\frac{1}{2}} \mathbf{f}_d \\ &\propto \mathbf{F}_d^{-1} \mathbf{f}_d - \mathbf{1} \mathbf{f}_d, \end{aligned}$$

where  $\mathbf{1}$  is a squared matrix filled with ones.

**A.1.10 Proof that  $\frac{1}{\mathbf{f}_d^T \mathbf{F}_d^{-1} \mathbf{f}_d} = \frac{1}{\sum_{j=1}^{N_s} \sum_{i(j) \in \mathcal{I}(j)} \frac{\tilde{f}_d^{(j)2} [i(j)]}{\tilde{F}_d^{(j)} [i(j)]}}$  in the fusion center approach.**

For simplifying notation the sensor superscript in  $\tilde{F}_d^{(l')} [i^{(l')}]$  and  $\tilde{f}_d^{(l')} [i^{(l')}]$  will not be written, the dependence on the sensor number will be done implicitly through the argument of the function  $\tilde{F}_d [i^{(l')}]$  and  $\tilde{f}_d [i^{(l')}]$ . Using the fact that that  $\mathbf{F}_d$  is diagonal, we can write

$$\mathbf{f}_d^T \mathbf{F}_d^{-1} \mathbf{f}_d = \sum_{\mathbf{i} \in \mathcal{I}^{\otimes N_s}} \frac{\left( \sum_{j=1}^{N_s} \tilde{f}_d [i^{(j)}] \prod_{\substack{j'=1 \\ j' \neq j}}^{N_s} \tilde{F}_d [i^{(j')}] \right)^2}{\prod_{j=1}^{N_s} \tilde{F}_d [i^{(j)}]},$$

where  $\mathcal{I}^{\otimes N_s}$  is the set of all possible  $\mathbf{i}$ . Developing the quadratic term, the sum above is equal to the sum of two terms

$$\mathbf{f}_d^T \mathbf{F}_d^{-1} \mathbf{f}_d = I_1 + I_2,$$

with

$$I_1 = \sum_{\mathbf{i} \in \mathcal{I}^{\otimes N_s}} \left\{ \frac{\sum_{j=1}^{N_s} \left\{ \tilde{f}_d [i^{(j)}] \right\}^2 \prod_{\substack{j'=1 \\ j' \neq j}}^{N_s} \left\{ \tilde{F}_d [i^{(j')}] \right\}^2}{\prod_{j=1}^{N_s} \tilde{F}_d [i^{(j)}]} \right\}$$

and

$$I_2 = \sum_{\mathbf{i} \in \mathcal{I}^{\otimes N_s}} \left\{ \frac{\sum_{l=1}^{N_s} \sum_{\substack{m=1 \\ m \neq l}}^{N_s} \tilde{f}_d [i^{(l)}] \tilde{f}_d [i^{(m)}] \left\{ \prod_{\substack{l'=1 \\ l' \neq l}}^{N_s} \tilde{F}_d [i^{(l')}] \right\} \left\{ \prod_{\substack{m'=1 \\ m' \neq m}}^{N_s} \tilde{F}_d [i^{(m')}] \right\}}{\prod_{j=1}^{N_s} \tilde{F}_d [i^{(j)}]} \right\}.$$

Dividing the common factors in  $I_2$  and rewriting the sum, we obtain

$$\begin{aligned} I_2 &= \sum_{\mathbf{i} \in \mathcal{I}^{\otimes N_s}} \sum_{l=1}^{N_s} \sum_{\substack{m=1 \\ m \neq l}}^{N_s} \left\{ \tilde{f}_d [i^{(l)}] \tilde{f}_d [i^{(m)}] \prod_{\substack{p=1 \\ p \neq l \\ p \neq m}}^{N_s} \tilde{F}_d [i^{(p)}] \right\} \\ &= \sum_{l=1}^{N_s} \sum_{\substack{m=1 \\ m \neq l}}^{N_s} \sum_{i^{(l)} \in \mathcal{I}^{(l)}} \sum_{i^{(m)} \in \mathcal{I}^{(m)}} \left\{ \tilde{f}_d [i^{(l)}] \tilde{f}_d [i^{(m)}] \left\{ \sum_{\mathbf{i} \in \mathcal{I}^{\otimes N_s^*}} \prod_{\substack{p=1 \\ p \neq l \\ p \neq m}}^{N_s} \tilde{F}_d [i^{(p)}] \right\} \right\}, \end{aligned}$$

where  $\mathcal{I}^{\otimes N_s^*}$  is the set of all combinations of  $\mathbf{i}$ , without considering  $i^{(l)}$  and  $i^{(m)}$ . The interior sum in the RHS of the last equality equals to one because  $\tilde{F}_d [i^{(p)}]$  is a probability. Thus  $I_2$

is given by

$$\begin{aligned}
I_2 &= \sum_{l=1}^{N_s} \sum_{\substack{m=1 \\ m \neq l}}^{N_s} \sum_{i^{(l)} \in \mathcal{I}^{(l)}} \sum_{i^{(m)} \in \mathcal{I}^{(m)}} \left\{ \tilde{f}_d [i^{(l)}] \tilde{f}_d [i^{(m)}] \right\} \\
&= \sum_{l=1}^{N_s} \sum_{\substack{m=1 \\ m \neq l}}^{N_s} \left\{ \sum_{i^{(l)} \in \mathcal{I}^{(l)}} \tilde{f}_d [i^{(l)}] \right\} \left\{ \sum_{i^{(m)} \in \mathcal{I}^{(m)}} \tilde{f}_d [i^{(m)}] \right\}.
\end{aligned}$$

From the symmetry assumptions  $\tilde{f}_d [i^{(j)}]$  is an odd function of  $i^{(j)}$ , therefore  $I_2 = 0$ .

The term  $I_1$  can be rewritten by dividing common factors from the numerator and denominator. This gives

$$I_1 = \sum_{\mathbf{i} \in \mathcal{I}^{\otimes N_s}} \sum_{j=1}^{N_s} \left\{ \frac{\tilde{f}_d^2 [i^{(j)}]}{\tilde{F}_d [i^{(j)}]} \prod_{\substack{j'=1 \\ j' \neq j}}^{N_s} \tilde{F}_d [i^{(j')}] \right\}.$$

Changing the order of summation and separating the sum for the sensor index  $j$  from the others, we obtain

$$I_1 = \sum_{j=1}^{N_s} \sum_{i^{(j)} \in \mathcal{I}^{(j)}} \left\{ \frac{\tilde{f}_d^2 [i^{(j)}]}{\tilde{F}_d [i^{(j)}]} \sum_{\mathbf{i} \in \mathcal{I}^{\otimes N_s^*}} \left\{ \prod_{\substack{j'=1 \\ j' \neq j}}^{N_s} \tilde{F}_d [i^{(j')}] \right\} \right\},$$

where now  $\mathcal{I}^{\otimes N_s^*}$  is the set of all possible  $\mathbf{i}$  without considering  $i^{(j)}$ . As the inner sum is equal to one, we finally have

$$\mathbf{f}_d^T \mathbf{F}_d^{-1} \mathbf{f}_d = \sum_{j=1}^{N_s} \sum_{i^{(j)} \in \mathcal{I}^{(j)}} \frac{\tilde{f}_d^2 [i^{(j)}]}{\tilde{F}_d [i^{(j)}]}$$

and consequently

$$\frac{1}{\mathbf{f}_d^T \mathbf{F}_d^{-1} \mathbf{f}_d} = \frac{1}{\sum_{j=1}^{N_s} \sum_{i^{(j)} \in \mathcal{I}^{(j)}} \frac{\tilde{f}_d^2 [i^{(j)}]}{\tilde{F}_d [i^{(j)}]}}.$$

## A.2 More? - Further details

### A.2.1 Discussion on the issues of finding the MLE

**Binary quantization.** In Subsection 1.3.6, we give an analytic expression for the MLE in the binary quantization case. In this case the MLE depends on the noise distribution mainly through the inverse of the CDF, thus existence and unicity of the MLE are guaranteed by the monotonicity of noise CDF (implicitly stated in the assumption AN2).

**Multibit and dynamic quantization: log-concave distributions.** In the multibit case, or even in the binary case when the threshold is not static, we cannot write a closed-form expression for the the MLE. In this case, we have to use a numerical method for the evaluation of the maximum.

In the case of log-concave distributions (the Gaussian distribution is an example), we can show that, as it is explained in Subsection 1.4.4, the log-likelihood with quantized measurements is concave. Thus, in this case the log-likelihood has only one maximum which can be found very efficiently using the Newton's algorithm.

**Multibit and dynamic quantization: general distributions.** If the distribution is not log-concave, then the Newton's algorithm does not necessarily converge. If it converges, it can converge very slowly when compared to the log-concave distribution case. It can also happen that the likelihood has multiple maxima, in this case, any technique based on the gradient may fail to find the global maxima and other types of maximization techniques must be used.

As a simple example of non log-concave noise distribution, we can consider the Cauchy distribution with PDF and CDF given by (1.33) and (1.34) respectively. The log-likelihood for estimating  $x$  with  $\delta = 1$ ,  $\boldsymbol{\tau} = [-3 \ -2 \ -1 \ 0 \ 1 \ 2 \ 3]^\top$  and  $i_k = \{-3, -3, -4, 3, 3, 3\}$  is shown in Fig. A.2.

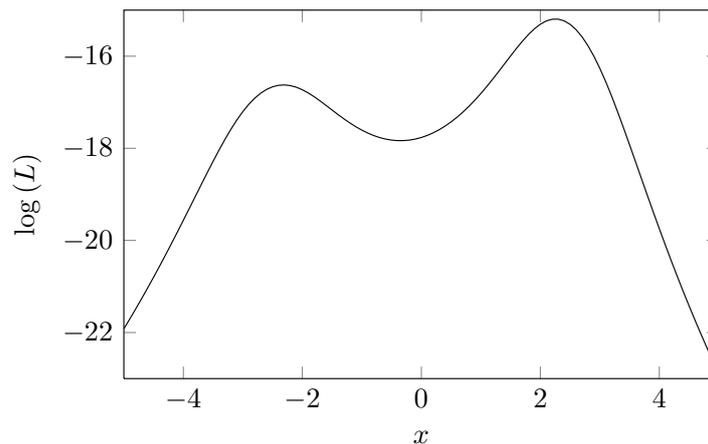


Figure A.2: Log-likelihood function for estimating  $x$  based on the quantized measurements  $i_k = \{-3, -3, -4, 3, 3, 3\}$ . The quantizer has  $N_I = 8$  and  $\boldsymbol{\tau} = [-3 \ -2 \ -1 \ 0 \ 1 \ 2 \ 3]^\top$ . The distribution of the noise is Cauchy with  $\delta = 1$ .

We can clearly note the multimodality of the log-likelihood function.

### A.2.2 MLE for estimation of a constant based on binary quantized measurements: uniform/Gaussian noise case.

The MLE for binary quantized measurements is given by (1.45)

$$\hat{X}_{ML} = \tau_0 - F^{-1} \left[ \frac{1}{2} \left( 1 - \frac{1}{N} \sum_{k=1}^N i_k \right) \right].$$

The function  $F^{-1}(\cdot)$  is the inverse of the noise CDF. For the uniform/Gaussian case, the CDF is given by (1.37)

$$F(\varepsilon) = \begin{cases} \frac{1}{C} \Phi \left( \frac{\varepsilon + \frac{\alpha}{2}}{\sigma} \right), & \text{for } \varepsilon < -\frac{\alpha}{2}, \\ \frac{1}{C} \left[ \frac{1}{2} + \frac{1}{\sqrt{2\pi}\sigma} \left( \varepsilon + \frac{\alpha}{2} \right) \right], & \text{for } -\frac{\alpha}{2} \leq \varepsilon \leq \frac{\alpha}{2}, \\ \frac{1}{C} \left[ \frac{\alpha}{\sqrt{2\pi}\sigma} + \Phi \left( \frac{\varepsilon + \frac{\alpha}{2}}{\sigma} \right) \right], & \text{for } \varepsilon > \frac{\alpha}{2}, \end{cases}$$

where  $C = 1 + \frac{\alpha}{\sqrt{2\pi}\sigma}$ . As the CDF is decomposed in three parts, for inverting the CDF we might distinguish three possible cases. Using the notation  $1 - \hat{P}_{ML} = \frac{1}{2} \left( 1 - \frac{1}{N} \sum_{k=1}^N i_k \right)$ , the cases are the following:

- $1 - \hat{P}_{ML} < \frac{1}{2C}$ ,
- $\frac{1}{2C} \leq 1 - \hat{P}_{ML} \leq \frac{1}{C} \left( \frac{1}{2} + \frac{\alpha}{\sqrt{2\pi}\sigma} \right)$ ,
- $1 - \hat{P}_{ML} > \frac{1}{C} \left( \frac{1}{2} + \frac{\alpha}{\sqrt{2\pi}\sigma} \right)$ .

Using the inverse of  $F(\cdot)$  for each case in the expression of the estimator above, we get

$$\hat{X}_{ML} = \begin{cases} \tau_0 + \frac{\alpha}{2} - \sigma \Phi^{-1} \left[ C \left( 1 - \hat{P}_{ML} \right) \right], & \text{for } 1 - \hat{P}_{ML} < \frac{1}{2C}, \\ \tau_0 + \frac{\alpha}{2} - \sqrt{2\pi}\sigma \left[ C \left( 1 - \hat{P}_{ML} \right) - \frac{1}{2} \right], & \text{for } \frac{1}{2C} \leq 1 - \hat{P}_{ML} \leq \frac{1}{C} \left( \frac{1}{2} + \frac{\alpha}{\sqrt{2\pi}\sigma} \right), \\ \tau_0 - \frac{\alpha}{2} - \sigma \Phi^{-1} \left[ C \left( 1 - \hat{P}_{ML} \right) - \frac{\alpha}{\sqrt{2\pi}\sigma} \right], & \text{for } 1 - \hat{P}_{ML} > \frac{1}{C} \left( \frac{1}{2} + \frac{\alpha}{\sqrt{2\pi}\sigma} \right). \end{cases}$$

The function  $\Phi^{-1}[\cdot]$  is the inverse of the standard Gaussian CDF.

### A.2.3 MLE for estimation of a constant based on binary quantized measurements: generalized Gaussian noise case.

For binary quantized measurements with a fixed threshold, the MLE is given by (1.45)

$$\hat{X}_{ML} = \tau_0 - F^{-1} \left[ \frac{1}{2} \left( 1 - \frac{1}{N} \sum_{k=1}^N i_k \right) \right],$$

where  $F^{-1}(\cdot)$  is the inverse of the noise CDF. In the GGD case the CDF is the following (1.40):

$$F(\varepsilon) = \frac{1}{2} \left[ 1 + \text{sign}(\varepsilon) \frac{\gamma\left(\frac{1}{\beta}, \left|\frac{\varepsilon}{\delta}\right|^\beta\right)}{\Gamma\left(\frac{1}{\beta}\right)} \right].$$

Therefore, denoting the average of the binary observations by  $\bar{i} = \frac{1}{N} \sum_{k=1}^N i_k$ , we have the following MLE:

$$\hat{X}_{ML} = \tau_0 + \delta \text{sign}(\bar{i}) \left\{ \gamma^{-1} \left[ \frac{1}{\beta}, |\bar{i}| \Gamma\left(\frac{1}{\beta}\right) \right] \right\}^{\frac{1}{\beta}},$$

where  $\gamma^{-1}[\cdot, \cdot]$  is the inverse of the incomplete gamma function.

### A.2.4 Adaptive binary threshold asymptotic probabilities when the threshold is defined in a grid.

We consider here that the parameter lies in an interval  $[-A, A]$ , where  $A$  is a positive real. For assimilating this information, we are going to change the update of the binary threshold. The following is assumed:

- The step size  $\gamma$  is chosen so that

$$\frac{A}{\gamma} = N,$$

with  $N$  a positive integer.

- The initial threshold  $\tau_{0,0}$  is chosen to be an integer multiple of  $\gamma$ ,  $\tau_{0,0} = j\gamma$ , so that  $\tau_{0,0} \in [-A, A]$ .
- The threshold cannot leave the interval  $[-A, A]$ . This means that when  $\tau_{0,k-1} = A$  and  $i_k = 1$ , we will set  $\tau_{0,k} = A$ . When  $\tau_{0,k-1} = -A$  and we have  $i_k = -1$ , we will set  $\tau_{0,k} = -A$ . This changes the adaptive update of the threshold (1.49) to

$$\tau_{0,k} = \begin{cases} -A, & \text{if } \tau_{0,k-1} = -A \text{ and } i_k = -1, \\ \tau_{0,k} = \tau_{0,k-1} + \gamma i_k, & \\ A, & \text{if } \tau_{0,k-1} = A \text{ and } i_k = 1. \end{cases} \quad (\text{A.7})$$





Using the expression for  $p_{N-1,\infty}$  above, we get

$$p_{N-2,\infty} = \frac{a_{N-1}a_N}{(1-a_{N-2})(1-a_{N-1})}.$$

Clearly, from the similarity of the equations for the other lines, we can proceed in the same way to obtain

$$p_{N-i,\infty} = \left[ \frac{\prod_{j=0}^{i-1} a_{N-j}}{\prod_{j=1}^i (1-a_{N-j})} \right] p_{N,\infty}, \quad \text{for } i \in \{-N, \dots, N-1\}. \quad (\text{A.9})$$

If we denote

$$c_i = \left[ \frac{\prod_{j=0}^{i-1} a_{N-j}}{\prod_{j=1}^i (1-a_{N-j})} \right],$$

then,  $p_{N,\infty}$  can be found by using the constraint that the vector must sum to one

$$\left( \sum_{i=1}^{2N} p_{N-i,\infty} \right) + p_{N,\infty} = 1.$$

Separating the factor  $p_{N,\infty}$  which appears in all terms (see (A.9)), we get

$$p_{N,\infty} = \frac{1}{1 + \sum_{i=1}^{2N} c_i}. \quad (\text{A.10})$$

Using this and (A.9), we can obtain a general expression for the probabilities

$$p_{N-i,\infty} = \frac{c'_i}{1 + \sum_{i=1}^{2N} c_i}, \quad \text{for } i \in \{0, \dots, 2N\}, \quad (\text{A.11})$$

where

$$c'_i = \begin{cases} 1, & \text{if } k = 0, \\ c_i, & \text{otherwise.} \end{cases}$$

By substituting the  $c'_i$  in (A.11), we get the following expressions for the asymptotic probabilities

$$\begin{aligned} p_{N,\infty} &= \frac{1}{P(x)} \prod_{i=1}^{2N} (1-a_{N-i}), \\ p_{-N,\infty} &= \frac{1}{P(x)} \prod_{i=0}^{2N-1} a_{N-i}, \\ p_{N-i,\infty} &= \frac{1}{P(x)} \left( \prod_{j=0}^{i-1} a_{N-j} \right) \left[ \prod_{j=i+1}^{2N} (1-a_{N-j}) \right], \end{aligned} \quad (\text{A.12})$$

where the normalization factor  $P(x)$  is

$$P(x) = \left[ \prod_{j=1}^{2N} (1 - a_{N-j}) \right] + \left( \prod_{j=0}^{2N-1} a_{N-j} \right) + \sum_{j=1}^{2N-1} \left\{ \left( \prod_{k=0}^{j-1} a_{N-k} \right) \left[ \prod_{l=j+1}^{2N} (1 - a_{N-l}) \right] \right\}. \quad (\text{A.13})$$

With the expressions above for the asymptotic probabilities of the thresholds, it is possible to obtain exact values for the asymptotic FI using (1.62).

### Maximum of the probability distribution

We are going to verify that the asymptotic threshold is indeed around the true parameter. We will analyze the position of the maximum probability threshold and the increasing and decreasing patterns of the asymptotic probabilities. For doing so, we will obtain the expressions for the signs of the differences between neighboring (in threshold position) asymptotic probabilities.

Starting at the negative extremum of the interval, the difference is the following:

$$p_{-N,\infty} - p_{-(N-1),\infty} = \frac{1}{P(x)} \left[ \left( \prod_{i=0}^{2N-1} a_{N-i} \right) - \left( \prod_{j=0}^{2N-2} a_{N-j} \right) (1 - a_{-N}) \right].$$

Making explicit the common factor, we have

$$p_{-N,\infty} - p_{-(N-1),\infty} = \frac{\left( \prod_{i=0}^{2N-1} a_{N-i} \right)}{P(x)} [a_{-(N-1)} - (1 - a_{-N})].$$

The factor  $\frac{\left( \prod_{i=0}^{2N-1} a_{N-i} \right)}{P(x)}$  is positive because  $\prod_{i=0}^{2N-1} a_{N-i}$  is positive, it is a product of probabilities, and  $P(x)$  is also positive, it is a sum of products of probabilities. Therefore, to obtain the sign of  $p_{-N,\infty} - p_{-(N-1),\infty}$  as a function of  $x$ , we need only to analyze the sign of the difference  $a_{-(N-1)} - (1 - a_{-N})$ . Using the expressions for the  $a_i$  terms, we have

$$\text{sign}(p_{-N,\infty} - p_{-(N-1),\infty}) = \text{sign} \left\{ \left[ 1 - F \left( x + A \frac{(N-1)}{N} \right) \right] - F(x + A) \right\}.$$

The difference in the sign on the RHS is the difference between a complementary CDF parametrized by  $x$  and centered on  $-A \frac{(N-1)}{N}$ ,  $1 - F \left( x + A \frac{(N-1)}{N} \right)$ , and a CDF centered on  $-A$ ,  $F(x + A)$ . Using the facts that the complementary CDF is a decreasing function (from one to zero), the CDF is an increasing function (from zero to one) and that CDF is simply a reversed and shifted version of the complementary CDF, we obtain the following conclusions:

- the sign of the probability difference is positive for  $x \in \left[ -A, -A \frac{(N-\frac{1}{2})}{N} \right)$ .

- The sign is negative for  $x > -A \frac{(N-\frac{1}{2})}{N}$ .
- The probability difference is zero when  $x = -A \frac{(N-\frac{1}{2})}{N}$ .

The difference between probabilities for  $i \in \{1, \dots, 2N-1\}$  is

$$P_{(N-i),\infty} - P_{(N-i+1),\infty} = \frac{1}{P(x)} \left\{ \left( \prod_{j=0}^{i-1} a_{N-j} \right) \left[ \prod_{j=i+1}^{2N} (1 - a_{N-j}) \right] - \left( \prod_{j=0}^{i-2} a_{N-j} \right) \left[ \prod_{j=i}^{2N} (1 - a_{N-j}) \right] \right\}$$

and after factorization

$$P_{(N-i),\infty} - P_{(N-i+1),\infty} = \frac{\left( \prod_{j=0}^{i-2} a_{N-j} \right) \left[ \prod_{j=i+1}^{2N} (1 - a_{N-j}) \right]}{P(x)} [a_{N-i+1} - (1 - a_{N-i})].$$

As the first factor on the RHS is positive, the sign of the difference is determined by

$$\text{sign}(p_{(N-i),\infty} - p_{(N-i+1),\infty}) = \text{sign} \left\{ \left[ 1 - F \left( x - A \frac{(N-i+1)}{N} \right) \right] - F \left( x - A \frac{(N-i)}{N} \right) \right\}.$$

The analysis of the sign above is similar to the negative extremum case. Thus we have the following conclusions:

- the sign of the difference is positive for  $x < A \frac{(N-i+\frac{1}{2})}{N}$ .
- We have a negative sign for  $x > A \frac{(N-i+\frac{1}{2})}{N}$ .
- The difference is zero when  $x = A \frac{(N-i+\frac{1}{2})}{N}$ .

Using a similar procedure for the positive extremum, we have that the sign of the difference is given by

$$\text{sign}(p_{(N-1),\infty} - p_{N,\infty}) = \text{sign} \left\{ [1 - F(x - A)] - F \left( x - A \frac{(N-1)}{N} \right) \right\},$$

which leads to the same conclusions as above, the exception is that  $i = 0$  in this case.

Joining all the results, we can see that the maximum of the asymptotic probability vector always occurs at the point of the grid that is closest to  $x$ . Moreover, the distribution always decreases when we consider thresholds with increasing distance to the maximum probability threshold. This means that the distribution is unimodal with its maximum close to the parameter, thus justifying the statement that the thresholds will be placed asymptotically around the parameter.

**Small noise approximation**

The analytical asymptotic probabilities expressions (A.13) are quite cumbersome to be evaluated when  $N$  is large. As the CDF are almost step functions (zero/one functions) for large arguments and as the asymptotic probabilities are products of CDF, in the case when the noise level is small compared with  $\gamma$ , we can obtain very simple approximate expressions for the asymptotic probabilities that involve only a few CDF terms.

The small noise approximations for the complementary CDF and CDF are the following:

$$\begin{aligned}
 a_{N-i} = 1 - F\left(x - A\frac{(N-i)}{N}\right) &= \begin{cases} 1, & x < A\frac{(N-i-1)}{N}, \\ 0, & x > A\frac{(N-i+1)}{N}, \\ 1 - F\left(x - A\frac{(N-i)}{N}\right), & A\frac{(N-i+1)}{N} < x < A\frac{(N-i-1)}{N}, \end{cases} \\
 1 - a_{N-i} = F\left(x - A\frac{(N-i)}{N}\right) &= \begin{cases} 0, & x < A\frac{(N-i-1)}{N}, \\ 1, & x > A\frac{(N-i+1)}{N}, \\ F\left(x - A\frac{(N-i)}{N}\right), & A\frac{(N-i+1)}{N} < x < A\frac{(N-i-1)}{N}. \end{cases} \quad (\text{A.14})
 \end{aligned}$$

Independently of the value of  $x$ , we can get the following approximations of the CDF products using (A.14):

$$\begin{aligned}
 \prod_{i=1}^{2N} (1 - a_{N-j}) &\approx 1 - a_{N-1}, \\
 \prod_{i=0}^{2N-1} a_{N-i} &\approx a_{-N+1}, \\
 \left(\prod_{i=0}^{j-1} a_{N-i}\right) \left[\prod_{i=j+1}^{2N} (1 - a_{N-i})\right] &\approx a_{N-j+1} (1 - a_{N-j-1}).
 \end{aligned}$$

We can now apply these approximations to the asymptotic probabilities (A.13). Note that the approximations will be dependent on the value of  $x$ . For  $x \in \left[-A, -A\frac{(N-1)}{N}\right]$  we have

$$\begin{aligned}
 p_{-N,\infty} &\approx \frac{a_{-N+1}}{2 - a_{-N}} = \frac{1 - F\left(x + A\frac{(N-1)}{N}\right)}{1 + F(x + A)}, \\
 p_{-(N-1),\infty} &\approx \frac{1 - a_{-N}}{2 - a_{-N}} = \frac{F(x + A)}{1 + F(x + A)}, \\
 p_{-(N-2),\infty} &\approx \frac{a_{-N+1}}{2 - a_{-N}} = \frac{F\left(x + A\frac{(N-1)}{N}\right)}{1 + F(x + A)} \\
 p_{(N-i),\infty} &\approx 0, \quad \text{for } i \in \{0, \dots, 2N - 3\}.
 \end{aligned}$$

For  $x \in \left( A \frac{(N-i-1)}{N}, A \frac{(N-i)}{N} \right]$ , we obtain 4 nonzero terms, all the others are approximately zero. The nonzero terms are

$$\begin{aligned} p_{(N-i-2),\infty} &\approx \frac{a_{N-i-1}}{2} = \frac{1 - F\left(x - A \frac{(N-i-1)}{N}\right)}{2}, \\ p_{(N-i-1),\infty} &\approx \frac{a_{N-i}}{2} = \frac{1 - F\left(x - A \frac{(N-i)}{N}\right)}{2}, \\ p_{(N-i),\infty} &\approx \frac{a_{N-i+1}}{2} = \frac{F\left(x - A \frac{(N-i+1)}{N}\right)}{2}, \\ p_{(N-i+1),\infty} &\approx \frac{a_{N-i+2}}{2} = \frac{F\left(x - A \frac{(N-i)}{N}\right)}{2}. \end{aligned}$$

Finally, for the positive extremum,  $x \in \left( -A \frac{(N-1)}{N}, -A \right]$ , the approximations give the following:

$$\begin{aligned} p_{(N-2),\infty} &\approx \frac{a_{N-1}}{1 - a_N} = \frac{1 - F\left(x - A \frac{(N-1)}{N}\right)}{2 - F(x - A)}, \\ p_{(N-1),\infty} &\approx \frac{a_N}{1 + a_N} = \frac{1 - F(x - A)}{2 - F(x - A)}, \\ p_{N,\infty} &\approx \frac{1 - a_{N-1}}{1 + a_N} = \frac{F\left(x - A \frac{(N-1)}{N}\right)}{2 - F(x - A)}, \\ p_{(N-i),\infty} &\approx 0, \quad \text{for } i \in \{3, \dots, 2N\}. \end{aligned}$$

Under the small noise assumption, these approximations are not only useful for evaluating the FI, but they can also be used for estimating the parameter when the number of measurements is very large. Suppose that after a number of samples  $M$ , the threshold probabilities reach approximately the asymptotic distribution, then, from this point on, we start to store the measurements forming an histogram of the threshold values that were used. After a large number of measurements, the histogram will be very close to the asymptotic threshold probabilities. We can then search for the two largest values of the histogram and using one of the correspondent empirical frequencies in the place of the true probability, we can inverse the corresponding approximate expression for the probability to obtain  $x$ .

For example, suppose we have obtained the largest empirical frequencies at the points  $N - i - 1$  and  $N - i$ . The empirical frequency at  $N - i - 1$  is  $\hat{p}_{(N-i-1),\infty}$ , then inverting the corresponding  $p_{(N-i-1),\infty}$  we get the estimate  $\hat{x}$

$$\hat{x} = A \frac{(N-1)}{N} + F^{-1}\left(2\hat{p}_{(N-i-1),\infty} - 1\right).$$

### A.2.5 Particle filter using rejection sampling for tracking a scalar Wiener process.

The optimal sampling distribution  $p(x_k|x_{k-1}, i_k)$  can be rewritten as

$$p(x_k|x_{k-1}, i_k) = \frac{p(x_k, x_{k-1}, i_k)}{p(x_{k-1}, i_k)} = \frac{\mathbb{P}(i_k|x_k) p(x_k|x_{k-1})}{\mathbb{P}(i_k|x_{k-1})} \propto \mathbb{P}(i_k|x_k) p(x_k|x_{k-1}), \quad (\text{A.15})$$

where the proportional relation comes from the fact that, for a given  $i_k$ , the probability  $\mathbb{P}(i_k|x_{k-1})$  is a constant independent of  $x_k$ . Note that as  $\mathbb{P}(i_k|x_k)$  is a probability, it can be bounded above by one, as a consequence  $\mathbb{P}(i_k|x_k) p(x_k|x_{k-1})$  can be bounded above by  $p(x_k|x_{k-1})$  which is a Gaussian PDF. Therefore, for each previous  $x_{k-1}^{(j)}$ , a standard rejection sampling method [Robert 1999, pp. 50] can be applied to generate a sample from  $p(x_k|x_{k-1}^{(j)}, i_k)$ . This can be done by sampling independently from the Gaussian distribution  $p(x_k|x_{k-1})$  and from the uniform distribution  $\mathcal{U}[0, 1]$ . The rejection sampling method that gives the optimal samples  $x_k^{(j)}$  is the following:

#### Rejection sampling for the optimal sampling distribution

**(app1) For  $j = 1$  to  $N_S$**

- **Set  $u_k^{(j)} = 1$  and  $l_k^{(j)} = 0$ .**
- **While  $l_k^{(j)} < u_k^{(j)}$ , do**
  - **Sample the Gaussian distribution**  
**(How? - App. A.3.3)**

$$p(x_k|x_{k-1}^{(j)}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp \left[ -\frac{1}{2} \left( \frac{x_k - x_{k-1}^{(j)} - u_k}{\sigma_w} \right)^2 \right].$$

- **Evaluate  $l_k^{(j)}$**   
$$l_k^{(j)} = \mathbb{P}(i_k|x_k^{(j)}).$$
- **Sample, independently from  $x_k^{(j)}$ , the uniform distribution  $\mathcal{U}[0, 1]$ .**

Note that we accept a sample  $x_k^{(j)}$  only when the its likelihood  $\mathbb{P}(i_k|x_k^{(j)})$  is larger than the uniform sample.

By replacing (A.15) in the place of  $q(x_k|x_{0:k-1}, i_{1:k})$  in the recursive expression for the weights (2.25), we have the following update equation for the weights

$$w(x_{1:k}^{(j)}) = \mathbb{P}(i_k|x_{k-1}^{(j)}) \tilde{w}(x_{1:k-1}^{(j)}).$$

Observe that we might evaluate  $\mathbb{P}(i_k|x_{k-1}^{(j)})$ , which can be obtained similarly as  $\mathbb{P}(i_k|x_k^{(j)})$  with

$$\mathbb{P}(i_k|x_{k-1}^{(j)}) = \begin{cases} F'(\tau_{i_k,k} - x_{k-1}^{(j)} - u_k) - F'(\tau_{i_k-1,k} - x_{k-1}^{(j)} - u_k), & \text{if } i_k > 0, \\ F'(\tau_{i_k+1,k} - x_{k-1}^{(j)} - u_k) - F'(\tau_{i_k,k} - x_{k-1}^{(j)} - u_k), & \text{if } i_k < 0, \end{cases} \quad (\text{A.16})$$

where  $F'$  is the CDF for the r.v. that is the sum of the noise r.v.  $V_k$  and the centered  $X_k$  increment  $W_k - u_k$ .

The procedure for tracking the Wiener process starts by sampling independently  $N_S$  times the prior distribution  $p(x_0)$  and setting the initial weights all to  $\frac{1}{N_S}$ . After obtaining the first measurement  $i_1$ , both the sampling with  $p(x_1|x_{k-1}, i_1)$  and the updates of the weights can be done. Then, after normalizing the weights, the estimate  $\hat{x}_1$  can be obtained with the weighted mean. The procedure is then repeated for each time  $k$  in a sequential way.

This procedure may also suffer from the degeneracy problem explained in Sec. 2.3.4 (p. 85), thus a resampling step (How? - App. A.3.4) (app4) must be carried out each time the number of effective samples is too low.

The performance of this sequential importance sampling algorithm can be obtained through a lower bound, as it is discussed in Sec. 2.4.

**Remark:** to reduce the complexity of this algorithm, we could use a technique based on local linearizations of the optimal proposal distribution [Doucet 2000]. The problem with this approach is that it requires the logarithm of the optimal proposal to have a positive second derivative and this cannot be guaranteed for all noise distributions considered here.

The sequential procedure with the resampling step (particle filter) for solving (b) (p. 29) is the following:

**Solution to (b) - Particle filter with rejection for  
a fixed threshold set sequence  $\tau_{1:k}$**

**(b1.2) 1) Estimator**

- Set uniform normalized weights  $\tilde{w}(x_0^{(j)}) = \frac{1}{N_S}$  and initialize  $N_S$  particles  $\{x_0^{(1)}, \dots, x_0^{(N_S)}\}$  by sampling the prior

$$p(x_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{x_0 - x'_0}{\sigma_0}\right)^2\right].$$

For each time  $k$ ,

- for  $j$  from 1 to  $N_S$ , sample the r.v.  $X_k^{(j)}$  with rejection sampling (app1).
- for  $j$  from 1 to  $N_S$ , evaluate and normalize the weights

$$w(x_{1:k}^{(j)}) = \mathbb{P}(i_k | x_{k-1}^{(j)}) \tilde{w}(x_{1:k-1}^{(j)}), \quad \tilde{w}(x_{1:k}^{(j)}) = \frac{w(x_{1:k}^{(j)})}{\sum_{j=1}^{N_S} w(x_{1:k}^{(j)})},$$

where  $\mathbb{P}(i_k | x_{k-1}^{(j)})$  is given by (A.16).

- Obtain the estimate with the weighted mean

$$\hat{x}_k \approx \sum_{j=1}^{N_S} x_k^{(j)} \tilde{w}(x_{1:k}^{(j)}).$$

- Evaluate the number of effective particles

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^{N_S} \tilde{w}^2(x_{1:k}^{(j)})},$$

if  $N_{\text{eff}} < N_{\text{thresh}}$ , then resample using multinomial resampling (How? - App. A.3.4) (app4).

**2) Performance (lower bound)**

The MSE can be lower bounded as follows

$$\text{MSE}_k \geq \frac{1}{J'_k},$$

with  $J'_k$  given recursively by

$$J'_k = \frac{1}{\sigma_w^2} + I_q(0) - \frac{1}{\sigma_w^A} \frac{1}{\left(\frac{1}{\sigma_w^2} + J'_{k-1}\right)}.$$

### A.3 How? - Algorithms and implementation issues

#### A.3.1 How to sample from a uniform/Gaussian distribution.

We are going to consider that we can generate easily and independently uniform and Gaussian variates. For generating uniform variates, one can use linear congruential generators (see [Knuth 1997, Sec. 3.2] for details), while for generating Gaussian variates one can use the Box-Muller transform which requires a pair of independent uniform variates [Box 1958].

By looking to the specific form of the PDF (1.36)

$$f(\varepsilon) = \begin{cases} f_{GL}(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon+\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{for } \varepsilon < -\frac{\alpha}{2}, \\ f_U(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma}, & \text{for } -\frac{\alpha}{2} \leq \varepsilon \leq \frac{\alpha}{2}, \\ f_{GR}(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon-\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{for } \varepsilon > \frac{\alpha}{2}, \end{cases}$$

where  $C = 1 + \frac{\alpha}{\sqrt{2\pi}\sigma}$ , we can see that we can generate samples from it by generating samples independently from the half Gaussian distributions

$$f'_{GL}(\varepsilon) = \begin{cases} \frac{2}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon+\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{for } \varepsilon < -\frac{\alpha}{2}, \\ 0, & \text{otherwise,} \end{cases}$$

$$f'_{GR}(\varepsilon) = \begin{cases} \frac{2}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon-\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{for } \varepsilon > \frac{\alpha}{2}, \\ 0, & \text{otherwise} \end{cases}$$

and from the central uniform distribution

$$f'_U(\varepsilon) = \begin{cases} \frac{1}{\alpha}, & \text{for } -\frac{\alpha}{2} \leq \varepsilon \leq \frac{\alpha}{2}, \\ 0, & \text{otherwise} \end{cases}$$

and then choosing one of the samples randomly. For having the samples distributed correctly, we will choose the sample from the left Gaussian r.v. with probability  $\frac{1}{2C}$ , or the sample from the uniform distribution with probability  $\frac{\alpha}{\sqrt{2\pi}\sigma C}$  or from the right Gaussian also with probability  $\frac{1}{2C}$ .

This gives the following algorithm for generating a sample from the uniform/Gaussian distribution with parameters  $\alpha$  and  $\sigma$ :

**Uniform/Gaussian sample generator**

**(app2) To generate a sample  $v$  do the following**

- **evaluate**

$$C = \frac{1}{1 + \frac{\alpha}{\sqrt{2\pi\sigma}}},$$

$$p_1 = \frac{1}{2C},$$

$$p_2 = \frac{1}{C} \left( \frac{1}{2} + \frac{\alpha}{\sqrt{2\pi\sigma}} \right).$$

- **Generate 2 independent uniform variates (from  $\mathcal{U}[0, 1]$ )  $u_0$  and  $u_1$  and 2 standard (zero mean and  $\sigma = 1$ ) Gaussian variates  $g_1$  and  $g_2$ .**
- **If  $u_0 < p_1$ , then**

$$v = - \left( \sigma |g_1| + \frac{\alpha}{2} \right),$$

**else if  $p_1 \leq u_0 \leq p_2$ , then**

$$v = \alpha \left( u_1 - \frac{1}{2} \right),$$

**else**

$$v = \sigma |g_2| + \frac{\alpha}{2}.$$

### A.3.2 How to sample from a GGD.

We consider that an easy method for generating independent binary samples (samples with values  $-1$  or  $1$  that have equal probability) and gamma samples is available. For obtaining binary samples one can simply take the sign of a sample from a uniform  $\mathcal{U}[-0.5, 0.5]$  distribution and for obtaining gamma variates one can use a rejection method [Marsaglia 2000].

It can be shown that a generalized Gaussian r.v.  $V'$  with shape parameter  $\beta$  and unit scale parameter can be obtained with the following transformation of two independent r.v. [Nardon 2009]:

$$V' = B \left( \Gamma_{\frac{1}{\beta}} \right)^{\frac{1}{\beta}},$$

where  $B$  is a binary r.v. and  $\Gamma_{\frac{1}{\beta}}$  is a gamma r.v. with shape parameter  $\frac{1}{\beta}$ . If we want a generalized Gaussian r.v.  $V$  with scale parameter  $\delta$ , we need only to multiply  $V'$  by  $\delta$ .

This gives the following algorithm for generating a sample from a GGD with parameters  $\beta$  and  $\delta$ :

**Generalized Gaussian sample generator**

(app3) To generate a sample  $v$  do the following

- generate independently a uniform sample  $u$  from  $\mathcal{U}[0, 1]$  and a gamma sample  $\gamma_{\frac{1}{\beta}}$  from  $\Gamma_{\frac{1}{\beta}}$  with unitary scale parameter.

- Transform the uniform sample  $u$  into a binary sample  $b$  with

$$b = \text{sign} \left( u - \frac{1}{2} \right).$$

- Apply the transformation

$$v = \delta b \left( \gamma_{\frac{1}{\beta}} \right)^{\frac{1}{\beta}}.$$

**A.3.3 How to sample from the distribution  $p(x_k | x_{k-1}^{(j)})$  using a Gaussian standard variate.**

Suppose we can generate a Gaussian standard variate  $W_n \sim \mathcal{N}(0, 1)$ , for example using the Box-Muller transform on a pair of independent uniform variates [Box 1958]. We want to generate a Gaussian variate with PDF

$$p(x_k | x_{k-1}^{(j)}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp \left[ -\frac{1}{2} \left( \frac{x_k - x_{k-1}^{(j)} - u_k}{\sigma_w} \right)^2 \right],$$

where  $x_{k-1}^{(j)}$ ,  $u_k$  and  $\sigma_w$  are known. Using the following properties of Gaussian r.v.:

- the product of a Gaussian r.v. by a constant gives a Gaussian r.v. with variance given by the initial variance multiplied by the square of the constant;
- the sum of a Gaussian r.v. and a constant gives a Gaussian r.v. with mean shifted by the value of the constant.

We have that the r.v.  $X_k^{(j)}$  distributed according to  $p(x_k | x_{k-1}^{(j)})$  can be generated as follows

$$X_k^{(j)} = \sigma_w W_n + x_{k-1}^{(j)} + u_k.$$

**A.3.4 Multinomial resampling algorithm.**

In order to sample from

$$\mathbb{P}(x_k) = \begin{cases} \tilde{w}(x_{1:k}^{(j)}), & \text{if } x_k = x_k^{(j)} \\ 0, & \text{otherwise,} \end{cases}$$

we can create an increasing sequence of cumulative weights

$$w_+^{(j)} = \sum_{i=1}^j \tilde{w} \left( x_{1:k}^{(i)} \right),$$

where we define  $\tilde{w} \left( x_{1:k}^{(0)} \right) = 0$ . Thus, the intervals defined by the neighboring pairs of the sequence  $\left( w_+^{(j-1)}, w_+^{(j)} \right]$  form a partition of the interval  $[0, 1]$  and their lengths equal the corresponding  $\tilde{w} \left( x_{1:k}^{(j)} \right)$ . If we sample from the uniform distribution defined on  $[0, 1]$ ,  $\mathcal{U}[0, 1]$ , and choose  $x_k^{(j)}$  with  $j$  corresponding to the interval in the sequence  $w_+$  in which the uniform sample is contained, then the chosen  $x_k^{(j)}$  are distributed according to the probability distribution  $\mathbb{P}(x_k)$  above.

Resetting equal sample weights at the end of the procedure, we have the multinomial resampling algorithm:

#### Multinomial resampling

**(app4) For  $j = 1$  to  $N_S$**

- **store the particle values in a sequence of auxiliary variables  $\tilde{x}_k^{(j)}$**

$$\tilde{x}_k^{(j)} = x_k^{(j)},$$

- **create the sequence of cumulative weights**

$$w_+^{(j)} = \sum_{i=1}^j \tilde{w} \left( x_{1:k}^{(i)} \right),$$

**with  $\tilde{w} \left( x_{1:k}^{(0)} \right) = 0$ .**

- **Create a sequence  $\{u'_1, \dots, u'_{N_S}\}$  by sampling independently  $N_S$  times from the distribution  $\mathcal{U}[0, 1]$ .**

**For  $j = 1$  to  $N_S$ ,**

- **set  $x_k^{(j)} = \tilde{x}_k^{(l_j)}$ , where  $l_j$  is chosen so that**

$$u_j \in \left( w_+^{(l_j-1)}, w_+^{(l_j)} \right],$$

- **reset the normalized weights to a uniform distribution**

$$\tilde{w} \left( x_{1:k}^{(j)} \right) = \frac{1}{N_S}.$$

### A.3.5 How to sample from a STD.

For generating samples from the STD, we consider that a simple method for generating uniform  $\mathcal{U}[0, 1]$  samples is available.

It is possible to show that a Student's-t r.v.  $V'$  with shape parameter  $\beta$  and unit scale parameter can be obtained with the following transformation of two independent r.v. [Bailey 1994]:

$$V' = \left[ \beta \left( U_1^{-\frac{2}{\beta}} - 1 \right) \right]^{\frac{1}{2}} \cos(2\pi U_2),$$

where  $U_1$  and  $U_2$  are independent r.v. with uniform  $\mathcal{U}[0, 1]$  distribution. If we want a Student's-t r.v.  $V$  with scale parameter  $\delta$ , we need only to multiply  $V'$  by  $\delta$ .

Thus we have the following algorithm for generating a sample from a STD with parameters  $\beta$  and  $\delta$ :

#### Student's-t sample generator

**(app5) To generate a sample  $v$  do the following**

- **generate independently two uniform samples  $u_1$  and  $u_2$  from  $\mathcal{U}[0, 1]$ .**
- **Apply the transformation**

$$v = \delta \left[ \beta \left( u_1^{-\frac{2}{\beta}} - 1 \right) \right]^{\frac{1}{2}} \cos(2\pi u_2).$$

# Résumé détaillé en français (extended abstract in French)

---

Ceci est un résumé détaillé en français des travaux réalisés dans cette thèse. L'introduction et les conclusions des travaux sont traduites directement du manuscrit en anglais pour une meilleure compréhension du contexte, les chapitres concernant les développements et résultats théoriques seront présentés sous forme synthétique, avec seulement les principaux développements et résultats.

## Contents

---

<b>B.1</b>	<b>Introduction</b>	<b>254</b>
<b>B.2</b>	<b>Estimation et quantification : algorithmes et performances</b>	<b>261</b>
B.2.1	Estimation d'un paramètre constant	261
B.2.2	Estimation d'un paramètre variable	270
B.2.3	Quantifieurs adaptatifs pour l'estimation	274
<b>B.3</b>	<b>Estimation et quantification : approximations à haute résolution</b>	<b>286</b>
B.3.1	Approximation à haute résolution de l'information de Fisher	286
<b>B.4</b>	<b>Conclusions</b>	<b>293</b>
B.4.1	Conclusions principales	293
B.4.2	Perspectives	294

---

## B.1 Introduction

### Quantification : une inconnue dans la salle

Ouvrez un livre, un livre quelconque sur les fondements du traitement numérique du signal, et comptez le nombre de pages dédiées au théorème de l'échantillonnage et au traitement du signal à temps discret : la transformée de Fourier rapide, la transformée en  $Z$ , le filtrage à réponse impulsionnelle finie et infinie. Maintenant, comptez le nombre de pages dédiées à la quantification. Même si la moitié du « monde numérique » est un résultat de la quantification, si on lit quelques livres fondamentaux en traitement numérique du signal, on a l'impression qu'elle est un sujet sans importance.

Toutefois, une personne curieuse peut se demander : la quantification est-elle un sujet vraiment dépourvu d'importance ? Peut-être qu'elle est si difficile à étudier et à expliquer de façon simple, que la plupart des références de base en traitement numérique du signal préfèrent omettre une explication plus détaillée. Nous croyons que cette explication est à l'origine de l'omniprésence de la quantification fine des signaux dans la plupart des livres sur le traitement numérique des signaux. En la considérant fine, les auteurs de ces livres peuvent reléguer la quantification à une note de bas de page. On constate que la quantification semble être l'étrange participant de la « fête du traitement numérique des signaux » et que personne ne veut discuter avec elle (même si elle est un des organisateurs de la fête). Quelques domaines du traitement du signal trouvent utile (et dans certaines circonstances ils n'ont pas tort) de refuser tout contact avec la quantification. Chaque fois qu'ils ont besoin de traiter des problèmes induits par la quantification, ils l'appellent de façon dépréciative – « le bruit de quantification ».

Dans cette thèse, nous espérons faire « discuter » de façon respectueuse, sans termes amoindrisants, un des participants de la fête du traitement du signal avec la quantification. Le sujet que nous avons choisi est l'estimation.

Dans la suite, on expliquera la motivation et les points principaux de cette « discussion ».

### Quantification et réseaux de capteurs : l'invitée d'honneur

Bien que nous ne traitons pas explicitement de la conception d'algorithmes d'estimation avec une architecture du type réseau de capteurs, avec cette thèse nous espérons contribuer au développement de techniques qui peuvent être utilisées ou étendues aux réseaux de capteurs.

**L'essor des réseaux de capteurs.** Avec la réduction des coûts et de la taille des dispositifs électroniques, tels que les capteurs et les émetteurs-récepteurs, un nouveau domaine a émergé sous le nom de « Réseaux de capteurs ». Ce terme, en général, désigne un groupe de capteurs capables de communiquer et de traiter des données pour réaliser une tâche donnée, *e.g.* : faire de l'estimation, de la détection, du suivi d'un signal, de la classification, etc.

Les réseaux de capteurs sont intéressants en pratique pour plusieurs raisons, parmi les plus mentionnées dans la littérature on peut trouver [Akyildiz 2002], [Intanagonwiwat 2000],

[Zhao 2004, pp. 7–8] :

- *tolérance aux défaillances et flexibilité.*
- *Déploiement facile.*
- *Possibilité d'utilisation en environnement dangereux.*
- *Possibilité d'utilisation sans maintenance.*
- *Utilisation de la communication pour réduire la quantité d'énergie utilisée.*
- *Rapport signal à bruit amélioré pour le suivi et détection d'événements dans une zone donnée.*

**Applications des réseaux de capteurs.** Les avantages cités plus haut ouvrent la voie pour l'utilisation des réseaux de capteurs dans un très large spectre de domaines [Arampatzis 2005], [Chong 2003], [Durisic 2012], [Puccinelli 2005]: *surveillance de l'environnement, surveillance pour l'agriculture, génie civil, surveillance urbaine, applications en santé, applications commerciales et applications militaires.*

**Le besoin de quantifier.** Même si le progrès des technologies de conception des capteurs et des dispositifs de communication nous amène à l'utilisation de réseaux à grand nombre de capteurs, des considérations pratiques tels que l'utilisation de batteries et des contraintes sur la taille maximale des capteurs imposent trois contraintes majeures pour la conception d'un réseau de capteur : la contrainte énergétique, la contrainte sur le débit de communication et la contrainte sur la complexité.

Pour respecter ces contraintes, on peut quantifier les mesures au niveau des capteurs. Ceci permet de :

- réduire la complexité des opérations grâce à des recherches dans des tableaux pré-stockés et limiter la quantité de mémoire utilisée.
- réduire directement le débit binaire en sortie des capteurs par le réglage du nombre d'intervalles de quantification.
- réduire la quantité d'énergie utilisée, comme conséquence de la réduction de la complexité et du débit.

Voilà les principales raisons pour lesquelles nous avons choisi d'étudier la quantification dans cette thèse.

## Différents objectifs et précisions sur le sujet de la thèse

Dans un réseau de capteurs, on s'intéresse principalement à l'inférence d'une certaine information enfouie dans les mesures. Les deux classes principales de problèmes d'inférence étudiées en traitement du signal sont la détection et l'estimation. Si on regarde la littérature sur les problèmes conjoints détection/quantification et estimation/quantification, on constate que, en comparaison avec la littérature pour les problèmes isolés (seulement détection ou seulement quantification), sa taille n'est pas importante, en revanche, comme conséquence de l'essor des réseaux de capteurs, elle ne cesse pas de grandir.

Quelques références sur ces problèmes conjoints sont :

- *Détection*: [Benitz 1989], [Gupta 2003], [Kassam 1977], [Longo 1990], [Picinbono 1988], [Poor 1977], [Poor 1988], [Tsitsiklis 1993], [Villard 2010], [Villard 2011].
- *Estimation*: [Aysal 2008], [Fang 2008], [Gubner 1993], [Luo 2005], [Marano 2007], [Papadopoulos 2001], [Poor 1988], [Ribeiro 2006a], [Ribeiro 2006b], [Ribeiro 2006c], [Wang 2010].

**Estimation à partir de mesures quantifiées.** Dans cette thèse on s'intéresse au second problème, l'estimation à partir de mesures quantifiées. On commence par la définition générale du problème d'estimation dans un réseau de capteurs pour, après une suite de simplifications, arriver au sujet précis de la thèse.

Dans le schéma général, chaque capteur : mesure une quantité à amplitude continue  $X^{(i)}$ , puis la mesure est traitée et transmise au point où l'estimation sera faite. Ce point peut être un centre de fusion, un des capteurs ou tous les capteurs. Dans le dernier cas, tous les capteurs diffuseront leurs mesures après traitement. Ce schéma est montré en Fig. B.1. La quantité mesurée peut être une suite de vecteurs, une suite de scalaires, un vecteur constant ou un scalaire constant.

Comme première hypothèse de travail, on considère que seulement un des terminaux (capteurs) est utilisé dans le réseau de capteurs, éventuellement on peut considérer plusieurs terminaux, mais dans ce cas la quantité à estimer sera la même pour tous les capteurs. On considère aussi que la quantité à estimer est une séquence de scalaires ou un seul scalaire, on utilise la notation  $X_k$  pour cette quantité dans les deux cas, l'indice  $k$  désigne l'échantillon en question ou le temps discret. Dans le cas où  $X_k$  est une constante scalaire, on a  $X_k = x$ . Le problème simplifié, qui peut être appelé problème d'estimation scalaire à distance, est montré en Fig. B.2.

Le paramètre  $X_k$  est mesuré avec du bruit additif  $V_k$ . La mesure à amplitude continue est notée  $Y_k = X_k + V_k$ . Le problème que nous traitons dans cette thèse est donc un problème d'estimation d'un paramètre de centrage.

En raison des contraintes de conception discutées plus haut, le bloc de traitement est remplacé par un quantifieur scalaire. Par conséquent, chaque mesure continue  $Y_k$  génère une mesure quantifiée  $i_k$  au travers d'une fonction de quantification  $Q(\cdot)$ . Chaque mesure quantifiée est définie dans un ensemble fini de valeurs, ceci permet de fixer le débit binaire en

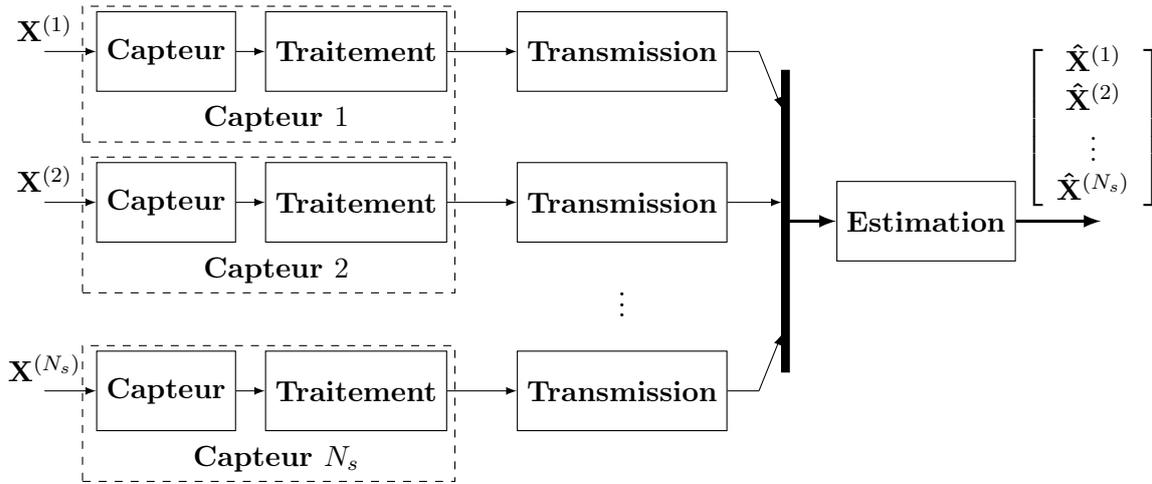


Figure B.1: Estimation avec un réseau de capteurs. Plusieurs capteurs transmettent des informations pré-traitées à l'estimateur final qui doit récupérer les quantités d'intérêt.

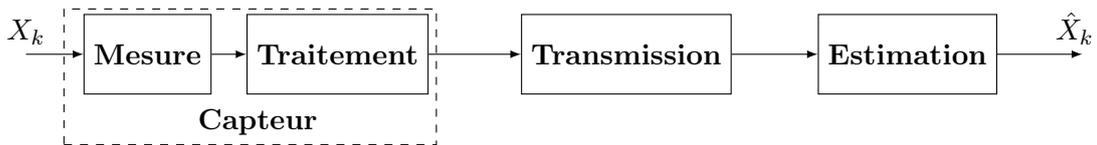


Figure B.2: Problème d'estimation scalaire à distance. Simplification scalaire et à un seul capteur du problème montré en Fig. B.1.

sortie du capteur. On suppose que le débit en bits par unité de temps est choisi de façon à ne pas dépasser la capacité du canal de transmission, de cette manière on peut considérer qu'un code suffisamment performant peut être mis en œuvre pour rendre le canal parfait.

A chaque instant  $k$ , on est intéressé par l'estimation de  $X_k$  à partir d'un bloc de mesures passées  $i_1, i_2, \dots, i_k$ . Ce problème est illustré en Fig. B.3.

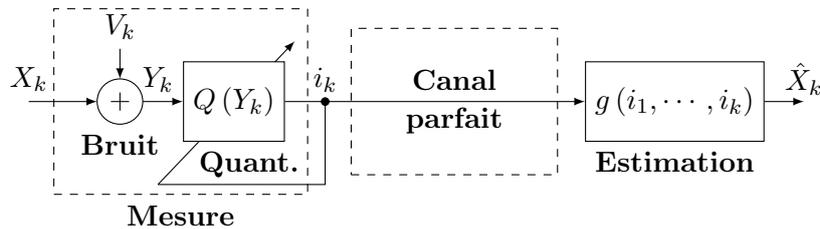


Figure B.3: Estimation à partir de mesures quantifiées. Un paramètre est mesuré avec du bruit additif, les mesures sont alors quantifiées et transmises à travers un canal de communication parfait. A partir de mesures passées, l'objectif est d'estimer  $X_k$  à chaque instant  $k$  avec la suite de fonctions  $g(\cdot)$ .

En Fig. B.3, on voit que la structure du quantifieur peut dépendre aussi de mesures quantifiées passées.

**Ce que l'on veut étudier.** On veut proposer des algorithmes pour l'estimation de  $X_k$  à partir des  $i_k$ . Le paramètre  $X_k$ , qui sera défini de façon plus précise dans la suite, peut être déterministe et constant ou aléatoire et lentement variable.

Après avoir proposé des algorithmes, on veut étudier leurs performances. Etant données les performances des algorithmes, on veut aussi étudier les effets de différents paramètres du quantifieur : seuils de quantification et résolution du quantifieur.

Pour évaluer l'impact de la quantification sur la performance d'estimation, on comparera la performance des algorithmes proposés avec leurs pendants à mesures continues.

L'objectif ici est d'estimer  $X_k$  seulement à partir des informations sur les intervalles où ses versions bruitées se trouvent.

**Ce que l'on ne veut pas étudier.** On ne veut pas reconstruire la mesure  $Y_k$  à partir de la mesure quantifiée pour ensuite estimer  $X_k$  à partir des mesures reconstruites comme si elles étaient continues. En faisant cela, on se ramènerait au groupement des solutions optimales des deux problèmes séparés, ces solutions ont déjà été abondamment étudiées dans la littérature.

On ne veut pas non plus considérer la quantification comme du bruit additif. On veut étudier le problème dans sa forme originale, c'est-à-dire, le problème d'estimation à partir des informations contenues dans des intervalles et pas dans des valeurs continues.

**Ce que l'on veut étudier mais que l'on n'étudiera pas.** Pour spécifier de façon plus précise le problème traité dans cette thèse, on doit aussi mentionner les problèmes que l'on a négligé sciemment pour rendre le sujet plus simple à traiter. Ces problèmes sont les suivants : paramètres vectoriels et quantification vectorielle, canaux de communication bruités et codage canal, signaux à variations rapides, estimation de signaux à temps continu et estimation Bayésienne d'une constante aléatoire.

## Plan du résumé

Le plan de ce résumé est le suivant :

- Estimation à partir de données quantifiées : algorithmes et performances

On détaille le problème à traiter (présentation des modèles de signaux à estimer, du bruit et du quantifieur), puis on étudie les algorithmes d'estimation et leurs performances.

- Estimation d'un paramètre constant.

D'abord, on se concentrera sur l'estimation d'un signal constant. On présentera un estimateur du maximum de vraisemblance pour deux types de quantification : binaire et multibit. Par l'analyse de sa performance asymptotique, donnée par la **borne de Cramér–Rao (BCR)** ou de façon équivalente par l'information de Fisher, on regardera l'impact du réglage de la dynamique de quantification. Comme conséquence de cette analyse, on montrera l'importance d'une approche adaptative pour le réglage du quantifieur. Finalement, on présentera des algorithmes adaptatifs de haute complexité qui, conjointement, estiment la constante et règlent le quantifieur. On montrera qu'asymptotiquement une de ces méthodes est équivalente à un algorithme récursif de basse complexité.

- Estimation d'un paramètre variable.

On passera ensuite au cas du paramètre variable. Après la présentation du modèle de variation utilisé, on définira le critère de performance d'estimation et l'estimateur optimal. Pour réaliser l'estimateur optimal, on utilisera une méthode numérique d'intégration, dans ce contexte (estimation Bayésienne) cette méthode est connue sous le nom de filtrage particulière. On étudiera ses performances avec la **borne de Cramér–Rao Bayésienne (BCRB)** et on montrera encore une fois l'importance de l'adaptativité du réglage du quantifieur. Avec l'approche adaptative, on montrera qu'asymptotiquement l'estimateur optimal ainsi obtenu pour un signal lentement variable peut être mis, lui aussi, sous une forme récursive simple.

- Quantifieurs adaptatifs pour l'estimation.

En se basant sur l'optimalité asymptotique des estimateurs vus précédemment, on proposera des algorithmes adaptatifs de basse complexité pour l'estimation et le réglage conjoint du quantifieur. On étudiera la performance de ces algorithmes pour deux modèles d'évolution de la quantité à estimer (constant ou lentement variable) et on les optimisera par rapport à ses paramètres libres. Pour la performance optimale, on étudiera la perte de performance d'estimation par rapport à des schémas équivalents pour des mesures continues.

On proposera deux extensions de l'algorithme adaptatif : une extension où l'on estime le paramètre  $x$  sans connaître l'échelle du bruit (équivalent de l'écart type) et une autre où plusieurs capteurs obtiennent des mesures quantifiées en parallèle et les transmettent à un centre de fusion qui applique un algorithme adaptatif pour l'estimation et diffuse son estimateur aux capteurs pour le réglage des quantifieurs.

- Estimation à partir de données quantifiées : approximations à haute résolution. Contrairement aux développements précédents où le réglage du quantifieur n'est fait qu'en fonction du seuil central, on se concentrera ici sur le placement de tous les seuils de quantification pour maximiser la performance d'estimation d'un paramètre arbitraire (pas seulement de centrage). Vu que ce problème est difficile à résoudre directement, on utilisera une approche asymptotique, *i.e.* on trouvera des approximations pour le quantifieur optimal quand le nombre d'intervalles de quantification est très grand.

- Approximation à haute résolution de l'information de Fisher. Après avoir montré l'importance de l'information de Fisher dans la performance d'estimation des algorithmes proposés, on appliquera cette approche asymptotique pour la maximiser en fonction des caractéristiques du quantifieur. Cette approche asymptotique permettra de trouver une caractérisation optimale du quantifieur et une expression analytique de l'information de Fisher optimale. On testera les résultats sur le problème d'estimation d'un paramètre de centrage. Pour avoir une approximation pratique des seuils de quantification optimaux, on proposera l'utilisation de l'algorithme adaptatif présenté précédemment.

Avec les expressions analytiques de l'information de Fisher, on pourra aussi étudier de façon approchée le problème d'allocation optimale de bits dans un réseau de capteurs, *i.e.* le nombre total de bits que les capteurs peuvent envoyer à un centre de fusion étant fixé, combien de bits faut-il allouer à chaque capteur ?

- Conclusions

On présentera les principaux points qui découlent des résultats de la thèse et on regardera les travaux qui peuvent être développés dans le futur : des extensions de problèmes traités ici ou des problèmes qui n'ont pas été traités pour avoir une première approche la plus simple possible.

## B.2 Estimation et quantification : algorithmes et performances

### B.2.1 Estimation d'un paramètre constant

Pour commencer cette section, on présente les modèles de mesure et de bruit utilisés.

#### Modèle de mesure

Le paramètre inconnu, constant et scalaire est

$$x \in \mathbb{R},$$

il est mesuré  $N$  fois,  $N \in \mathbb{N}^*$ , avec du bruit **indépendant et identiquement distribué (i.i.d.)**  $V_k$ . Pour  $k \in \{1, \dots, N\}$ , les mesures continues sont données par

$$Y_k = x + V_k. \quad (\text{B.1})$$

#### Modèle de bruit, hypothèses sur la distribution du bruit

Pour simplifier la suite, on considérera les hypothèses suivantes sur la distribution du bruit :

**AN1** La fonction de répartition marginale du bruit, notée  $F$ , admet une **densité de probabilité (d.d.p.)**  $f$  par rapport à la mesure de Lebesgue standard en  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

**AN2** La d.d.p.  $f(v)$  est une fonction paire, strictement positive et elle décroît strictement avec  $|v|$ .

#### Modèle du quantifieur

La sortie du quantifieur est donnée par

$$i_k = Q(Y_k),$$

où  $i_k$  est choisi dans un ensemble fini de valeurs  $\mathcal{I}$  de  $\mathbb{R}$ , cet ensemble possède  $N_I$  éléments. Le nombre d'intervalles de quantification est par conséquent noté  $N_I$ . Un exemple simple de quantifieur  $Q$  avec seuils uniformes est donné en Fig. B.4.

A l'exception de la quantification uniforme, que l'on n'imposera pas, cet exemple illustre les principales hypothèses de travail sur la structure du quantifieur :

#### Hypothèses (sur le quantifieur) :

**AQ1**  $N_I$  est un nombre naturel pair et l'ensemble  $\mathcal{I}$ , auquel  $i_k$  appartient, est

$$\mathcal{I} = \left\{ -\frac{N_I}{2}, \dots, -1, 1, \dots, \frac{N_I}{2} \right\}.$$

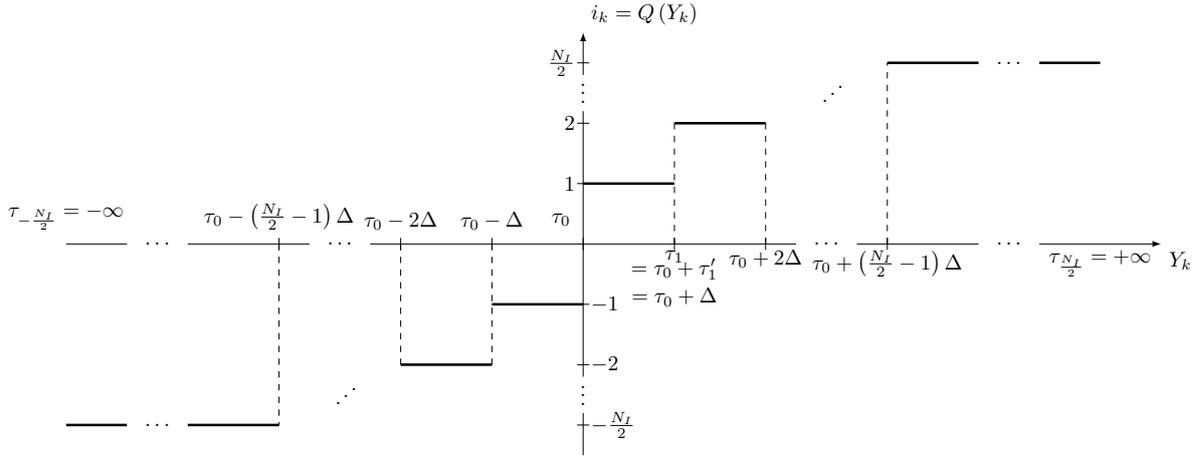


Figure B.4: Fonction de quantification  $Q(Y_k)$  avec  $N_I$  intervalles de quantification uniformes de taille  $\Delta$ . Le nombre d'intervalles de quantification  $N_I$  est pair, le quantifieur est symétrique autour d'un seuil central  $\tau_0$  et ses indices de sortie sont des entiers non nuls.

**AQ2** Le quantifieur est symétrique autour d'un seuil central. Par conséquent le vecteur de seuils  $\boldsymbol{\tau}$  peut être écrit sous la forme suivante ( $\top$  est l'opérateur transposé)

$$\boldsymbol{\tau} = \left[ \tau_{-\frac{N_I}{2}} = \tau_0 - \tau'_{\frac{N_I}{2}} \cdots \tau_{-1} = \tau_0 - \tau'_1 \quad \tau_0 \quad \tau_1 = \tau_0 + \tau'_1 \cdots \tau_{\frac{N_I}{2}} = \tau_0 + \tau'_{\frac{N_I}{2}} \right]^\top.$$

Les éléments de ce vecteur forment une séquence strictement positive et le vecteur de variations de seuil par rapport au seuil central est donné par

$$\boldsymbol{\tau}' = \left[ 0 \quad \tau'_1 \quad \cdots \quad \tau'_{\frac{N_I}{2}} = +\infty \right]^\top.$$

Avec les variations de seuils  $\tau'_i$ , on peut écrire la relation entrée-sortie du quantifieur sous une forme plus compacte :

$$i_k = i \operatorname{sign}(Y_k - \tau_0), \quad \text{pour } |Y_k - \tau_0| \in [\tau'_{i-1}, \tau'_i]. \quad (\text{B.2})$$

### Maximum de vraisemblance, borne de Cramér–Rao et information de Fisher

On veut estimer  $x$  à partir de  $i_{1:N} = \{i_1, \dots, i_N\}$ , on cherche donc un estimateur

$\hat{X}(i_{1:N})$  - qui est aléatoire, vu que les  $i_{1:N}$  sont aléatoires aussi,

le plus proche de  $x$ . Proche dans ce cas peut être traduit de façon quantitative par un critère de performance. Dans notre cas on considère comme critère de performance l'**erreur quadratique moyenne (EQM)**

$$\text{EQM} = \mathbb{E} \left[ \left( \hat{X} - x \right)^2 \right].$$

Si l'on impose que l'estimateur soit non biaisé *i.e.*

$$\mathbb{E} \left[ \hat{X} \right] = x,$$

au moins quand  $N \rightarrow \infty$ , on sait que l'estimateur qui minimise l'EQM asymptotiquement (et donc qui maximise la performance asymptotiquement) est l'estimateur du **maximum de vraisemblance (MV)** [Kay 1993, p. 160]. Le MV consiste à maximiser la fonction de vraisemblance par rapport au paramètre inconnu. La vraisemblance est la distribution conjointe des mesures (celles-ci étant figées après observation) et elle est une fonction du paramètre inconnu (celui-ci considéré comme une variable). Pour le problème que l'on traite ici, la vraisemblance pour un bloc de mesures indépendantes  $i_{1:N}$  est

$$L(x; i_{1:N}) = \prod_{k=1}^N \mathbb{P}(i_k; x),$$

où  $\mathbb{P}(i_k; x)$  est la probabilité d'avoir une valeur quantifiée  $i_k$  à l'instant  $k$  pour un paramètre  $x$ . On peut réécrire cette probabilité en fonction des seuils et de la fonction de répartition :

$$\mathbb{P}(i_k; x) = \begin{cases} \mathbb{P}(\tau_{i_k-1} \leq Y_k < \tau_{i_k}), & \text{si } i_k > 0, \\ \mathbb{P}(\tau_{i_k} \leq Y_k < \tau_{i_k+1}), & \text{si } i_k < 0, \end{cases}$$

avec la définition  $Y_k = x + V_k$  donnée par (B.1)

$$\begin{aligned} \mathbb{P}(i_k; x) &= \begin{cases} \mathbb{P}(\tau_{i_k-1} \leq x + V_k < \tau_{i_k}), & \text{si } i_k > 0, \\ \mathbb{P}(\tau_{i_k} \leq x + V_k < \tau_{i_k+1}), & \text{si } i_k < 0, \end{cases} \\ &= \begin{cases} F(\tau_{i_k} - x) - F(\tau_{i_k-1} - x), & \text{si } i_k > 0, \\ F(\tau_{i_k+1} - x) - F(\tau_{i_k} - x), & \text{si } i_k < 0. \end{cases} \end{aligned}$$

L'estimateur du MV est donné par

$$\hat{X}_{MV,q} = \hat{X}_{MV}(i_{1:N}) = \underset{x}{\operatorname{argmax}} L(x; i_{1:N}),$$

ou de façon équivalente par

$$\hat{X}_{MV,q} = \underset{x}{\operatorname{argmax}} \log L(x; i_{1:N}).$$

On se concentre maintenant sur les performances de cet estimateur, qui, à cause du manque de résultats à taille d'échantillon finie, ne sont connues qu'en régime asymptotique.

L'EQM du MV peut être écrit, en général, sous la forme suivante

$$\mathbb{E} \left[ \left( \hat{X}_{MV,q} - x \right)^2 \right] = \left[ \mathbb{E} \left( \hat{X}_{MV,q} - x \right) \right]^2 + \mathbb{V}\text{ar} \left( \hat{X}_{MV,q} \right) = \text{biais}^2 + \text{variance}.$$

Comme mentionné auparavant, le MV est asymptotiquement non biaisé:

$$\mathbb{E} \left[ \hat{X}_{MV,q} \right] \underset{N \rightarrow \infty}{=} x.$$

Par conséquent, son EQM n'est caractérisée que par sa variance.

La variance asymptotique du MV atteint la BCR [Kay 1993, p. 160] (qui est aussi une borne inférieure sur la variance des estimateurs non biaisés dans un contexte non asymptotique [Kay 1993, p. 30]) :

$$\text{Var} \left( \hat{X}_{MV,q} \right) \underset{N \rightarrow \infty}{\sim} \text{BCR}_q,$$

où le symbole  $\underset{N \rightarrow \infty}{\sim}$  est utilisé pour représenter une équivalence.

La BCR est l'inverse de l'information de Fisher [Kay 1993, p. 30]  $I_q$ , qui est la variance de la fonction score  $S_q$ . En partant de la fonction score pour  $N$  mesures quantifiées, on a les expressions suivantes

$$\begin{aligned} S_{q,1:N} &= \frac{\partial \log L(x; i_{1:N})}{\partial x} && \text{- fonction score,} \\ I_{q,1:N} &= \mathbb{E} [S_{q,1:N}^2] = \mathbb{E} \left\{ \left[ \frac{\partial \log L(x; i_{1:N})}{\partial x} \right]^2 \right\} && \text{- Fisher,} \\ \text{Var} \left( \hat{X}_{MV,q} \right) \underset{N \rightarrow \infty}{\sim} \text{BCR}_q &= \frac{1}{I_{q,1:N}} = \frac{1}{\mathbb{E} \left\{ \left[ \frac{\partial \log L(x; i_{1:N})}{\partial x} \right]^2 \right\}} && \text{- variance et BCR.} \end{aligned}$$

L'indice  $1:N$  est utilisé pour indiquer que ces quantités sont relatives à  $N$  mesures. Pour simplifier, on utilisera la notation  $S_q$  et  $I_q$  dans le contexte d'une mesure quantifiée arbitraire.

Sous l'hypothèse de mesures indépendantes on a

$$\text{Var} \left( \hat{X}_{MV,q} \right) \underset{N \rightarrow \infty}{\sim} \text{BCR}_q = \frac{1}{NI_q}.$$

La fonction score pour une mesure  $S_q$  est

$$S_q = \frac{\partial \log L(x; i_k)}{\partial x} = \frac{\frac{\partial \mathbb{P}(i_k; x)}{\partial x}}{\mathbb{P}(i_k; x)}$$

et l'information de Fisher correspondante est

$$\begin{aligned} I_q &= \mathbb{E} \left\{ \left[ \frac{\partial \log L(x; i_k)}{\partial x} \right]^2 \right\} = \sum_{i_k \in \mathcal{I}} \left[ \frac{\frac{\partial \mathbb{P}(i_k; x)}{\partial x}}{\mathbb{P}(i_k; x)} \right]^2 \mathbb{P}(i_k; x), \\ &= \sum_{i_k \in \mathcal{I}} \frac{\left[ \frac{\partial \mathbb{P}(i_k; x)}{\partial x} \right]^2}{\mathbb{P}(i_k; x)}. \end{aligned}$$

Si on note  $\varepsilon = \tau_0 - x$  la différence entre le seuil central et le paramètre, on peut réécrire  $I_q$  sous la forme suivante :

$$I_q = \sum_{i_k=1}^{i_k=\frac{NI}{2}} \left\{ \frac{[f(\varepsilon + \tau'_{i_k-1}) - f(\varepsilon + \tau'_{i_k})]^2}{F(\varepsilon + \tau'_{i_k}) - F(\varepsilon + \tau'_{i_k-1})} + \frac{[f(\varepsilon - \tau'_{i_k}) - f(\varepsilon - \tau'_{i_k-1})]^2}{F(\varepsilon - \tau'_{i_k-1}) - F(\varepsilon - \tau'_{i_k})} \right\}. \quad (\text{B.3})$$

### Influence du quantifieur sur la performance

La performance de l'estimateur est donc caractérisée par  $\text{BCR}_q$  ou de façon équivalente par  $I_q$ , par conséquent, pour étudier l'influence du quantifieur sur la performance d'estimation on peut, de façon quantitative, étudier comment  $\text{BCR}_q$  ou  $I_q$  se comportent en fonction de  $N_I$  et  $\tau$ . On commence par quelques propriétés générales de  $I_q$  :

**Perte induite par la quantification :** si on note  $S_c$  et  $I_c$  la fonction score et l'information de Fisher du problème d'estimation équivalent avec des mesures continues, on peut montrer que

$$I_c - I_q = \mathbb{E} \left[ (S_c - S_q)^2 \right] \geq 0.$$

Ce qui veut dire que  $I_q$  est majorée par  $I_c$  et qu'il existe une perte de performance inhérente à la quantification donnée de manière quantitative par  $\mathbb{E} \left[ (S_c - S_q)^2 \right]$ .

**Monotonie de  $I_q$  :** on peut montrer aussi que si l'on ajoute un seuil à un vecteur de seuils  $\tau$ , alors l'information de Fisher correspondant au nouveau vecteur de seuils est toujours plus grande ou égale à l'information de Fisher précédente. Cela veut dire que l'information de Fisher croît de façon monotone en fonction de  $N_I$  (pour une séquence de seuils construite en ajoutant des seuils).

Une question qui se pose pour la suite est : comme on peut construire une séquence de seuils telle que  $I_q$  croît de façon monotone en  $N_I$  et comme on sait que  $I_q$  est majorée par  $I_c$ , est-ce que  $I_q$  converge vers  $I_c$  ? On répondra à cette question plus loin dans ce résumé.

Maintenant, on passe à l'étude de la performance d'estimation en fonction de la position des seuils. On commence par le cas binaire.

**Cas binaire :** dans le cas binaire on peut utiliser l'expression de l'information de Fisher (B.3) pour obtenir la BCR suivante

$$\text{BCR}_q^B = \frac{F(\varepsilon) [1 - F(\varepsilon)]}{N f^2(\varepsilon)}.$$

L'analyse de la performance se réduit alors à l'analyse de la fonction

$$B(\varepsilon) = N \text{BCR}_q^B = \frac{F(\varepsilon) [1 - F(\varepsilon)]}{f^2(\varepsilon)}.$$

L'étude de cette fonction dans le cas Gaussien ( $f(\varepsilon) = \frac{1}{\sqrt{\pi}\delta} \exp \left[ -\left(\frac{\varepsilon}{\delta}\right)^2 \right]$ ) a été réalisée par [Papadopoulos 2001] et [Ribeiro 2006a], son comportement est illustré en Fig. B.5.

On peut noter que la valeur minimale de  $B$  est atteinte lorsque  $\varepsilon = 0$  et que  $B(\varepsilon)$  augmente lorsque  $|\varepsilon|$  augmente. Par conséquent la valeur optimale du seuil  $\tau_0^*$  est égale à  $x$  et la valeur minimale de  $B$  est  $B^* = \frac{1}{4f^2(0)} = \frac{\pi\delta^2}{4}$ . Si on compare cette valeur avec la BCR pour les mesures continues,  $\text{BCR}_c \times N = \frac{\delta^2}{2}$ , on peut constater que la perte produite par la quantification binaire est d'environ 2dB, ce qui est, de façon surprenante, très peu.

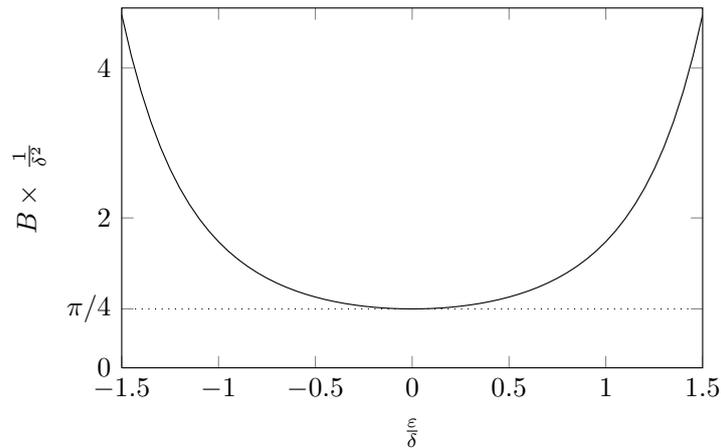


Figure B.5: BCR normalisée  $B$  en fonction de la différence normalisée  $\frac{\varepsilon}{\delta}$  entre le seuil et le paramètre. La distribution du bruit est Gaussienne et le facteur de normalisation  $\delta$  est le paramètre d'échelle de la Gaussienne. Des normalisations sont réalisées sur les deux axes pour que la courbe affichée soit indépendante de  $\delta$ .

Notez que pour avoir cette petite perte, il faut que  $\tau_0 = x$ , ce qui est impossible en pratique, puisque  $x$  est le paramètre inconnu à estimer. Notez aussi que  $B$  est une fonction assez sensible par rapport à la position du seuil, si l'on place  $\tau_0$  loin du paramètre la performance d'estimation est très dégradée.

On peut montrer que pour d'autres distributions couramment utilisées comme modèle de bruit, tels que la distribution de Laplace et la distribution de Cauchy, des conclusions similaires peuvent être obtenues :

- La valeur optimale du seuil de quantification est  $\tau_0^* = x$ .
- La perte due à la quantification est petite, si on utilise  $\tau_0^*$ .
- La performance se dégrade lorsque  $\tau_0$  s'éloigne de  $x$ .

**Cas asymétriques :** même si pour plusieurs distributions de bruit couramment utilisées la fonction  $B$  a un comportement symétrique en forme de « u », ce comportement ne se généralise pas à toutes les distributions symétriques, comme on s'y attend intuitivement. Il suffit que la condition suivante ne soit pas satisfaite

$$-f^{(2)}(0) > 4f^3(0),$$

pour que la fonction  $B$  ait  $\varepsilon = 0$  comme maximum local. Ceci veut dire que pour des densités ne respectant pas cette condition, le point de quantification optimal n'est pas  $x$  et la quantification optimale doit être faite de manière asymétrique par rapport à la distribution des mesures.

Un cas simple de distribution symétrique qui ne respecte pas cette condition est la distribution *ad hoc* suivante

$$f(\varepsilon) = \begin{cases} f_{GL}(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon+\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{pour } \varepsilon < -\frac{\alpha}{2}, \\ f_U(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma}, & \text{pour } -\frac{\alpha}{2} \leq \varepsilon \leq \frac{\alpha}{2}, \\ f_{GR}(\varepsilon) = \frac{1}{C\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon-\frac{\alpha}{2}}{\sigma}\right)^2\right], & \text{pour } \varepsilon > \frac{\alpha}{2}. \end{cases}$$

Un exemple de BCR obtenue avec cette distribution (et de la performance pratique du MV) est donné en Fig. B.6

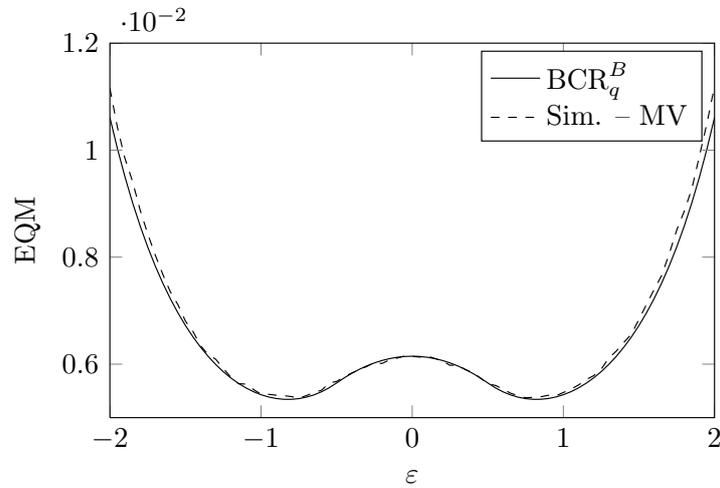


Figure B.6:  $BCR_q^B$  et EQM simulée du MV pour un bruit distribué selon la loi *ad hoc*. La borne et l'EQM simulée ont été évaluées pour  $N = 500$  et  $\varepsilon$  dans l'intervalle  $[-2, 2]$ . L'EQM du MV a été évaluée par une simulation Monte Carlo avec  $10^5$  réalisations de blocs de 500 échantillons. On a utilisé aussi :  $\alpha = 1$  et  $\sigma = 1$ .

**Cas multibit :** pour l'estimation avec MV et une quantification multibit, la performance en fonction du quantifieur peut être étudiée au travers de l'analyse de l'information de Fisher (B.3). Comme résultat de cette analyse on trouve que :

- la dynamique de quantification doit être proche du paramètre pour maximiser la performance d'estimation.
- Pour des variations de seuils symétriques bien choisies, le choix  $\tau_0 = x$  est optimal pour plusieurs types de bruit (pour une classe plus large que dans le cas binaire).
- La performance se dégrade rapidement quand la dynamique de quantification est placée loin du paramètre à estimer.
- Le problème d'optimisation de  $I_q$  en fonction de  $\tau'$  est difficile à résoudre pour  $N_B = \log_2(N_I) > 3$ .

### Quantification adaptative : l'approche à haute complexité

La conclusion directe des résultats précédents est la suivante : on doit placer le seuil central le plus proche possible du paramètre  $x$ . Or comme  $x$  est inconnu, on peut se baser sur les dernières mesures quantifiées pour estimer  $x$ , et comme on s'attend à ce que l'estimateur soit, au moins après un certain moment, proche de  $x$ , on placera le seuil central de quantification exactement sur cette dernière estimation. Ceci équivaut donc à une approche d'estimation où le processus de mesure, le quantifieur, est à tout instant adapté pour améliorer la performance d'estimation.

Dans la littérature cette approche adaptative a été proposée en [Li 2007] et [Fang 2008], dans le cas binaire et Gaussien.

La première méthode, proposée en [Li 2007], consiste à générer des estimations simples du paramètre au niveau du capteur avec la mise à jour du seuil central donnée par

$$\tau_{0,k} = \tau_{0,k-1} + \gamma i_k,$$

où  $\gamma$  est un pas d'adaptation. Les mesures quantifiées sont donc transmises à l'estimateur distant qui possède suffisamment de puissance de calcul pour générer des estimations plus précises en utilisant le MV. Dans ce cas, le MV consiste à maximiser la vraisemblance suivante

$$\begin{aligned} L(x; i_{1:N}) &= \mathbb{P}(i_{1:N}; x) = \prod_{k=1}^N \mathbb{P}(i_k | i_{k-1}, \dots, i_1; x) \\ &= \prod_{k=1}^N \mathbb{P}(i_k | \tau_{0,k-1}; x) \\ &= \prod_{k=1}^N [1 - F(\tau_{0,k-1} - x)]^{\frac{1+i_k}{2}} F(\tau_{0,k-1} - x)^{\frac{1-i_k}{2}}, \quad (\text{B.4}) \\ \log L(x; i_{1:N}) &= \sum_{k=1}^N \left\{ \frac{1+i_k}{2} \log [1 - F(\tau_{0,k-1} - x)] + \frac{1-i_k}{2} \log F(\tau_{0,k-1} - x) \right\}. \end{aligned}$$

Du fait de la symétrie du problème, on espère que le seuil  $\tau_{0,k}$  va tendre en moyenne vers le point  $x$ , de cette façon le seuil central va fluctuer autour du vrai paramètre et donnera une performance d'estimation proche de l'optimum.

La performance asymptotique de l'algorithme a été étudiée plus en détail dans [Fang 2008]. Elle est obtenue à partir de l'inverse de l'information de Fisher

$$I_{q,1:N} = \sum_{k=1}^N \mathbb{E} \left[ \frac{f^2(\tau_{0,k-1} - x)}{F(\tau_{0,k-1} - x) [1 - F(\tau_{0,k-1} - x)]} \right],$$

où l'espérance est évaluée par rapport à la distribution de  $\tau_{0,k-1}$ , qui maintenant n'est plus fixé, ni déterministe. Sachant que la distribution des seuils tend vers une distribution asymptotique, quand  $N \rightarrow \infty$ , cette information de Fisher peut être approchée par l'information de Fisher avec la distribution asymptotique des seuils  $\mathbf{p}_\infty$  :

$$\lim_{N \rightarrow \infty} \frac{I_{q,1:N}}{N} = \tilde{\mathbf{I}}_q^\top \tilde{\mathbf{p}}_\infty,$$

où

$$\mathbf{I}'_q = \left[ \cdots, \frac{f^2(-\gamma - x)}{F(-\gamma - x)[1 - F(-\gamma - x)]}, \frac{f^2(0 - x)}{F(0 - x)[1 - F(0 - x)]}, \frac{f^2(\gamma - x)}{F(\gamma - x)[1 - F(\gamma - x)]}, \cdots \right]^\top.$$

$\tilde{\mathbf{p}}_\infty$  étant de taille infinie, pour avoir la performance asymptotique en pratique, [Fang 2008] propose de tronquer le vecteur  $\tilde{\mathbf{p}}_\infty$ .

Un problème avec l'approche adaptative qui vient d'être présentée réside dans la présence d'une fluctuation asymptotique sur l'emplacement de la dynamique de quantification, ceci entraîne une sous optimalité asymptotique vu que l'on devrait avoir  $\tau_{0,k} = x$  pour une performance optimale. Ce problème peut être résolu en utilisant un algorithme qui converge vers  $x$  quand  $N \rightarrow \infty$ . En ajoutant de la complexité au niveau du capteur ou un lien de retour de l'estimateur vers le capteur, une idée assez directe consiste à utiliser la dernière estimation du MV comme nouveau seuil central :

$$\tau_{0,k} = \hat{X}_{MV,k}.$$

Cette idée a été proposée initialement pour le cas binaire Gaussien dans [Fang 2008] où les auteurs prétendent qu'asymptotiquement la performance en terme de variance est équivalente à  $\frac{1}{NI_q(0)}$ . Ce qui équivaut à dire que l'algorithme est asymptotiquement optimal.

On peut étendre de façon assez naturelle cette approche adaptative au cas multibit et non Gaussien. Moyennant certaines contraintes sur la fonction  $I_q(\varepsilon)$ , on peut montrer que la performance asymptotique est aussi optimale

$$\text{BCR}_q \underset{N \rightarrow \infty}{\sim} \frac{1}{NI_q(0)}.$$

Pour réduire la complexité de l'algorithme, qui résout un problème d'optimisation à chaque nouvelle mesure, dans [Papadopoulos 2001] une étude heuristique de la forme asymptotique du MV dans le cas Gaussien binaire avec le seuil adaptatif a été réalisée, elle montre qu'asymptotiquement le MV adaptatif a une forme récursive de très basse complexité :

$$\hat{X}_k = \tau_{0,k} = \hat{X}_{k-1} + \frac{\delta\sqrt{\pi}}{2k} i_k.$$

En se basant sur les propriétés asymptotiques du MV et en considérant que l'estimateur est suffisamment proche du vrai paramètre, on peut montrer que, même dans un contexte non Gaussien, le MV adaptatif est équivalent à une forme asymptotique simple

$$\hat{X}_k \approx \hat{X}_{k-1} + \frac{i_k}{2kf(0)}.$$

On peut se poser quelques questions sur la forme équivalente simple donnée ci-dessus :

- peut-elle converger quand l'erreur initiale  $|\varepsilon| = |\tau_0 - x|$  est arbitraire (pas nécessairement petite) ?
- Peut-on étendre cet algorithme à basse complexité au cas  $N_I > 2$  ?

On donnera les réponses à ces questions en Sous-section B.2.3.

### B.2.2 Estimation d'un paramètre variable

On passe maintenant à l'estimation d'un paramètre variable.

#### Modèle du paramètre

Le paramètre à estimer est défini comme un processus stochastique, il n'est donc pas seulement variable mais aussi aléatoire. A chaque instant  $k \in \mathbb{N}^*$ , la **variable aléatoire (v.a.)**  $X_k$  est donnée par le modèle de Wiener suivant :

$$X_k = X_{k-1} + W_k, \quad k > 0,$$

où  $W_k$  est le  $k$ -ème élément d'une séquence indépendante de v.a. Gaussiennes. Sa moyenne est donnée par  $u_k$  et sa variance est une constant connue  $\sigma_w^2$ . Si  $u_k = 0$ , alors  $X_k$  forme un processus de Wiener à temps discret classique, sinon, on l'appelle processus de Wiener avec dérive. La distribution initiale  $X_k$  est supposée Gaussienne de moyenne  $x'_0$  et de variance connue  $\sigma_0^2$ .

#### Modèle du quantifieur

Pour poursuivre le paramètre, on suppose que le quantifieur peut être dynamique avec un vecteur de seuils donné par  $\tau_k$  :

$$\tau_k = \left[ \tau_{-\frac{N_I}{2},k} \cdots \tau_{-1,k} \quad \tau_{0,k} \quad \tau_{1,k} \cdots \tau_{\frac{N_I}{2},k} \right]^\top.$$

Les mesures quantifiées sont encore données par  $Q(\cdot)$  définie en (B.2)

$$i_k = Q(Y_k),$$

mais dans ce cas, cette fonction peut varier dans le temps.

#### Estimateur optimal

De façon analogue au cas constant, on veut un estimateur

$$\hat{X}(i_{1:k})$$

qui minimise l'EQM

$$\text{EQM}_k = \mathbb{E} \left[ \left( \hat{X}_k - X_k \right)^2 \right].$$

Il est connu que l'estimateur optimal minimisant l'EQM est donné par la moyenne *a posteriori*

$$\hat{X}_k = \mathbb{E}_{X_k|i_{1:k}}(X_k) = \int_{\mathbb{R}} x_k p(x_k|i_{1:k}) dx_k, \quad (\text{B.5})$$

où  $p(x_k|i_{1:k})$  est la densité *a posteriori*. Cet estimateur est non biaisé et sa performance est donnée par

$$\begin{aligned} \text{EQM}_k &= \mathbb{E}_{i_{1:k}} [\text{Var}_{X_k|i_{1:k}}(X_k)] \\ &= \sum_{i_{1:k} \in \mathcal{I}^{\otimes k}} \left\{ \int_{\mathbb{R}} (x_k - \mathbb{E}_{X_k|i_{1:k}}(X_k))^2 p(x_k|i_{1:k}) dx_k \right\} \mathbb{P}(i_{1:k}). \end{aligned} \quad (\text{B.6})$$

### Filtrage particulaire

La solution donnée par (B.5) est difficile à mettre en œuvre de façon analytique car, dans la plupart des cas, l'évaluation directe de la densité  $p(x_k|i_{1:k})$  et de l'intégrale n'est pas possible. On doit donc utiliser une méthode numérique pour l'évaluation de la densité et de l'intégrale. Dans notre cas, on peut utiliser la méthode de Monte Carlo, qui est une méthode d'intégration numérique basée sur la simulation.

Dans le contexte du problème étudié, l'application de la méthode de Monte Carlo est connue sous le nom d'échantillonnage d'importance avec rééchantillonnage, populairement, elle est aussi connue sous le nom **filtrage particulaire (FP)**, que l'on utilise pour la suite.

Un algorithme particulaire pour l'estimation du paramètre variable à partir de données quantifiées est donné ci-dessous :

0. Définition des poids  $\tilde{w}(x_0^{(j)}) = \frac{1}{N_S}$  et initialisation de  $N_S$  particules (échantillons)  $\{x_0^{(1)}, \dots, x_0^{(N_S)}\}$  par l'échantillonnage de la densité

$$p(x_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[ -\frac{1}{2} \left( \frac{x_0 - x'_0}{\sigma_0} \right)^2 \right].$$

A chaque instant  $k$ ,

1. pour  $j$  de 1 à  $N_S$ , l'échantillonnage de  $X_k^{(j)}$  est réalisé avec la d.d.p.

$$p(x_k|x_{k-1}^{(j)}) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp \left[ -\frac{1}{2} \left( \frac{x_k - x_{k-1}^{(j)} - u_k}{\sigma_w} \right)^2 \right],$$

2. pour  $j$  de 1 à  $N_S$ , on évalue et on normalise les poids

$$w(x_{1:k}^{(j)}) = \mathbb{P}(i_k|x_k^{(j)}) \tilde{w}(x_{1:k-1}^{(j)}), \quad \tilde{w}(x_{1:k}^{(j)}) = \frac{w(x_{1:k}^{(j)})}{\sum_{j=1}^{N_S} w(x_{1:k}^{(j)})},$$

où  $\mathbb{P}(i_k|x_k^{(j)})$  est donnée par

$$\mathbb{P}(i_k|x_k) = \begin{cases} F(\tau_{i_k,k} - x_k) - F(\tau_{i_k-1,k} - x_k), & \text{si } i_k > 0, \\ F(\tau_{i_k+1,k} - x_k) - F(\tau_{i_k,k} - x_k), & \text{si } i_k < 0. \end{cases}$$

3. L'estimation est donc donnée par

$$\hat{x}_k \approx \sum_{j=1}^{N_S} x_k^{(j)} \tilde{w} \left( x_{1:k}^{(j)} \right).$$

4. Finalement, on évalue le nombre effectif de particules utilisées

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^{N_S} \tilde{w}^2 \left( x_{1:k}^{(j)} \right)},$$

si  $N_{\text{eff}} < N_{\text{seuil}}$ , alors une procédure de rééchantillonnage multinomial doit être réalisée (rééchantillonnage des valeurs  $x_{1:k}^{(j)}$  avec les poids  $w \left( x_{1:k}^{(j)} \right)$  comme probabilités de tirage).

### Evaluation de la performance

Quand  $N_S$  tend vers l'infini, il est connu que le filtrage particulaire converge vers l'estimateur optimal. Dans ce cas, si l'on considère qu'un  $N_S$  suffisamment grand est utilisé, on peut utiliser (B.6) pour obtenir la performance du filtrage particulaire. Cependant, l'expression (B.6) souffre du même problème que (B.5), l'impossibilité d'être évaluée analytiquement dans la plupart des cas. Comme solution, on pourrait donc avoir recours à une procédure d'intégration numérique similaire à celle utilisée pour obtenir l'estimateur. Le problème avec cette dernière approche réside dans le fait que si l'on voulait utiliser la performance de simulation pour la conception d'un système de mesure (choix de la qualité du capteur, choix de  $N_I$ , etc) on serait obligé de réaliser des simulations pour plusieurs valeurs possibles des paramètres. Ceci demanderait un temps de simulation très long. Comme alternative, on utilise une borne inférieure sur l'EQM, qui peut être obtenue de façon analytique. Cette borne est la version Bayésienne de la BCR, la BCRB.

$$\text{EQM}_k \geq \text{BCRB}_k = \frac{1}{J_k}, \quad (\text{B.7})$$

où  $J_k$  est l'information Bayésienne donnée sous forme récursive par

$$J_k = \frac{1}{\sigma_w^2} + \mathbb{E} [I_q(\varepsilon_k)] - \frac{1}{\sigma_w^4 \left( \frac{1}{\sigma_w^2} + J_{k-1} \right)}. \quad (\text{B.8})$$

### L'innovation quantifiée

Pour les modèles symétriques de bruit couramment utilisés (Gaussien, Laplacien et Cauchy), on sait que  $I_q(\varepsilon)$  est maximisée quand  $\varepsilon = 0$  et que  $I_q(\varepsilon)$  décroît quand  $|\varepsilon|$  augmente. Or, d'après (B.8), on voit que plus  $\tau_{0,k}$  est proche de la réalisation du paramètre  $x_k$ , plus grande est l'information Bayésienne et, par conséquent, plus petite est la BCRB. Si l'on suppose que la BCRB est une borne suffisamment serrée, de façon à ce que l'on puisse accepter son comportement comme une approximation de l'EQM, alors plus proche  $\tau_{0,k}$  est de  $x_k$ , plus

petite est l'EQM. Ceci indique que pour avoir une performance d'estimation améliorée, la dynamique de quantification doit se déplacer dans le temps de façon à suivre le paramètre.

L'approche  $-\tau_{0,k} = x_k^-$  est, encore une fois, impossible à mettre en œuvre car on ne connaît pas  $x_k$ . On doit alors accepter une perte de performance et utiliser la valeur la plus proche de  $x_k$  disponible, dans notre cas, la prédiction de  $x_k$ . Ceci consiste donc à quantifier l'innovation apportée par la nouvelle mesure. Avec le modèle d'évolution de  $X_k$  utilisé ici, on peut montrer que cette prédiction est

$$\hat{X}_{k|k-1} = \tau_{0,k} = \hat{X}_{k-1} + u_k. \quad (\text{B.9})$$

On peut donc modifier l'algorithme particulière, pour inclure la mise à jour adaptative du centre du quantifieur (B.9). La performance du nouvel algorithme peut être approchée par la BCRB (B.7).

Si l'on suppose que le signal est lent, *i.e.* que  $\sigma_w$  est petit devant l'écart type du bruit de mesure, alors on peut s'attendre à ce que l'erreur d'estimation soit petite après un certain temps, vu que l'estimateur a le temps de « moyenner » les mesures avant un changement important d'amplitude du signal. On peut donc remplacer  $\mathbb{E}[I_q(\varepsilon_k)]$  par sa borne supérieure  $I_q(0)$  dans l'expression récursive pour l'information Bayésienne. Si on calcule le point fixe de l'expression résultante pour  $\sigma_w$  petit, on trouve une expression asymptotique simple pour la borne sur l'EQM :

$$\text{EQM}_\infty \geq \frac{\sigma_w}{\sqrt{I_q(0)}} + o(\sigma_w), \quad (\text{B.10})$$

où la notation  $o(\sigma_w)$  est utilisée pour représenter un terme qui est négligeable devant  $\sigma_w$  quand  $\sigma_w \rightarrow 0$ , c'est-à-dire, quand le signal est lent.

### Estimateur asymptotique optimal d'un signal lent

On peut se demander si, de la même façon que pour le MV, il existe une forme asymptotique simple pour l'estimateur optimal d'un paramètre variable quand on utilise  $\tau_{0,k} = \hat{X}_{k-1}$ . En effet, si l'on suppose encore une fois que  $\sigma_w$  est petit devant l'écart type du bruit, on peut montrer que l'estimateur asymptotique optimal est donné par la forme récursive suivante :

$$\hat{X}_k \approx \hat{X}_{k-1} + u_k - \frac{\sigma_w}{\sqrt{I_q(0)}} \frac{f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}{\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}}, \quad (\text{B.11})$$

où  $\mathbb{P}(i_k|X_k) \Big|_{\hat{X}_{k|k-1}=X_k}$  est la probabilité d'avoir la sortie  $i_k$  quand  $\hat{X}_{k|k-1} = X_k$ . Sa dérivée par rapport à l'erreur  $\varepsilon_k$  évaluée au point  $\varepsilon_k = 0$  est  $f_d(i_k, \hat{X}_{k|k-1}, X_k) \Big|_{\hat{X}_{k|k-1}=X_k}$ . Notez que cette forme a une complexité encore plus basse que celle du MV adaptatif, car maintenant le gain qui corrige l'estimateur à chaque instant est constant.

On peut montrer qu'une approximation de l'EQM asymptotique de cet algorithme est

$$\text{EQM} \approx \frac{\sigma_w}{\sqrt{I_q(0)}}.$$

Cette performance pour  $\sigma_w$  petit se raccorde bien avec la BCRB asymptotique donnée en (B.10). On constate aussi que, de la même façon que pour le MV adaptatif, la performance de l'estimateur optimal avec seuil central adaptatif dépend des caractéristiques de la mesure (bruit et vecteur de seuils  $\tau'$ ) au travers de l'information de Fisher  $I_q(0)$ . Par conséquent, pour caractériser complètement le quantifieur optimal pour l'estimation on doit, encore une fois, maximiser  $I_q(0)$  par rapport à  $\tau'$ . Comme on l'a mentionné précédemment, ce problème est difficile à résoudre de façon directe pour  $N_B > 3$ , on doit donc essayer de trouver une approximation de la solution, cette approximation sera le sujet de la Section B.3.

Comme dans le cas du paramètre constant, on peut se poser la question suivante :

- est-ce que la forme réursive (B.11) peut converger quand l'erreur initiale sur  $\hat{X}_k$ ,  $|\varepsilon_0| = |\hat{X}_0 - X_0|$ , est arbitraire (pas nécessairement petite) ?

On répondra à cette question dans la suite.

### B.2.3 Quantifieurs adaptatifs pour l'estimation

Dans cette sous-section on traite des questions posées précédemment au sujet de l'application des algorithmes asymptotiques de basse complexité. Pour cela, on impose d'abord la structure de quantification adaptative, puis on définit un algorithme d'estimation général qui a comme cas spécifiques les formes asymptotiquement optimales vues précédemment. Après la définition de l'algorithme, on s'intéresse à l'analyse de sa performance : son biais et sa variance asymptotique. Suite à l'optimisation de sa performance par rapport aux paramètres libres de l'algorithme, on analyse la perte de performance d'estimation par rapport au cas continu (mesures continues). A la fin, on présente aussi des extensions de l'algorithme à d'autres problèmes : estimation conjointe de  $x$  et de l'échelle du bruit  $\delta$  et estimation à partir de mesures obtenues par plusieurs capteurs.

#### Modèle du signal

Dans la suite, le paramètre à estimer est considéré soit constant  $X_k = x$ , soit lentement variable  $X_k = X_{k-1} + W_k$  (avec  $\sigma_w$  petit et  $u_k = u$  petite ou nulle).

#### Modèle du quantifieur

On a vu que le seuil central du quantifieur doit être mis à jour dynamiquement pour améliorer la performance d'estimation. Pour rendre explicite cette caractéristique du quantifieur, on imposera un biais réglable  $b_k$  à l'entrée du quantifieur, pour régler l'amplitude de l'entrée on appliquera aussi un gain  $\frac{1}{\Delta}$ . La fonction de quantification est donc donnée par

$$i_k = Q \left( \frac{Y_k - b_k}{\Delta} \right).$$

Avec un biais réglable et un gain d'entrée, on peut fixer la structure du quantifieur avec un seuil central statique à zéro et d'autres seuils qui seront égaux aux décalages  $\tau'$ .

La sortie du quantifieur réglable est donnée par

$$i_k = Q\left(\frac{Y_k - b_k}{\Delta_k}\right) = i \operatorname{sign}(Y_k - b_k), \quad \text{pour } \frac{|Y_k - b_k|}{\Delta} \in [\tau'_{i-1}, \tau'_i).$$

A partir des mesures quantifiées, l'objectif est d'estimer le paramètre  $X_k$ , un objectif secondaire est de régler les paramètres  $b_k$  et  $\Delta$  pour avoir une performance d'estimation améliorée. Comme l'estimateur  $\hat{X}_k$  de  $X_k$  peut être utilisé dans des applications temps réel, il serait intéressant de l'estimer en ligne.

Dans les sous-sections précédentes on a vu que :

- dans le cas où  $X_k = x$ , si on place le centre du quantifieur sur la dernière estimation, on peut avoir un algorithme asymptotiquement optimal.
- dans le cas où  $X_k$  est variable, la performance peut être améliorée si on place le seuil central sur la prédiction du signal. Quand le signal a pour modèle un processus de Wiener, la prédiction et donc le seuil central sont donnés par  $\hat{X}_{k-1}$  et quand le modèle a une dérive  $u_k$ , le seuil est donné par  $\hat{X}_{k-1} + u_k$ .

Etant données ces observations et pour simplifier le problème (avoir une seule forme d'algorithme pour tous les signaux), on posera  $b_k = \hat{X}_{k-1}$ .

Notez que ce choix entraîne une possible perte de performance quand le modèle a une dérive. En réalité, si l'on utilise la prédiction  $\hat{X}_{k-1} + u_k$  au lieu de la dernière estimation, les deux cas, sans et avec dérive, peuvent être traités de façon conjointe (ici on les traitera sans perte de généralité comme étant le cas sans dérive). Le choix  $b_k = \hat{X}_{k-1}$  nous permet d'étudier le comportement d'une approche sous-optimale.

Le schéma général d'estimation est donné en Fig. B.7. L'objectif maintenant est de définir l'algorithme qui sera placé dans le bloc **Mise à jour**.

### Algorithme d'estimation

On utilise comme estimateur l'algorithme adaptatif suivant :

$$\hat{X}_k = \hat{X}_{k-1} + \gamma_k \eta \left[ Q\left(\frac{Y_k - \hat{X}_{k-1}}{\Delta}\right) \right], \quad (\text{B.12})$$

où  $\gamma_k$  est une séquence de gains réels positifs et  $\eta[\cdot]$  est une application de  $\mathcal{I}$  vers  $\mathbb{R}$

$$\begin{aligned} \eta : \mathcal{I} &\rightarrow \mathbb{R} \\ j &\rightarrow \eta_j \end{aligned}$$

caractérisée par  $N_I$  coefficients  $\left\{ \eta_{-\frac{N_I}{2}}, \dots, \eta_{-1}, \eta_1, \dots, \eta_{\frac{N_I}{2}} \right\}$ . Les coefficients  $\eta[\cdot]$  peuvent être vus comme des équivalents pour l'estimation des niveaux de sortie des quantifieurs dans un contexte de quantification classique (quantification pour la reconstruction des mesures).

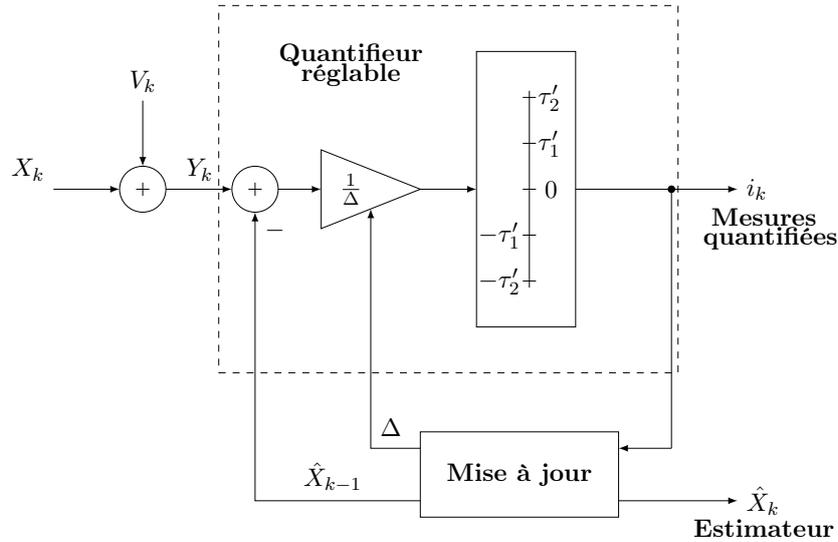


Figure B.7: Schéma général d'estimation. L'algorithme d'estimation est réalisé dans le bloc **Mise à jour**.

Cet algorithme a les avantages d'être un algorithme en ligne, d'avoir une basse complexité et d'inclure comme des cas spéciaux les formes récursives optimales des estimateurs avec quantification adaptative.

A cause de la symétrie du problème et pour simplifier les développements présentés dans la suite, on impose ce qui suit :

**Hypothèse (sur les niveaux de sortie du quantifieur) :**

**AQ3** Les niveaux ont une symétrie impaire en  $i$ :

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

avec  $\eta_i > 0$  pour  $i > 0$ .

La non linéarité non différentiable en (B.12) rend difficile l'analyse directe de l'algorithme. Pour s'en sortir, on peut utiliser les techniques présentées en [Benveniste 1990]. Ces techniques d'analyse sont basées sur des approximations de la moyenne de l'algorithme et sont valables pour une classe assez générale d'algorithmes adaptatifs. Dans le contexte des algorithmes étudiés en [Benveniste 1990], la fonction  $\eta$  peut être une fonction non linéaire et non différentiable et il est montré que les gains  $\gamma_k$  qui optimisent l'estimation de  $X_k$  ont les formes suivantes :

- $\gamma_k \propto \frac{1}{k}$  pour  $X_k$  constant.
- $\gamma_k$  est constant pour  $X_k$  avec un modèle de Wiener.
- $\gamma_k$  est une constante proportionnelle à  $u^{\frac{2}{3}}$  pour  $X_k$  avec un modèle de Wiener qui contient une dérive.

Dans ce qui suit, on utilise en (B.12) les séquences de gains données ci-dessus et on obtient la performance de l'algorithme avec les techniques présentées en [Benveniste 1990].

### Performance d'estimation

L'analyse de la performance de l'algorithme adaptatif est séparée en deux parties : l'analyse de la trajectoire moyenne de l'algorithme et, par conséquent, de son biais et l'analyse de l'EQM ou de la variance asymptotique.

**Analyse de la moyenne : cas constant et Wiener.** Une approximation de la moyenne de l'algorithme  $\mathbb{E}(\hat{X}_k)$  dans le cas constant et Wiener est donnée par  $\hat{x}(t_k)$ , où  $\hat{x}(t)$  est la solution de l'équation différentielle ordinaire (EDO) suivante :

$$\frac{d\hat{x}}{dt} = h(\hat{x}).$$

La correspondance entre temps discret et continu est donnée par la relation  $t_k = \sum_{j=1}^k \gamma_j$  et  $h(\hat{x})$  est

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

où l'espérance est évaluée par rapport à la distribution marginale du bruit  $V$ .

Cette approximation est valable lorsque les gains  $\gamma_k$  sont petits, ce qui veut dire que l'approximation est valable après un certain temps dans le cas constant (vu que les gains décroissent en  $k$ ) et pour tout  $k$  dans le cas Wiener, si on choisit un petit gain  $\gamma_k = \gamma$  (ce qui doit être le cas, car pour poursuivre avec peu d'erreur un signal lentement variable, il faut que les variations de l'estimateur soient petites).

On peut utiliser cette approximation de la moyenne pour obtenir une approximation du biais  $\varepsilon(t)$  :

$$\frac{d\varepsilon}{dt} = \tilde{h}(\varepsilon),$$

où  $\tilde{h}(\varepsilon) = h(\varepsilon + x)$  est une fonction qui ne dépend pas du paramètre  $x$ .

En utilisant les hypothèses de symétrie AN2, AQ2 et AQ3, on peut démontrer que l'EDO est globalement asymptotiquement stable, *i.e.*, pour tout  $\varepsilon(0)$ , on a  $\varepsilon(t) \rightarrow 0$  quand  $t \rightarrow \infty$ . Comme  $\varepsilon(t)$  approche le biais, l'algorithme est donc asymptotiquement non biaisé.

**Analyse de la moyenne : cas Wiener avec dérive.** Pour une dérive  $u$  petite, on s'attend à ce que le gain  $\gamma_k = \gamma$  soit petit aussi (pour suivre le signal qui est lentement variable), dans ce cas on peut aussi utiliser une approximation par EDO. Par contre, contrairement aux cas précédents, on doit prendre en compte la variabilité de la moyenne de  $X_k$ . Ce qui fait que maintenant, la moyenne de l'algorithme est obtenue en échantillonnant la solution de la paire

d'EDO suivante :

$$\begin{aligned}\frac{dx}{dt} &= \frac{u}{\gamma}, \\ \frac{d\hat{x}}{dt} &= \tilde{h}(\hat{x} - x).\end{aligned}$$

Si l'on soustrait les deux expressions, on a une EDO pour le biais

$$\frac{d\varepsilon}{dt} = \tilde{h}(\varepsilon) - \frac{u}{\gamma}. \quad (\text{B.13})$$

La principale différence dans ce cas est le second terme à droite, qui fait que le biais n'est pas asymptotiquement nul. Si l'EDO sans le second terme est globalement asymptotiquement stable, on s'attend à une convergence du biais vers une valeur petite. Pour des petites valeurs de biais, on peut linéariser (B.13) autour de zéro et obtenir une approximation du biais asymptotique. Cette approximation est

$$\mathbb{E} \left( \hat{X}_k - X_k \right)_{k \rightarrow \infty} \approx \frac{u}{\gamma h_\varepsilon}.$$

où  $h_\varepsilon$  est la dérivé de  $\tilde{h}(\varepsilon)$  évaluée en zéro.

**EQM et variance normalisée.** Les résultats de [Benveniste 1990] peuvent être utilisés pour la caractérisation des fluctuations asymptotiques de l'algorithme adaptatif. Les fluctuations asymptotiques dans ce cas sont la variance asymptotique normalisée de l'erreur d'estimation de la constante  $\sigma_\infty^2 = \lim_{k \rightarrow \infty} \text{Var} \left[ \sqrt{k} \left( \hat{X}_k - x \right) \right]$  et l'EQM asymptotique pour l'estimation de  $X_k$  variable EQM $_{q,\infty} = \lim_{k \rightarrow \infty} \mathbb{E} \left[ \left( \hat{X}_k - X_k \right)^2 \right]$ .

Les expressions asymptotiques des fluctuations étant dépendantes de  $\gamma$ , on peut les minimiser par rapport aux gains. Les expressions des paires (gain optimal  $\gamma^*$ , performance optimale) sont données en Tab B.1.

Signal	Gain optimal	Performance
Constant	$\gamma^* = -\frac{1}{h_\varepsilon}$	$\sigma_\infty^2 = \frac{R}{h_\varepsilon^2}$
Wiener	$\gamma^* = \frac{\sigma_w}{\sqrt{R}}$	EQM $_{q,\infty} = \frac{\sigma_w \sqrt{R}}{-h_\varepsilon} + o(\gamma^*) = \sigma_w \sigma_\infty + o(\gamma^*)$
Wiener avec dérive	$\gamma^* = \left( \frac{4u^2}{-h_\varepsilon R} \right)^{\frac{1}{3}}$	EQM $_{q,\infty} \approx 3 \left( \frac{uR}{4h_\varepsilon^2} \right)^{\frac{2}{3}} + o(\gamma^*) = 3 \left( \frac{u}{4} \sigma_\infty^2 \right)^{\frac{2}{3}} + o(\gamma^*)$

Table B.1: Gains optimaux, EQM asymptotique et variance normalisée asymptotique de l'algorithme adaptatif.

La quantité  $R$  dans ce tableau est la variance asymptotique normalisée des corrections de

l'algorithme quand  $\hat{X}_k = X_k$

$$\begin{aligned} R &= \text{Var} \left[ \eta \left( Q \left( \frac{x - \hat{x} + V}{\Delta} \right) \right) \right] \Big|_{\hat{x}=x} \\ &= 2 \sum_{i=1}^{\frac{N_I}{2}} \eta_i^2 \tilde{F}_d(i, 0), \end{aligned}$$

$\tilde{F}_d(i, 0)$  est la probabilité d'avoir la sortie  $i$  du quantifieur aussi quand  $\hat{X}_k = X_k$ .

### Algorithme optimal et performance

Les performances asymptotiques présentées ci-dessus indiquent que la performance de l'algorithme dépend, dans les trois cas, de la quantité  $\sigma_\infty^2$ , qui est une fonction du vecteur de coefficients  $\boldsymbol{\eta} = \left[ \eta_1 \cdots \eta_{\frac{N_I}{2}} \right]^\top$ . Par conséquent, pour maximiser la performance asymptotique, on doit résoudre le problème de maximisation suivant

$$\underset{\boldsymbol{\eta}}{\text{argmin}} \frac{R}{h_\varepsilon^2} = \underset{\boldsymbol{\eta}}{\text{argmin}} \frac{\boldsymbol{\eta}^\top \mathbf{F}_d \boldsymbol{\eta}}{2(\boldsymbol{\eta}^\top \mathbf{f}_d)^2},$$

où  $\mathbf{F}_d$  est une matrice diagonale donnée par

$$\mathbf{F}_d = \text{diag} \left[ \tilde{F}_d(1, 0), \dots, \tilde{F}_d\left(\frac{N_I}{2}, 0\right) \right],$$

et  $\mathbf{f}_d$  est le vecteur des dérivées des probabilités de sortie du quantifieur par rapport à  $\hat{X}_k$  quand  $\hat{X}_k = X_k$

$$\mathbf{f}_d = \left[ \tilde{f}_d(1, 0) \cdots \tilde{f}_d\left(\frac{N_I}{2}, 0\right) \right]^\top.$$

Ce dernier peut être vu comme le vecteur des différences entre les valeurs de la d.d.p. du bruit pour des variations de seuil consécutives  $\tau'_{i-1}$  et  $\tau'_i$ .

Ce problème peut être résolu facilement à l'aide de l'inégalité de Cauchy-Schwarz. En tenant compte de la contrainte de positivité sur les coefficients, on trouve

$$\boldsymbol{\eta}^* = -\mathbf{F}_d^{-1} \mathbf{f}_d.$$

Le  $\sigma_\infty^2$  minimum est donc

$$\sigma_\infty^2 = \frac{1}{2(\mathbf{f}_d^\top \mathbf{F}_d^{-1} \mathbf{f}_d)} = \left[ 2 \sum_{i=1}^{\frac{N_I}{2}} \frac{\tilde{f}_d^2(i, 0)}{\tilde{F}_d(i, 0)} \right]^{-1} = \frac{1}{I_q(0)}.$$

Dans le tableau suivant, on donne les gains et les performances asymptotiques optimales. Notez que dans le cas de l'estimation d'une constante et d'un processus de Wiener lent, l'algorithme a des performances asymptotiques optimales. Il est donc une alternative à basse complexité aux algorithmes vus précédemment (le MV adaptatif et l'estimateur optimal adaptatif).

Signal	Gain optimal	Performance
Constant	$\gamma^* = \frac{1}{I_q(0)}$	$\sigma_\infty^2 = \frac{1}{I_q(0)}$
Wiener	$\gamma^* = \frac{\sigma_w}{\sqrt{I_q(0)}}$	$\text{EQM}_{q,\infty} = \frac{\sigma_w}{\sqrt{I_q(0)}} + o(\sigma_w)$
Wiener avec dérive	$\gamma^* = \left(\frac{4u^2}{I_q^2(0)}\right)^{\frac{1}{3}}$	$\text{EQM}_{q,\infty} \approx 3 \left(\frac{u}{4I_q(0)}\right)^{\frac{2}{3}} + o(\gamma^*)$

Table B.2: Gains optimaux, EQM et variance normalisée asymptotique de l'algorithme adaptatif pour  $\boldsymbol{\eta}$  optimal.

**Choix du gain d'entrée :** pour simplifier le choix de la constante  $\Delta$ , on peut considérer que la fonction de répartition du bruit est caractérisée par un paramètre d'échelle  $\delta$  :

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

où  $F_n$  est la fonction de répartition pour  $\delta = 1$ . Dans ce cas,  $\frac{\Delta}{\delta}$  est un facteur clé pour l'évaluation des coefficients  $\boldsymbol{\eta}$ . Par conséquent, l'évaluation des coefficients peut être simplifiée si on choisit

$$\Delta = c_\Delta \delta.$$

La constante  $c_\Delta$  peut être utilisée pour régler le gain d'entrée du quantifieur ou pour régler le pas de quantification quand les seuils  $\boldsymbol{\tau}'$  sont uniformes et fixés à des valeurs qui ne peuvent pas être modifiées.

**Seuils optimaux.** On voit que, dans les expressions pour les performances de l'algorithme (Tab. B.2), l'influence des variations de seuil  $\boldsymbol{\tau}'$  se fait à travers la quantité  $I_q(0)$ , donc, pour optimiser les performances par rapport aux seuils on doit résoudre le problème d'optimisation suivant :

$$I_q^* = \underset{\boldsymbol{\tau}'}{\operatorname{argmax}} I_q(0).$$

Or, comme on l'a mentionné précédemment, ce problème est difficile à résoudre en général (pour  $N_B > 3$ ) et une approximation de la solution optimale sera présentée dans la Section B.3. Pour les simulations qui seront présentées dans la suite, on imposera les variations de seuil d'être uniformes :

$$\boldsymbol{\tau}' = \left[ -\tau'_{\frac{N_I}{2}} = -\infty \quad \dots \quad -\tau'_1 = -1 \quad 0 \quad +\tau'_1 = +1 \quad \dots \quad +\tau'_{\frac{N_I}{2}} = +\infty \right]^\top.$$

Sous cette contrainte, l'optimisation de la performance est faite par rapport à  $c_\Delta$ . La valeur optimale de  $c_\Delta$  peut être obtenue de façon simple par recherche exhaustive.

**Perte induite par la quantification.** On peut comparer les performances asymptotiques de l'algorithme adaptatif avec les performances asymptotiques de son équivalent qui utilise des mesures continues :

- *Cas constant* : dans ce cas, on compare la performance asymptotique de l'algorithme adaptatif à pas décroissant avec la performance du MV avec des mesures continues –  $\text{Var}(\hat{X}_k) \underset{k \rightarrow \infty}{\sim} \frac{1}{kI_c}$ .
- *Cas Wiener lent* : la performance est comparée avec celle de l'estimateur optimal du processus de Wiener lent avec des mesures continues –  $\text{EQM}_{c,\infty} = \frac{\sigma_w}{\sqrt{I_c}} + o(\sigma_w)$ .
- *Cas Wiener lent avec dérive* : dans ce cas, on compare la performance de l'algorithme adaptatif avec des mesures quantifiées avec celle de l'algorithme adaptatif avec des mesures continues –  $\text{EQM}_{c,\infty} \approx 3 \left( \frac{u}{4I_c(0)} \right)^{\frac{2}{3}} + o\left(u^{\frac{2}{3}}\right)$ .

Les pertes de performance relative en dB sont données en Tab. B.3.

Signal	Perte
Constant	$L_q = -10 \log_{10} \left( \frac{I_q(0)}{I_c} \right)$
Wiener	$L_q^W \approx \frac{1}{2} L_q$ ( $\sigma_w$ petit)
Wiener avec dérive	$L_q^{WD} \approx \frac{2}{3} L_q$ ( $\sigma_w$ et $u$ petits)

Table B.3: Pertes de performance asymptotique induites par la quantification.

Ce qui est surprenant dans ces résultats est le fait que la perte de performance est plus petite dans le cas variable que dans le cas constant, ceci indique une certaine ressemblance avec le phénomène de « dithering », connu en quantification classique. Dans la quantification classique, un ajout de variabilité à l'entrée du quantifieur (ajout de bruit) peut améliorer les performances de reconstruction après quantification. Dans les résultats présentés ci-dessus, on voit que la variabilité intrinsèque au signal induit une perte de performance d'estimation plus petite que dans le cas constant.

### Simulations

**Modèle de bruit** : pour les résultats de simulation présentés dans la suite deux modèles de bruit (respectant les hypothèses de travail) ont été utilisés. Ces modèles sont caractérisés par les distributions suivantes :

- *Gaussienne généralisée (GG)*. Cette distribution a pour d.d.p.

$$f_{GGD}(x) = \frac{\beta}{2\delta\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left|\frac{x}{\delta}\right|^\beta\right),$$

où  $\beta$  est un paramètre de forme (réel et positif).

- *Student-t (ST)*. Sa d.d.p. est

$$f_{STD}(x) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\delta\sqrt{\beta\pi}\Gamma\left(\frac{\beta}{2}\right)} \left[1 + \frac{1}{\beta} \left(\frac{x}{\delta}\right)^2\right]^{-\frac{\beta+1}{2}}.$$

**Perte théorique :** les résultats de simulation de l'algorithme seront comparés aux pertes théoriques qui dépendent toutes de  $L_q$ , l'évolution de cette quantité en fonction de  $N_B$  est donnée en Fig. B.8.

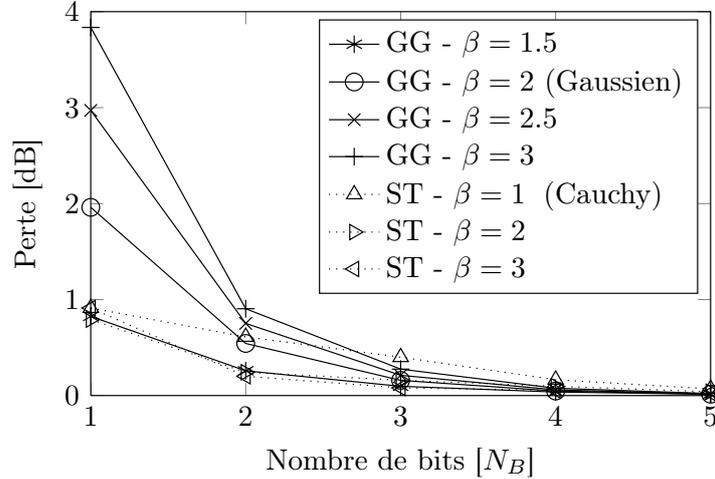


Figure B.8: Perte  $L_q$  induite par la quantification dans le cas constant pour différents nombres de bits et différents types de bruit.

Notez que dans tous les cas, la perte est faible pour la quantification binaire (de 1 à 4 dB) et qu'elle décroît très rapidement avec  $N_B$  pour devenir négligeable pour 4 ou 5 bits de quantification.

**Simulation pour le cas constant :** on vérifie la convergence des pertes simulées pour  $N_B$  de 2 à 5 et plusieurs distributions de bruit dans la Fig. B.9.

**Simulation pour le cas Wiener :** dans la Fig. B.10, on vérifie que si le signal est lent ( $\sigma_w = 0.001$ ), alors les résultats asymptotiques simulés sont très proches des résultats théoriques. Par contre, dès que l'on s'éloigne de l'hypothèse de signal lent ( $\sigma_w = 0.1$ ), les résultats théoriques et simulés ont un certain écart.

**Simulation pour le cas Wiener avec dérive :** la Fig. B.11 montre les performances asymptotiques simulées de l'algorithme adaptatif pour la poursuite du processus de Wiener avec dérive. Le petit écart entre les résultats théoriques et simulés vient du fait que le gain optimal  $\gamma^*$  est calculé avec une estimation en ligne de  $u$  (qui est inconnue en pratique).

**Comparaison avec les algorithmes à haute complexité :** avant de passer aux extensions de l'algorithme adaptatif on discutera rapidement des différences entre l'algorithme adaptatif proposé ici et les algorithmes vus dans les sous sections précédentes.

Etant donnée l'équivalence en termes de performance asymptotique de l'algorithme adaptatif et des solutions à haute complexité (le MV adaptatif et le FP adaptatif), des simulations

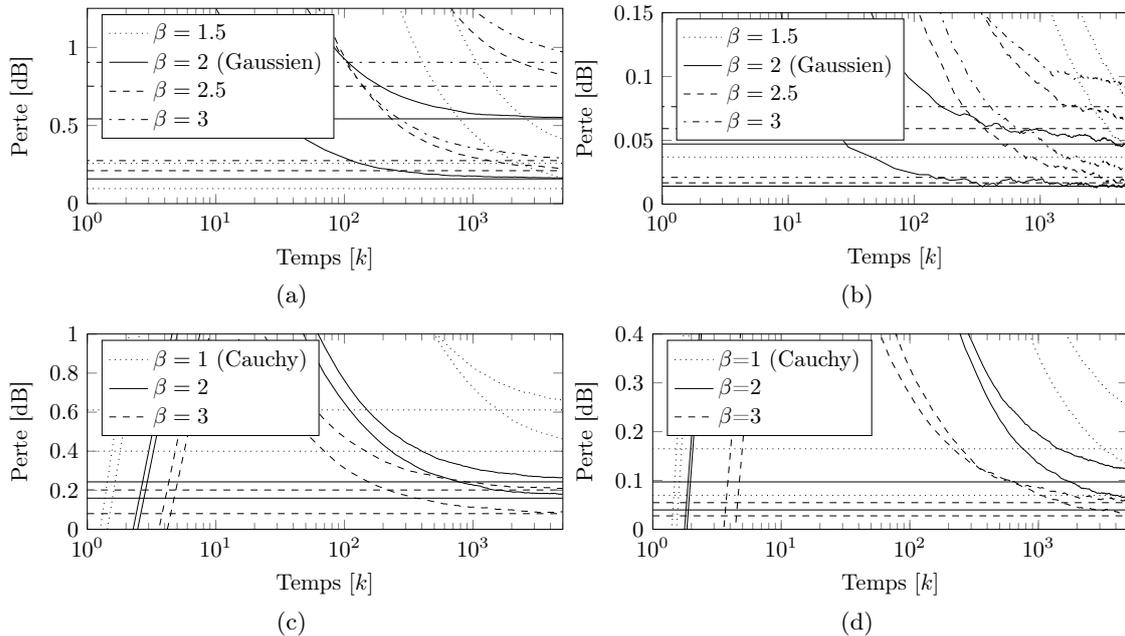


Figure B.9: Perte induite par la quantification pour des distributions GG et ST et pour  $N_B \in \{2, 3, 4, 5\}$  quand  $X_k$  est constant. Pour chaque type de bruit il y a 4 courbes, les constantes sont les résultats théoriques et les courbes décroissantes sont les résultats simulés avec l'algorithme adaptatif. Pour chaque paire de courbes, les résultats plus hauts correspondent à moins de bits de quantification. En (a) on a les résultats pour un bruit GG avec  $N_B = 2$  et 3 et en (b) avec  $N_B = 4$  et 5. En (c) on présente les résultats pour un bruit ST avec  $N_B = 2$  et 3 et en (d) avec  $N_B = 4$  et 5.

des transitoires des algorithmes ont été réalisées pour les différencier de façon plus précise. Les résultats de simulation indiquent que l'algorithme adaptatif peut atteindre des performances similaires voire meilleures que le MV adaptatif pour l'estimation d'une constante et que, dans le cas de la poursuite d'un signal variable, le FP semble être dans la plupart des cas plus performant. En conclusion, pour l'estimation d'une constante, l'algorithme adaptatif semble être la meilleure solution, car il a une complexité très basse en comparaison avec le MV. Toutefois, dans le cas de l'estimation du processus de Wiener, l'algorithme adaptatif ne sera la meilleure solution que si les contraintes de complexité empêchent d'utiliser le FP (aussi très complexe).

### Extensions de l'algorithme adaptatif

**Paramètre d'échelle inconnu :** une extension possible du problème d'estimation d'une constante consiste à considérer que le paramètre d'échelle  $\delta$  est inconnu et donc que l'on doit estimer conjointement la paire  $(x, \delta)$  à partir de mesures quantifiées.

Pour améliorer la performance d'estimation on peut envisager non seulement l'utilisation de  $\hat{X}_{k-1}$  comme biais du quantifieur, mais aussi de  $\hat{\delta}_{k-1}$ , pour régler le gain d'entrée du quantifieur. Ceci est montré en Fig. B.12.

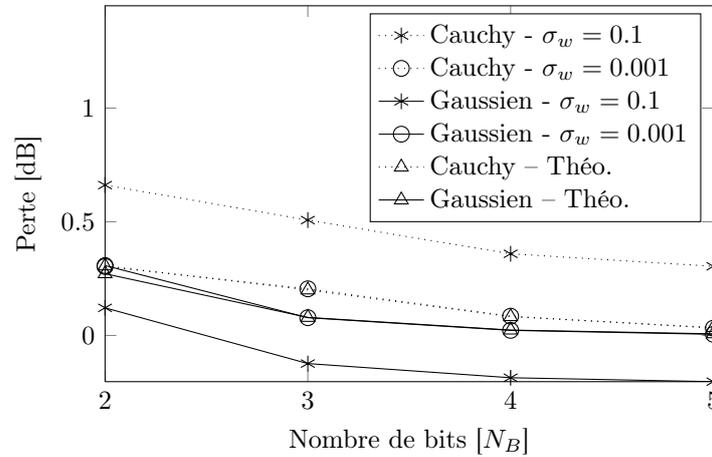


Figure B.10: Perte induite par la quantification dans le cas Wiener pour différents nombres de bits, écarts types des incréments du signal et types de bruits (Gaussien – GG avec  $\beta = 2$  et Cauchy – ST avec  $\beta = 1$ ).

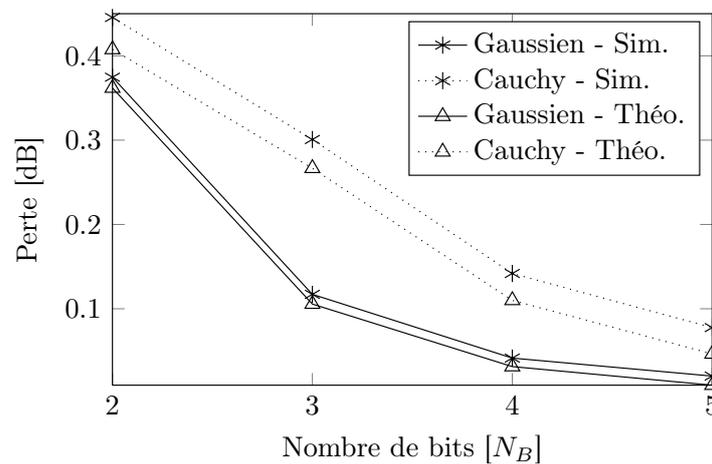


Figure B.11: Perte induite par la quantification dans le cas Wiener avec dérive pour différents nombre de bits et types de bruit (Gaussien et Cauchy).

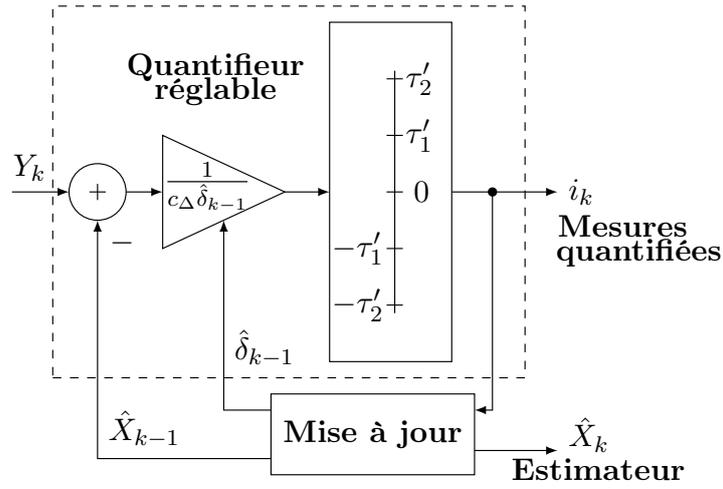


Figure B.12: Schéma d'estimation/quantification pour retrouver conjointement le paramètre de centrage  $x$  et d'échelle  $\delta$ .

Pour la **mise à jour**, on peut encore une fois utiliser un algorithme adaptatif de basse complexité :

$$\begin{bmatrix} \hat{X}_k \\ \hat{\delta}_k \end{bmatrix} = \begin{bmatrix} \hat{X}_{k-1} \\ \hat{\delta}_{k-1} \end{bmatrix} + \frac{\mathbf{\Gamma}}{k} \hat{\delta}_{k-1} \begin{bmatrix} \eta_x(i_k) \\ \eta_\delta(i_k) \end{bmatrix},$$

où  $\mathbf{\Gamma}$  est une matrice  $2 \times 2$  de gains.

Sous certaines hypothèses de convergence en moyenne de l'algorithme, on peut montrer que les coefficients optimaux  $\eta_x$  et  $\eta_\delta$  sont donnés par

$$\begin{aligned} \eta_x &= -\mathbf{F}_d^{-1} \mathbf{f}_d^{(x)}, \\ \eta_\delta &= -\mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)}, \end{aligned}$$

où  $\mathbf{F}_d$  a déjà été détaillée plus haut et où  $\mathbf{f}_d^{(x)}$  et  $\mathbf{f}_d^{(\delta)}$  sont des vecteurs de dérivées des probabilités des sorties du quantifieur par rapport à  $\hat{X}_k$  et  $\hat{\delta}_k$  respectivement. Ces dérivées sont évaluées au point  $(\hat{X}_k = x, \hat{\delta}_k = \delta)$ .

Avec les coefficients optimaux on trouve les valeurs asymptotiques optimales de la covariance normalisée d'estimation  $\mathbf{P}$  et du gain  $\mathbf{\Gamma}^*$  :

$$\mathbf{P} = \delta^2 \mathbf{\Gamma}^* = \frac{\delta^2}{2} \begin{bmatrix} \frac{1}{\mathbf{f}_d^{(x)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(x)}} & 0 \\ 0 & \frac{1}{\mathbf{f}_d^{(\delta)T} \mathbf{F}_d^{-1} \mathbf{f}_d^{(\delta)}} \end{bmatrix}.$$

Les éléments de la diagonale de  $P$  sont les informations de Fisher pour l'estimation de  $x$  et  $\delta$  à partir des données quantifiées, l'algorithme est donc optimal asymptotiquement et on voit que le fait de ne pas connaître  $\delta$  ne dégrade pas les performances asymptotiques de l'estimateur de  $x$ .

**Approche multi capteur :** une autre extension consiste à utiliser des mesures obtenues de façon simultanée par plusieurs capteurs pour estimer un paramètre constant  $x$ .

Dans cette approche, chaque capteur quantifie une mesure continue

$$Y_k^{(j)} = x + V_k^{(j)}, \quad \text{for } j \in \{1, \dots, N_s\},$$

où  $N_s$  est le nombre de capteurs, et transmet la mesure quantifiée à un centre de fusion. Le centre de fusion utilise toutes les mesures des capteurs pour générer une estimation  $\hat{X}_k$  de  $x$  qui est diffusée à tous les capteurs pour être utilisée comme seuil central des quantifieurs. Le schéma qui représente cette approche est montré en Fig. B.13.

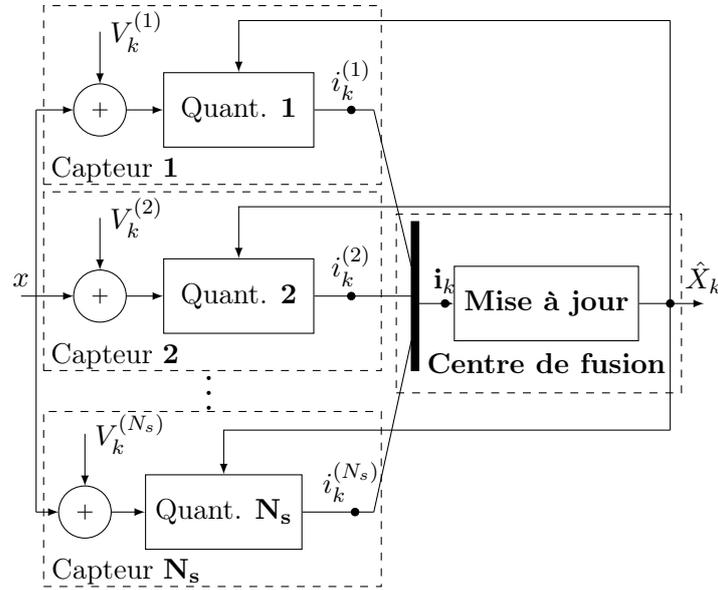


Figure B.13: Schéma d'estimation/quantification multicapteur avec un centre de fusion.

Pour la **mise à jour**, on peut utiliser l'extension suivante de l'algorithme adaptatif :

$$\hat{X}_k = \hat{X}_{k-1} + \frac{\gamma}{k} \eta(\mathbf{i}_k),$$

où  $\mathbf{i}_k$  est le vecteur des mesures quantifiées  $[i_k^{(1)} \dots i_k^{(N_s)}]^T$ .

Les coefficients  $\eta$  optimaux sont donnés cette fois par

$$\eta(\mathbf{i}) = - \sum_{j=1}^{N_s} \frac{\tilde{f}_d^{(j)}[i^{(j)}]}{\tilde{F}_d^{(j)}[i^{(j)}]}. \quad (\text{B.14})$$

Pour ces coefficients, on trouve la variance asymptotique d'estimation normalisée et le gain optimal suivants

$$\sigma_\infty^2 = \gamma^* = \frac{1}{\sum_{j=1}^{N_s} \sum_{i^{(j)} \in \mathcal{I}^{(j)}} \frac{\tilde{f}_d^{(j)2}[i^{(j)}]}{\tilde{F}_d^{(j)}[i^{(j)}]}}. \quad (\text{B.15})$$

Les expressions (B.14) et (B.15) nous montrent que l'algorithme adaptatif pour l'approche à un seul capteur s'étend de façon très naturelle à l'approche multicapteur : le coefficient de l'algorithme  $\eta$  en multicapteur est la somme des coefficients en monocapteur et la performance ainsi que le gain sont donnés par l'inverse de la somme des informations de Fisher en monocapteur.

## B.3 Estimation et quantification : approximations à haute résolution

On présente maintenant des résultats concernant la caractérisation asymptotique des quantifieurs optimaux pour l'estimation. Le mot « asymptotique » dans ce cas vient du fait que l'on suppose que le nombre d'intervalles de quantification  $N_I$  est très grand. Comme on impose aussi que les tailles  $\Delta_i$  des intervalles de quantification tendent vers zéro, on les appelle aussi approximations à haute résolution.

### B.3.1 Approximation à haute résolution de l'information de Fisher

Pour trouver la caractérisation asymptotique des quantifieurs optimaux et la performance correspondante en termes d'estimation, on s'intéresse aux questions suivantes :

- comment décrire l'information de Fisher pour l'estimation d'un paramètre  $x$  en fonction du quantifieur quand  $N_I$  est grand ?
- Comment maximiser l'information de Fisher par rapport à la caractérisation du quantifieur ?
- Quelle est la performance optimale correspondante ?

**Remarque :** notez que dans la suite on n'impose pas que le problème soit un problème d'estimation de paramètre de centrage.

#### Approximation asymptotique

Pour répondre à la première question, on va commencer par une réécriture de l'information de Fisher :

$$I_q = I_c - \mathbb{E} \left[ (S_c - S_q)^2 \right]. \quad (\text{B.16})$$

Le membre de droite dans (B.16) peut être vu comme la perte  $L$  induite par la quantification. L'espérance en  $L$  peut être écrite comme une somme d'intégrales sur les différents intervalles de quantification  $q_i$  :

$$L = \sum_{i=1}^{N_I} \int_{q_i} \left[ \frac{\partial \log f(y; x)}{\partial x} - \frac{\partial \log \mathbb{P}(i; x)}{\partial x} \right]^2 f(y; x) dy.$$

Des développements en série de Taylor nous donnent

$$L = \sum_{i=1}^{N_I} \left\{ \left( S_{c,i}^{(y)} \right)^2 f_i \frac{\Delta_i^3}{12} + o(\Delta_i^3) \right\}, \quad (\text{B.17})$$

où  $S_{c,i}^{(y)}$  est la dérivée du score par rapport à  $y$  évaluée au centre de l'intervalle de quantification  $q_i$  et  $f_i$  est la d.d.p. des mesures continues aussi évaluée au centre de l'intervalle.

Pour obtenir la caractérisation de la perte en fonction du quantifieur, on définit la densité d'intervalles  $\lambda$  :

$$\lambda(y) = \lambda_i = \frac{1}{N_I \Delta_i}, \quad \text{pour } y \in q_i. \quad (\text{B.18})$$

La densité d'intervalles est une fonction qui, si on l'intègre dans un intervalle donné, donne la fraction d'intervalles de quantification dans cet intervalle.

Si l'on utilise (B.18) dans (B.17), si l'on fait  $N_I \rightarrow \infty$  et si les  $\Delta_i$  convergent uniformément vers zéro, on obtient le résultat suivant

$$\lim_{N_I \rightarrow \infty} N_I^2 L = \frac{1}{12} \int \frac{\left( \frac{\partial S_c(y;x)}{\partial y} \right)^2 f(y;x)}{\lambda^2(y)} dy.$$

Si on revient à (B.16), le résultat asymptotique ci-dessus nous amène à l'approximation asymptotique de l'information de Fisher :

$$I_q \approx I_c - \frac{1}{12 N_I^2} \int \frac{\left( \frac{\partial S_c(y;x)}{\partial y} \right)^2 f(y;x)}{\lambda^2(y)} dy. \quad (\text{B.19})$$

Celle-ci est la réponse de la première question. On constate que si l'intégrale du membre à droite converge, alors  $I_q$  converge vers  $I_c$  quand  $N_I \rightarrow \infty$  et, de cette façon, on répond aussi à une question qui avait été posée précédemment (p. 265).

Si toutes les valeurs possibles en sortie du quantifieur sont quantifiées avec des mots binaires de même taille  $N_B = \log_2(N_I)$ , alors (B.19) peut être réécrite de la façon suivante :

$$I_q \approx I_c - \frac{2^{-2N_B}}{12} \int \frac{\left( \frac{\partial S_c(y;x)}{\partial y} \right)^2 f(y;x)}{\lambda^2(y)} dy.$$

On voit de façon explicite la convergence exponentielle en  $N_B$  de  $I_q$  vers  $I_c$ .

**Densité d'intervalles optimale :** maintenant on répond à la deuxième question.

L'expression (B.19) nous montre directement que, pour maximiser la performance par rapport au quantifieur, on doit minimiser l'intégrale de droite par rapport à  $\lambda$ . Ce problème de minimisation peut être facilement résolu avec l'inégalité de Hölder, ce qui donne

$$\lambda^*(y) = \frac{\left( \frac{\partial S_c(y;x)}{\partial y} \right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x)}{\int \left( \frac{\partial S_c(y;x)}{\partial y} \right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x) dy} \propto \left( \frac{\partial S_c(y;x)}{\partial y} \right)^{\frac{2}{3}} f^{\frac{1}{3}}(y;x). \quad (\text{B.20})$$

Notez que, contrairement aux résultats asymptotiques pour la reconstruction des mesures où  $\lambda^*(y) \propto f^{\frac{1}{3}}(y; x)$ , en quantification optimale pour l'estimation le score du problème d'estimation intervient sur la densité d'intervalles.

Si l'on remplace (B.20) en (B.19), on peut donner une réponse à la troisième question. L'expression analytique de l'approximation asymptotique de l'information de Fisher optimale est

$$I_q^* \approx I_c - \frac{1}{12N_I^2} \left[ \int \left( \frac{\partial S_c(y; x)}{\partial y} \right)^{\frac{2}{3}} f^{\frac{1}{3}}(y; x) dy \right]^3. \quad (\text{B.21})$$

**Approximation pratique des seuils optimaux :** la définition de la densité d'intervalles nous dit que le pourcentage d'intervalles jusqu'à l'intervalle  $q_i, \frac{i}{N_I}$  doit être égal à l'intégrale de la densité d'intervalles jusqu'à  $\tau_i$ . Par conséquent, une approximation pratique des seuils optimaux est donnée par

$$\tau_i^* = F_\lambda^{-1} \left( \frac{i}{N_I} \right), \quad \text{pour } i \in \{1, \dots, N_I - 1\}, \quad (\text{B.22})$$

où  $F_\lambda^{-1}$  est l'inverse de la fonction de répartition obtenue par intégration de la densité d'intervalles  $\lambda$ .

**Remarque sur la solution à débit variable :** on pourrait aussi considérer que les sorties du quantifieur sont encodées avec des mots de tailles égales au logarithme de leur probabilité, ceci entraînerait une possible réduction de la taille moyenne des mots en sortie du quantifieur. Cette solution est connue sous le nom d'encodage à débit variable.

La taille moyenne des mots en sortie du quantifieur avec l'encodage à débit variable est donnée par l'entropie des mots de sortie. De la même manière que précédemment, où en imposant un  $N_B$  on a trouvé la densité d'intervalles optimale pour des mots de sortie de taille égale, on peut s'intéresser au problème de quantification optimale avec encodage à débit variable. Si l'on impose un débit moyen  $R$ , on peut montrer que la densité optimale est

$$\lambda^*(y) = \frac{\left| \frac{\partial S_c(y; x)}{\partial y} \right|}{\int \left| \frac{\partial S_c(y; x)}{\partial y} \right| dy}$$

et l'information de Fisher maximale

$$I_q \approx I_c - \frac{1}{12} 2^{-2} \left\{ R - h_y - \int \log_2 \left[ \left| \frac{\partial S_c(y; x)}{\partial y} \right| \right] f(y; x) dy \right\},$$

où  $h_y = - \int f(y; x) \log_2 [f(y; x)] dy$  est l'entropie différentielle des mesures.

Le problème avec cette solution est que l'encodage des sorties du quantifieur dépend du paramètre qui est inconnu. Même si on utilise une approche adaptative pour la quantification avec une convergence vers l'encodage optimal, on ne respectera pas les contraintes de débit moyen pendant toute la phase de convergence de l'algorithme adaptatif.

**Application à l'estimation d'un paramètre de centrage :** pour la distribution Gaussienne, la densité d'intervalles optimale et l'approximation de l'information de Fisher maximale sont données par

$$\lambda_G^x(y) = \frac{1}{\delta\sqrt{3\pi}} \exp\left[-\left(\frac{y-x}{\sqrt{3}\delta}\right)^2\right], \quad I_{q,G}^x \approx \frac{2}{\delta^2} \left[1 - \pi\sqrt{3}2^{-(2N_B-1)}\right]. \quad (\text{B.23})$$

Pour la distribution de Cauchy on a

$$\lambda_C^x(y) = \frac{1}{\delta B\left(\frac{1}{2}; \frac{5}{6}\right)} \frac{\left[1 - \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{2}{3}}}{\left[1 + \left(\frac{y-x}{\delta}\right)^2\right]^{\frac{5}{3}}}, \quad I_{q,C}^x \approx \frac{1}{2\delta^2} \left[1 - \frac{B\left(\frac{1}{2}; \frac{5}{6}\right)^3}{3\pi} 2^{-2N_B+1}\right]. \quad (\text{B.24})$$

Afin de valider les résultats théoriques, l'information de Fisher (B.3) a été évaluée avec  $\delta = 1$  pour les deux distributions et pour

- les seuils optimaux pour  $N_B \in \{1, 2, 3\}$ . Les seuils optimaux ont été obtenus par recherche exhaustive. Pour  $N_B \in \{4, 5, 6, 7, 8\}$  les résultats théoriques (B.23) et (B.24) sont utilisés comme une approximation.
- la quantification uniforme pour  $N_B \in \{1, \dots, 8\}$ . En plaçant le seuil central sur  $x$ , l'intervalle de quantification optimal  $\Delta^*$  est obtenu par maximisation de l'information de Fisher. Dans ce cas aussi, le maximum est trouvé par recherche exhaustive.
- l'approximation pratique des seuils optimaux donnée par (B.22), pour  $N_B \in \{1, \dots, 8\}$ .

Les résultats sont montrés en Tab. B.4.

$N_B$	Gaussien ( $I_{c,n}^x = 2$ )			Cauchy ( $I_{c,n}^x = 0.5$ )		
	Optimal	Uniforme	Approx. pratique	Optimal	Uniforme	Approx. pratique
1	1.27323954 <sup>†</sup>	–	1.27323954	0.40528473 <sup>†</sup>	–	0.40528473
2	1.76503630 <sup>†</sup>	1.76503630	1.75128300	0.43433896 <sup>†</sup>	0.43433896	0.40528473
3	1.93090199 <sup>†</sup>	1.92837814	1.92740111	0.48474865 <sup>†</sup>	0.45600797	0.47893785
4	1.97874454 <sup>*</sup>	1.97841622	1.98038526	0.49533850 <sup>*</sup>	0.48136612	0.49504170
5	1.99468613 <sup>*</sup>	1.99353005	1.99489906	0.49883463 <sup>*</sup>	0.49204506	0.49879785
6	1.99867153 <sup>*</sup>	1.99807736	1.99869886	0.49970866 <sup>*</sup>	0.49656712	0.49970408
7	1.99966788 <sup>*</sup>	1.99943563	1.99967136	0.49992716 <sup>*</sup>	0.49851056	0.49992659
8	1.99991697 <sup>*</sup>	1.99983649	1.99991741	0.49998179 <sup>*</sup>	0.49935225	0.49998172

Table B.4: Information de Fisher  $I_q$  pour l'estimation d'un paramètre de centrage des distributions Gaussienne et Cauchy. En **Optimal**<sup>†</sup> se trouve l'information de Fisher maximale obtenue par recherche exhaustive des seuils optimaux. **Optimal**<sup>\*</sup> est l'approximation asymptotique de l'information de Fisher maximale. En **Uniforme**, les valeurs de l'information de Fisher pour la quantification uniforme optimale sont montrées. Les colonnes **Approx. pratique** correspondent à l'information de Fisher obtenue avec l'approximation pratique des seuils asymptotiquement optimaux.

On constate que, dans tous les cas,  $I_q$  converge rapidement vers  $I_c$  quand  $N_B$  augmente.

Ici encore, on voit que 4 ou 5 bits sont suffisants pour obtenir une performance d'estimation proche de celle obtenue avec des mesures continues. La différence de performance entre la quantification uniforme et non uniforme semble être plus importante pour la distribution de Cauchy, mais pratiquement négligeable dans le cas Gaussien, ceci indique que la quantification uniforme est probablement une meilleure solution en pratique (étant donnée sa simplicité d'implantation). Finalement, on observe aussi que l'approximation asymptotique de l'information de Fisher et sa valeur obtenue avec l'approximation pratique des seuils optimaux sont très proches, même pour des valeurs petites de  $N_B$  ( $N_B = 4$ ).

**Utilisation de l'algorithme adaptatif :** pour la réalisation pratique du quantifieur optimal dans l'estimation d'un paramètre de centrage, un problème important est la dépendance explicite au paramètre  $x$  de l'approximation pratique des seuils optimaux  $\tau_i^*$ . Une solution pour résoudre ce problème et atteindre une performance asymptotique optimale, même en ne connaissant pas  $x$ , consiste à utiliser l'algorithme adaptatif proposé en Sous-section B.2.3 :

$$\hat{X}_k = \hat{X}_{k-1} + \frac{1}{kI_q} \eta(i_k),$$

avec le vecteur de variations de seuil  $\tau'$  donné par  $\tau^*$  avec  $x$  en (B.22) considéré comme étant égal à zéro et  $\eta(i_k)$  donnés par  $\eta(i) = -\frac{f(\tau_{i-1}^*;x) - f(\tau_i^*;x)}{F(\tau_i^*;x) - F(\tau_{i-1}^*;x)}$ . Si  $N_B \geq 4$ , pour  $k$  grand, on s'attend à une performance d'estimation proche de l'optimale et, par conséquent, proche de l'approximation suivante :

$$\text{Var} [\hat{X}_k] \approx \text{BCR}_q \approx \frac{1}{kI_q},$$

où  $I_q$  est l'approximation asymptotique (B.21).

Les résultats de simulation pour des distributions de Gauss et de Cauchy avec  $N_B = 4$  et 5 indiquent la validité cette approche. Ils sont montrés en Fig. B.14.

**Allocation de bits pour l'estimation d'un paramètre de centrage scalaire**

On suppose maintenant que  $N_s$  capteurs mesurent, avec du bruit additif et indépendant d'un capteur à l'autre, une constante  $x$ . En raison des contraintes de communication, la somme des nombres de bits par mesure alloués à chaque capteur  $N_{B,i}$  est contrainte à être égale à une valeur  $N_B$ . La question que l'on se pose est la suivante : quelle est l'allocation de bits qui maximise la performance d'estimation sous la contrainte de communication ?

Ceci équivaut de façon quantitative à résoudre le problème de maximisation suivant :

$$\begin{array}{ll} \text{maximiser} & I_q = \sum_{i=1}^{N_s} I_{q,i}(N_{B,i}), \\ \text{en } N_{B,i} & \\ \text{sujet à} & \sum_{i=1}^{N_s} N_{B,i} = N_B, \\ & N_{B,i} \in \mathbb{N}, \end{array}$$

où  $I_{q,i}(N_{B,i})$  est l'information de Fisher maximale pour  $N_{B,i}$ .

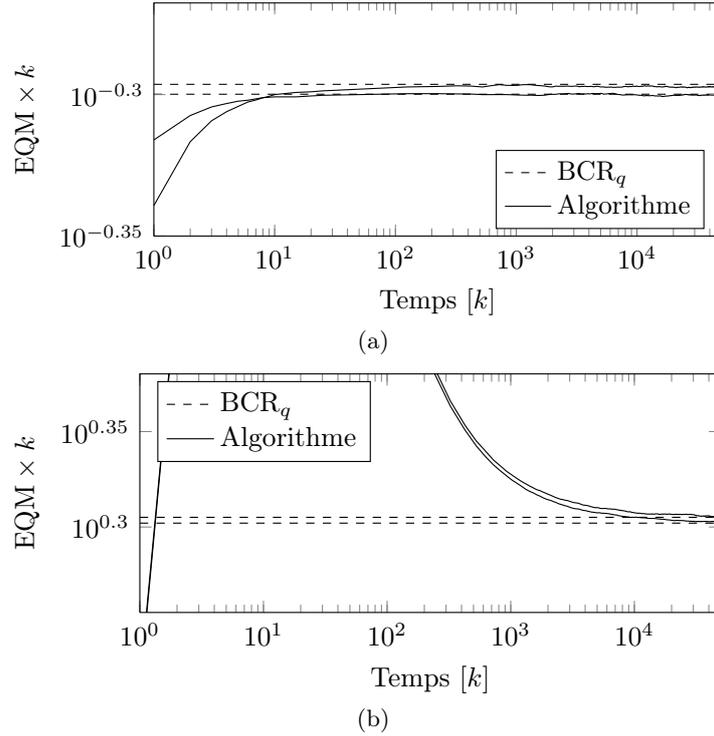


Figure B.14: EQM simulée pour l'algorithme adaptatif avec les seuils non uniformes asymptotiquement optimaux. Les mesures continues sont distribuées selon la loi de Gauss (a) et de Cauchy (b). Les nombres de bits de quantification utilisés sont  $N_B = 4$  et 5. Les courbes qui ont des valeurs asymptotiques plus hautes correspondent à  $N_B = 4$ .

On peut trouver la solution analytique de ce problème par la comparaison de toutes les combinaisons possibles des  $N_{B,i}$ . Cependant, l'aspect combinatoire de la solution rend impossible en pratique l'application de cette solution, même pour quelques dizaines de capteurs.

Une autre solution possible consiste à utiliser les expressions asymptotiques analytiques des informations de Fisher en levant la contrainte  $N_{B,i} \in \mathbb{N}$ . La solution du problème d'optimisation sous ces nouvelles conditions peut être trouvée sous forme analytique et, si on arrondi les  $N_{B,i}$  trouvés de cette manière, on a une approximation pratique de la solution.

**Solution à  $N_{B,i}$  réels :** si l'on considère que les distributions de bruit ont la même forme mais des paramètres d'échelle  $\delta_i$  différents, alors, en utilisant les approximations asymptotiques on trouve

$$N_{B,i} = \frac{N_B}{N_s} - \log_2 \left( \frac{\delta_i}{\sqrt[N_s]{\prod_{j=1}^{N_s} \delta_j}} \right).$$

On voit que les  $N_{B,i}$  optimaux ne dépendent que des paramètres d'échelle des bruits.

L'approximation de l'information de Fisher dans ce cas est

$$I_q \approx N_s \left[ \frac{I_{c,n}^x}{HM(\delta_1^2, \dots, \delta_{N_s}^2)} - \frac{\kappa'(f_n)}{12} \frac{2^{-2\bar{N}_B}}{GM(\delta_1^2, \dots, \delta_{N_s}^2)} \right],$$

où  $I_{c,n}^x$  est l'information de Fisher pour un paramètre d'échelle unitaire,  $\kappa'(f_n)$  est une fonctionnelle de la d.d.p. du bruit aussi pour  $\delta = 1$ ,  $\bar{N}_B = \frac{N_B}{N_s}$  et  $HM(\delta_1^2, \dots, \delta_{N_s}^2)$  et  $GM(\delta_1^2, \dots, \delta_{N_s}^2)$  sont les moyennes harmoniques et géométriques des paramètres d'échelle.

On peut démontrer que l'approximation de  $I_q$  optimale ainsi obtenue est toujours plus grande que celle donnée par une allocation de bits uniforme.

**Solution à  $N_{B,i}$  réels et positifs :** si l'on contraint les  $N_{B,i}$  à être positifs, on peut démontrer que la solution optimale est obtenue en deux étapes. D'abord on choisit un  $\nu$  qui satisfait

$$\sum_{i=1}^{N_s} [\nu - \log_2(\delta_i)]_+ = N_B,$$

où  $[x]_+ = \max(x, 0)$ , puis, on obtient les  $N_{B,i}$  avec

$$N_{B,i} = [\nu - \log_2(\delta_i)]_+.$$

On peut facilement vérifier que cette solution est équivalente à la procédure de « water-filling » qui est utilisée pour l'allocation de puissance aux sous porteuses dans les modulations multi porteuses.

## B.4 Conclusions

Dans cette thèse, nous avons traité le problème d'estimation à partir de mesures quantifiées, un problème qui attire depuis quelque temps l'attention de la communauté de traitement du signal, en raison de l'essor des réseaux de capteurs. Nous avons traité, plus spécifiquement, le problème d'estimation d'un paramètre de centrage scalaire, soit constant, soit lentement variable avec un modèle de Wiener.

### B.4.1 Conclusions principales

Nous avons observé que, pour la plupart des modèles de bruit considérés en pratique, la performance d'estimation se dégrade lorsque la dynamique de quantification est loin de la vraie valeur du paramètre. Ceci indique qu'une bonne performance d'estimation peut être obtenue par une approche adaptative, où l'on place la dynamique de quantification grâce à l'information donnée par l'estimation la plus récente du paramètre.

Avec le schéma adaptatif, nous avons vu que la perte de performance d'estimation induite par la quantification est petite. Pour tous les cas testés, nous avons observé une perte de performance petite pour 1–3 bits de quantification et une perte négligeable pour 4 ou 5 bits.

Ceci indique que dans un contexte d'estimation à distance, où le nombre de bits total est contraint, il est possible qu'une solution multicapteur/basse résolution soit préférable à la solution classique monocapteur/haute résolution.

Nous avons proposé des alternatives à basse complexité pour les algorithmes trouvés dans la littérature et leurs extensions. Nous avons démontré que les algorithmes à basse complexité proposés atteignent les mêmes performances asymptotiques que leurs pendants à haute complexité. En utilisant les approches à basse complexité, nous avons présenté des solutions assez naturelles pour traiter des extensions du problème de base : l'extension au cas d'un paramètre d'échelle inconnu et l'extension à plusieurs capteurs.

Pour traiter le problème de placement des seuils optimaux pour l'estimation quand un grand nombre d'intervalles de quantification est utilisé, nous avons étudié une approche asymptotique. Cette approche asymptotique nous a permis d'obtenir une approximation pratique des seuils optimaux ainsi qu'une expression analytique de la performance d'estimation optimale, dans ce cas l'information de Fisher optimale. Nous avons vu aussi avec cette approche que la performance d'estimation avec des mesures quantifiées converge exponentiellement vite vers la performance avec des mesures continues quand le nombre de bits de quantification augmente. En appliquant les résultats sur un problème d'estimation de paramètre de centrage, nous avons constaté que l'approximation proposée, censée être valable seulement asymptotiquement, est valable pour un nombre petit de bits (4 dans ce cas), ceci indique que les résultats asymptotiques peuvent être utilisés en pratique.

Nous avons montré, avec l'approche asymptotique, la dépendance des seuils asymptotiquement optimaux par rapport au paramètre inconnu. Ceci indique, encore une fois, l'importance de l'approche adaptative qui permet de placer asymptotiquement les seuils sur leurs valeurs optimales et donc d'obtenir une performance asymptotiquement optimale.

Nous voudrions aussi attirer l'attention sur le fait que la différence de performance entre un schéma de quantification uniforme et un schéma non uniforme semble être petite pour l'estimation d'un paramètre de centrage. Par conséquent, en pratique, si une forte contrainte sur la complexité est présente, la quantification uniforme peut être préférable.

### B.4.2 Perspectives

Cette « discussion » entre quantification et estimation sera terminée par la présentation des possibles sujets de travaux futurs.

- *Paramètre vectoriel et quantification vectorielle* : ce sujet est une extension naturelle du problème. Tandis que l'extension à la quantification vectorielle est assez directe (en termes d'algorithmes d'estimation et de leurs performances asymptotiques), l'extension aux paramètres vectoriels est moins directe car elle nécessitera une nouvelle définition pour la performance d'estimation et un changement de la structure des algorithmes pour prendre en compte les corrélations possibles entre les composantes vectorielles.
- *Canaux bruités* : un canal de communication bruité peut être intégré au problème de différentes façons. La plus simple consistant à introduire un indice binaire pour chaque

mesure quantifiée et un modèle de canal binaire symétrique. Avec un étiquetage fixe des mesures quantifiées, des extensions de l'algorithme de basse complexité proposé peuvent être directement conçues. Cependant, si les indices ne sont pas fixés, le problème qui en résulte, avec l'étiquetage des sorties du quantifieur, peut être très difficile à traiter.

D'autres extensions peuvent être envisagées par l'introduction d'un canal à amplitude continue, par exemple des canaux à bruit additif et à évanouissements. Dans ce cas, on sera obligé, encore une fois, d'ajouter le problème d'étiquetage et donc de traiter un problème conjoint de conception d'encodeur/estimation.

- *Estimation avec distribution de bruit inconnue*: on a supposé depuis le début que la distribution du bruit est connue, en pratique ceci ne sera pas toujours le cas et on sera obligé de trouver d'autres approches d'estimation.
- *Variations rapides* : dans certaines parties de cette thèse nous avons supposé que le signal est lentement variable. Sous cette hypothèse, nous avons vu que la perte de performance induite par la quantification est petite. On peut se poser la question de savoir si cette conclusion reste vraie lorsque le signal varie rapidement.
- *Problème distribué* : pour arriver aux applications classiques des réseaux de capteur, on doit généraliser les algorithmes et résultats obtenus ici pour un capteur à un contexte partiellement distribué, où certains capteurs recueillent l'information des capteurs qui sont autour, ou complètement distribué, où tous les capteurs recueillent l'information.
- *Temps continu* : dans le cas d'un paramètre variable, nous avons considéré, depuis le début, que le temps est discret et nous n'avons pas traité de l'échantillonnage. Un sujet qui reste ouvert est donc l'estimation d'un signal à temps continu, échantillonné et quantifié.



# Bibliography

- [Akyildiz 2002] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci. *A survey on sensor networks*. Communications magazine, IEEE, vol. 40, no. 8, pages 102–114, 2002. (Cited in page(s) 17, 19 and 254.)
- [Ali 1966] S.M. Ali and S.D. Silvey. *A general class of coefficients of divergence of one distribution from another*. Journal of the Royal Statistical Society. Series B (Methodological), pages 131–142, 1966. (Cited in page(s) 207.)
- [Arampatzis 2005] T. Arampatzis, J. Lygeros and S. Manesis. *A survey of applications of wireless sensors and wireless sensor networks*. In Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation, pages 719–724. Ieee, 2005. (Cited in page(s) 18 and 255.)
- [Aysal 2008] T.C. Aysal and K.E. Barner. *Constrained decentralized estimation over noisy channels for sensor networks*. Signal Processing, IEEE Transactions on, vol. 56, no. 4, pages 1398–1410, 2008. (Cited in page(s) 20 and 256.)
- [Bailey 1994] R.W. Bailey. *Polar Generation of Random Variates with the t-Distribution*. Mathematics of Computation, pages 779–781, 1994. (Cited in page(s) 252.)
- [Baker 2004] E.T. Baker and C.R. German. *On the global distribution of hydrothermal vent fields*. Mid-ocean ridges: hydrothermal interactions between the lithosphere and oceans, vol. 148, pages 245–266, 2004. (Cited in page(s) 173.)
- [Benitz 1989] G.R. Benitz and J.A. Bucklew. *Asymptotically optimal quantizers for detection of iid data*. Information Theory, IEEE Transactions on, vol. 35, no. 2, pages 316–325, 1989. (Cited in page(s) 19 and 256.)
- [Benveniste 1990] A. Benveniste, M. Métivier and P. Priouret. *Adaptive algorithms and stochastic approximations*. Springer-Verlag New York, Inc., 1990. (Cited in page(s) 113, 114, 115, 121, 122, 123, 136, 150, 151, 152, 158, 159, 166, 276 and 278.)
- [Berzuni 1997] C. Berzuni, N.G. Best, W.R. Gilks and C. Larizza. *Dynamic conditional independence models and Markov chain Monte Carlo methods*. Journal of the American Statistical Association, vol. 92, no. 440, pages 1403–1412, 1997. (Cited in page(s) 86.)
- [Blahut 1987] R.E. Blahut. *Principles and practice of information theory*. Addison-Wesley Longman Publishing Co., Inc., 1987. (Cited in page(s) 208.)
- [Borkar 1995] V.S. Borkar, S.K. Mitter *et al.* *LQG control with communication constraints*. Technical report, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 1995. (Cited in page(s) 95.)
- [Box 1958] G.E.P. Box and M.E. Muller. *A note on the generation of random normal deviates*. The Annals of Mathematical Statistics, vol. 29, no. 2, pages 610–611, 1958. (Cited in page(s) 248 and 250.)

- [Boyd 2004] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, New York, NY, USA, 2004. (Cited in page(s) 57, 58 and 59.)
- [Chong 2003] C.Y. Chong and S.P. Kumar. *Sensor networks: evolution, opportunities, and challenges*. Proceedings of the IEEE, vol. 91, no. 8, pages 1247–1256, 2003. (Cited in page(s) 18 and 255.)
- [Costa 2003] J. Costa, A. Hero and C. Vignat. *On Solutions to Multivariate Maximum  $\alpha$ -Entropy Problems*. In Anand Rangarajan, Mário Figueiredo and Josiane Zerubia, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2683 of *Lecture Notes in Computer Science*, pages 211–226. Springer Berlin/Heidelberg, 2003. (Cited in page(s) 136.)
- [Cover 2006] T.M. Cover and J.A. Thomas. *Elements of information theory* 2<sup>nd</sup> edition. Wiley-Interscience, 2006. (Cited in page(s) 136, 187 and 188.)
- [Crisan 2000] D. Crisan and A. Doucet. *Convergence of sequential Monte Carlo methods*. Technical report, Signal Processing Group, Department of Engineering, University of Cambridge, 2000. (Cited in page(s) 91.)
- [Crowder 1976] M.J. Crowder. *Maximum likelihood estimation for dependent observations*. Journal of the Royal Statistical Society. Series B (Methodological), pages 45–53, 1976. (Cited in page(s) 62, 67 and 70.)
- [Curry 1970] R. Curry, W.V. Velde and J. Potter. *Nonlinear estimation with quantized measurements—PCM, predictive quantization, and data compression*. Information Theory, IEEE Transactions on, vol. 16, no. 2, pages 152–161, March 1970. (Cited in page(s) 95.)
- [Doucet 1998] A. Doucet *et al.* *On sequential simulation-based methods for Bayesian filtering*. Technical report, 1998. (Cited in page(s) 83, 84, 85 and 86.)
- [Doucet 2000] A. Doucet, S. Godsill and C. Andrieu. *On sequential Monte Carlo sampling methods for Bayesian filtering*. Statistics and computing, vol. 10, no. 3, pages 197–208, 2000. (Cited in page(s) 246.)
- [Durisic 2012] M.P. Durisic, Z. Tafa, G. Dimic and V. Milutinovic. *A survey of military applications of wireless sensor networks*. In *Embedded Computing (MECO), 2012 Mediterranean Conference on*, pages 196–199. IEEE, 2012. (Cited in page(s) 18 and 255.)
- [Fang 2008] J. Fang and H. Li. *Distributed adaptive quantization for wireless sensor networks: from delta modulation to maximum likelihood*. Signal Processing, IEEE Transactions on, vol. 56, no. 10, pages 5246–5257, 2008. (Cited in page(s) 20, 32, 62, 63, 64, 66, 106, 112, 118, 256, 268 and 269.)
- [Fine 1968] T. Fine. *The response of a particular nonlinear system with feedback to each of two random processes*. Information Theory, IEEE Transactions on, vol. 14, no. 2, pages 255–264, 1968. (Cited in page(s) 63 and 239.)

- [Gallager 1996] R.G. Gallager. Discrete stochastic processes, volume 101. Kluwer Academic Publishers, 1996. (Cited in page(s) 239.)
- [Gastpar 2008] M. Gastpar. *Uncoded transmission is exactly optimal for a simple Gaussian “sensor” network*. Information Theory, IEEE Transactions on, vol. 54, no. 11, pages 5247–5251, 2008. (Cited in page(s) 19.)
- [Gersho 1992] A. Gersho and R.M. Gray. Vector quantization and signal compression. Springer, 1992. (Cited in page(s) 91, 109, 148, 174, 184 and 197.)
- [Golub 1973] G.H. Golub. *Some modified matrix eigenvalue problems*. SIAM Review, pages 318–334, 1973. (Cited in page(s) 153 and 231.)
- [Golub 1991] G.H. Golub and J.M. Ortega. Scientific computing and differential equations: An introduction to numerical methods. Academic Press, Inc., 1991. (Cited in page(s) 116.)
- [Gordon 1993] N.J. Gordon, D.J. Salmond and A.F.M. Smith. *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*. In Radar and Signal Processing, IEE Proceedings F, volume 140, pages 107–113. IET, 1993. (Cited in page(s) 86.)
- [Gubner 1993] J.A. Gubner. *Distributed estimation and quantization*. Information Theory, IEEE Transactions on, vol. 39, no. 4, pages 1456–1459, 1993. (Cited in page(s) 20 and 256.)
- [Gupta 2003] R. Gupta and A.O. Hero III. *High-rate vector quantization for detection*. Information Theory, IEEE Transactions on, vol. 49, no. 8, pages 1951–1969, 2003. (Cited in page(s) 19, 208, 212 and 256.)
- [Hardy 1988] G.H. Hardy, J.E. Littlewood and G. Polya. Inequalities. Cambridge University Press, 1988. (Cited in page(s) 43 and 184.)
- [Herzig 2002] P. Herzig, M.D. Hannington and S. Petersen. *Polymetallic massive sulphide deposits at the modern seafloor and their resources potential*. Technical report, 2002. (Cited in page(s) 173.)
- [Hoagland 2010] P. Hoagland, S. Beaulieu, M.A. Tivey, R.G. Eggert, C. German, L. Glowka and J. Lin. *Deep-sea mining of seafloor massive sulfides*. Marine Policy, vol. 34, no. 3, pages 728–732, 2010. (Cited in page(s) 173.)
- [Hol 2006] J.D. Hol, T.B. Schon and F. Gustafsson. *On resampling algorithms for particle filters*. In Nonlinear Statistical Signal Processing Workshop, 2006 IEEE, pages 79–82. IEEE, 2006. (Cited in page(s) 86.)
- [Intanagonwiwat 2000] C. Intanagonwiwat, R. Govindan and D. Estrin. *Directed diffusion: a scalable and robust communication paradigm for sensor networks*. In Proceedings of the 6th annual international conference on Mobile computing and networking, pages 56–67. ACM, 2000. (Cited in page(s) 17 and 254.)

- [Jazwinski 1970] A.H. Jazwinski. *Stochastic processes and filtering theory*. Academic Press, New York, 1970. (Cited in page(s) 29, 78, 91 and 97.)
- [Karlsson 2005] G.R. Karlsson and F. Gustafsson. *Particle Filtering for Quantized Sensor Information*. In 13th European Signal Processing Conference, EUSIPCO. EURASIP, 2005. (Cited in page(s) 87.)
- [Kassam 1977] S. Kassam. *Optimum quantization for signal detection*. Communications, IEEE Transactions on, vol. 25, no. 5, pages 479–484, 1977. (Cited in page(s) 19, 53 and 256.)
- [Kay 1993] S.M. Kay. *Fundamentals of statistical signal processing, volume 1: Estimation theory*. PTR Prentice Hall, 1993. (Cited in page(s) 38, 39, 40, 53, 62, 70, 88, 91, 95, 227, 263 and 264.)
- [Khalil 1992] H.K. Khalil and J.W. Grizzle. *Nonlinear systems*. Macmillan Publishing Company New York, 1992. (Cited in page(s) 116.)
- [Knuth 1997] D.E. Knuth. *The art of computer programming, volume 2: Seminumerical algorithms*. Addison-Wesley, 1997. (Cited in page(s) 248.)
- [Kong 1994] A. Kong, J.S. Liu and W.H. Wong. *Sequential imputations and Bayesian missing data problems*. Journal of the American Statistical Association, vol. 89, no. 425, pages 278–288, 1994. (Cited in page(s) 85.)
- [Lange 1989] K.L. Lange, R.J.A. Little and J.M.G. Taylor. *Robust statistical modeling using the  $t$  distribution*. Journal of the American Statistical Association, pages 881–896, 1989. (Cited in page(s) 136.)
- [Li 1999] J. Li, N. Chaddha and R.M. Gray. *Asymptotic performance of vector quantizers with a perceptual distortion measure*. Information Theory, IEEE Transactions on, vol. 45, no. 4, pages 1082–1091, 1999. (Cited in page(s) 188.)
- [Li 2007] H. Li and J. Fang. *Distributed adaptive quantization and estimation for wireless sensor networks*. Signal Processing Letters, IEEE, vol. 14, no. 10, pages 669–672, 2007. (Cited in page(s) 60, 62, 106, 118 and 268.)
- [Longo 1990] M. Longo, T.D. Lookabaugh and R.M. Gray. *Quantization for decentralized hypothesis testing under communication constraints*. Information Theory, IEEE Transactions on, vol. 36, no. 2, pages 241–255, 1990. (Cited in page(s) 19 and 256.)
- [Luo 2005] Z.Q. Luo. *Universal decentralized estimation in a bandwidth constrained sensor network*. Information Theory, IEEE Transactions on, vol. 51, no. 6, pages 2210–2219, 2005. (Cited in page(s) 20 and 256.)
- [Marano 2007] S. Marano, V. Matta and P. Willett. *Asymptotic design of quantizers for decentralized MMSE estimation*. Signal Processing, IEEE Transactions on, vol. 55, no. 11, pages 5485–5496, 2007. (Cited in page(s) 20, 42, 208 and 256.)

- [Marsaglia 2000] G. Marsaglia and W.W. Tsang. *A simple method for generating gamma variables*. ACM Transactions on Mathematical Software (TOMS), vol. 26, no. 3, pages 363–372, 2000. (Cited in page(s) 249.)
- [Molden 2007] D. Molden. Water for food, water for life: a comprehensive assessment of water management in agriculture. Earthscan/James & James, 2007. (Cited in page(s) 27.)
- [Nardon 2009] M. Nardon and P. Pianca. *Simulation techniques for generalized Gaussian densities*. Journal of Statistical Computation and Simulation, vol. 79, no. 11, pages 1317–1329, 2009. (Cited in page(s) 249.)
- [Papadopoulos 2001] H.C. Papadopoulos, G.W. Wornell and A.V. Oppenheim. *Sequential signal encoding from noisy measurements using quantizers with dynamic bias control*. Information Theory, IEEE Transactions on, vol. 47, no. 3, pages 978–1002, 2001. (Cited in page(s) 20, 44, 53, 60, 66, 70, 72, 105, 106, 256, 265 and 269.)
- [Picinbono 1988] B. Picinbono and P. Duvaut. *Optimum quantization for detection*. Communications, IEEE Transactions on, vol. 36, no. 11, pages 1254–1258, 1988. (Cited in page(s) 19 and 256.)
- [Poor 1977] H.V. Poor and J. Thomas. *Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems*. Communications, IEEE Transactions on, vol. 25, no. 9, pages 893–900, 1977. (Cited in page(s) 19 and 256.)
- [Poor 1988] H.V. Poor. *Fine quantization in signal detection and estimation*. Information Theory, IEEE Transactions on, vol. 34, no. 5, pages 960–972, 1988. (Cited in page(s) 19, 20, 178, 207, 209, 214 and 256.)
- [Puccinelli 2005] D. Puccinelli and M. Haenggi. *Wireless sensor networks: applications and challenges of ubiquitous sensing*. Circuits and Systems Magazine, IEEE, vol. 5, no. 3, pages 19–31, 2005. (Cited in page(s) 18 and 255.)
- [Rhodes 1971] I. Rhodes. *A tutorial introduction to estimation and filtering*. Automatic Control, IEEE Transactions on, vol. 16, no. 6, pages 688–706, 1971. (Cited in page(s) 98.)
- [Ribeiro 2006a] A. Ribeiro and G.B. Giannakis. *Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case*. Signal Processing, IEEE Transactions on, vol. 54, no. 3, pages 1131–1143, 2006. (Cited in page(s) 20, 44, 53, 58, 60, 63, 106, 256 and 265.)
- [Ribeiro 2006b] A. Ribeiro and G.B. Giannakis. *Bandwidth-constrained distributed estimation for wireless sensor networks-part II: Unknown probability density function*. Signal Processing, IEEE Transactions on, vol. 54, no. 7, pages 2784–2796, 2006. (Cited in page(s) 20, 106 and 256.)
- [Ribeiro 2006c] A. Ribeiro, G.B. Giannakis and S.I. Roumeliotis. *SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations*. Signal Processing, IEEE Transactions on, vol. 54, no. 12, pages 4782–4795, 2006. (Cited in page(s) 20, 75, 94, 95, 106 and 256.)

- [Robert 1999] C.P. Robert and G. Casella. Monte Carlo statistical methods. Springer New York, 1999. (Cited in page(s) 81, 82 and 245.)
- [Ruan 2004] Y. Ruan, P. Willett, A. Marrs, S. Marano and F. Palmieri. *Practical fusion of quantized measurements via particle filtering*. In Target Tracking 2004: Algorithms and Applications, IEE, pages 13–18. IET, 2004. (Cited in page(s) 87.)
- [Rubin 1988] D.B. Rubin *et al.* *Using the SIR algorithm to simulate posterior distributions*. Bayesian statistics, vol. 3, pages 395–402, 1988. (Cited in page(s) 86.)
- [Samorodnitsky 1994] G. Samorodnitsky and M.S. Taqqu. Stable non-Gaussian random processes: stochastic models with infinite variance. Chapman and Hall/CRC, 1994. (Cited in page(s) 34.)
- [Sigman 1999] K. Sigman. *Appendix: A primer on heavy-tailed distributions*. Queueing Systems, vol. 33, no. 1, pages 261–275, 1999. (Cited in page(s) 47.)
- [Sukhavasi 2009a] R.T. Sukhavasi and B. Hassibi. The Kalman like particle filter : Optimal estimation with quantized innovations/measurements. arXiv:0909.0996, September 2009. (Cited in page(s) 95.)
- [Sukhavasi 2009b] R.T. Sukhavasi and B. Hassibi. *Particle filtering for Quantized Innovations*. In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, pages 2229–2232, April 2009. (Cited in page(s) 91.)
- [Tichavsky 1998] P. Tichavsky, C.H. Muravchik and A. Nehorai. *Posterior Cramér-Rao bounds for discrete-time nonlinear filtering*. Signal Processing, IEEE Transactions on, vol. 46, no. 5, pages 1386–1396, May 1998. (Cited in page(s) 88.)
- [Tsitsiklis 1993] J.N. Tsitsiklis. *Extremal properties of likelihood-ratio quantizers*. Communications, IEEE Transactions on, vol. 41, no. 4, pages 550–558, 1993. (Cited in page(s) 19 and 256.)
- [Van Trees 1968] H. L. Van Trees. Detection, estimation, and modulation theory. Part 1. New York: John Wiley and Sons, Inc., 1968. (Cited in page(s) 37, 78, 88 and 208.)
- [Varanasi 1989] M.K. Varanasi and B. Aazhang. *Parametric generalized Gaussian density estimation*. The Journal of the Acoustical Society of America, vol. 86, pages 1404–1415, 1989. (Cited in page(s) 136.)
- [Villard 2010] J. Villard, P. Bianchi, E. Moulines and P. Piantanida. *High-rate quantization for the Neyman-Pearson detection of hidden Markov processes*. In Information Theory Workshop (ITW), 2010 IEEE, pages 1–5. IEEE, 2010. (Cited in page(s) 19 and 256.)
- [Villard 2011] J. Villard and P. Bianchi. *High-rate vector quantization for the Neyman-Pearson detection of correlated processes*. Information Theory, IEEE Transactions on, vol. 57, no. 8, pages 5387–5409, 2011. (Cited in page(s) 19 and 256.)
- [Wang 2010] L.Y. Wang, G. Yin, J.F. Zhang and Y. Zhao. System identification with quantized observations. Birkhauser, 2010. (Cited in page(s) 20, 32 and 256.)

- 
- [Wasserman 2003] L. Wasserman. All of statistics: a concise course in statistical inference. Springer, 2003. (Cited in page(s) 53, 68 and 72.)
- [You 2008] K. You, L. Xie, S. Sun and W. Xiao. *Multiple-level quantized innovation Kalman filter*. In IFAC World Congress, volume 17, pages 1420–1425, 2008. (Cited in page(s) 75, 95 and 106.)
- [Zhao 2004] F. Zhao and L. Guibas. Wireless sensor networks: an information processing approach. Morgan Kaufmann, 2004. (Cited in page(s) 17 and 254.)



---

**Abstract:** With recent advances in sensing and communication technology, sensor networks have emerged as a new field in signal processing. One of the applications of this new field is remote estimation, where the sensors gather information and send it to some distant point where estimation is carried out. For overcoming the new design challenges brought by this approach (constrained energy, bandwidth and complexity), quantization of the measurements can be considered. Based on this context, we study the problem of estimation based on quantized measurements. We focus mainly on the scalar location parameter estimation problem, the parameter is considered to be either constant or varying according to a slow Wiener process model. We present estimation algorithms to solve this problem and, based on performance analysis, we show the importance of quantizer range adaptiveness for obtaining optimal performance. We propose a low complexity adaptive scheme that jointly estimates the parameter and updates the quantizer thresholds, achieving in this way asymptotically optimal performance. With only 4 or 5 bits of resolution, the asymptotically optimal performance for uniform quantization is shown to be very close to the continuous measurement estimation performance. Finally, we propose a high resolution approach to obtain an approximation of the optimal nonuniform quantization thresholds for parameter estimation and also to obtain an analytical approximation of the estimation performance based on quantized measurements.

**Keywords:** estimation, quantization, compression, adaptive algorithms.

---

**Résumé :** L'essor des nouvelles technologies de télécommunication et de conception des capteurs a fait apparaître un nouveau domaine du traitement du signal : les réseaux de capteurs. Une application clé de ce nouveau domaine est l'estimation à distance : les capteurs acquièrent de l'information et la transmettent à un point distant où l'estimation est faite. Pour relever les nouveaux défis apportés par cette nouvelle approche (contraintes d'énergie, de bande et de complexité), la quantification des mesures est une solution. Ce contexte nous amène à étudier l'estimation à partir de mesures quantifiées. Nous nous concentrons principalement sur le problème d'estimation d'un paramètre de centrage scalaire. Le paramètre est considéré soit constant, soit variable dans le temps et modélisé par un processus de Wiener lent. Nous présentons des algorithmes d'estimation pour résoudre ce problème et, en se basant sur l'analyse de performance, nous montrons l'importance de l'adaptativité de la dynamique de quantification pour l'obtention d'une performance optimale. Nous proposons un schéma adaptatif de faible complexité qui, conjointement, estime le paramètre et met à jour les seuils du quantifieur. L'estimateur atteint de cette façon la performance asymptotique optimale. Avec 4 ou 5 bits de résolution, nous montrons que la performance optimale pour la quantification uniforme est très proche des performances d'estimation à partir de mesures continues. Finalement, nous proposons une approche à haute résolution pour obtenir les seuils de quantification non-uniformes optimaux ainsi qu'une approximation analytique des performances d'estimation.

**Mots clés :** estimation, quantification, compression, algorithmes adaptatifs.

---