



HAL
open science

Consistency of weighted least-square estimators for parameter estimation problems based on binary measurements

Jérôme Juillard, Kian Jafaridinani, Eric Colinet

► **To cite this version:**

Jérôme Juillard, Kian Jafaridinani, Eric Colinet. Consistency of weighted least-square estimators for parameter estimation problems based on binary measurements. 15th IFAC Symposium on System Identification (SYSID'09), Jul 2009, Saint-Malo, France. pp.72-77. hal-00445435

HAL Id: hal-00445435

<https://centralesupelec.hal.science/hal-00445435v1>

Submitted on 8 Jan 2010

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.

Consistency of weighted least-square estimators for parameter estimation problems based on binary measurements

J. Juillard*, K. Jafaridinani*, E. Colinet**

* SUPELEC - SSE, Gif-sur-Yvette, France

(e-mail: Jerome.Juillard@supelec.fr, Kian.Jafaridinani@supelec.fr).

** CEA-LETI - MINATEC, Grenoble, France

(e-mail: Eric.Colinet@cea.fr)

Abstract: In this paper, we present a new weighted least-squares (WLS) approach for parameter estimation based on binary data. Two WLS criteria are studied. We show that these two criteria do not have the same asymptotical behavior although they are closely related. Particularly, in the presence of noise, one of the criteria used for determining the system parameters provides an appropriate estimation, whereas the other one leads to an underestimation of the system parameters. These asymptotical results are illustrated by simulations in Gaussian and non-Gaussian contexts.

1. INTRODUCTION

In this paper, we give some theoretical results concerning the asymptotic bias of least-square approaches to parameter estimation methods based on binary measurements. This work is based on (Colinet and Juillard, 2008). It is originally motivated by the need to add integrated low-cost self-test features to micro fabricated devices, such as MEMS and NEMS. Even though there exists a wide range of applications where identification methods based on binary observations are necessary or desirable (Wang *et al.*, 2003), the focus is brought here on the test of microelectronic devices.

1.1 Context and state of the art

It is well-known that, as characteristic dimensions become smaller, the dispersions afflicting electronic devices tend to become larger. Typical sources of dispersion are variations in the fabrication process or changes in the operating conditions, such as temperature and ambient pressure. Imperfect knowledge of semiconductor physics or of nano-mechanics is also a large source of uncertainty. As a consequence, it is usually impossible to guarantee a priori that a given device will function properly. The tests that are run after fabrication, under different operating conditions, ensure that only suitable devices are commercialized. However, these tests are very costly: a typical figure is that one third of the cost of a micro fabricated device is due to testing (Charlot *et al.*, 2001). An alternate solution is to integrate low-cost self-test (and self-tuning) features, such as parameter estimation routines, in each device, so that it can adapt to changing conditions.

However, most parameter estimation methods (Ljung, 1999; Walter and Pronzato, 1997) do not easily scale down to the micro- or nano-world. This is because these methods rely on high-resolution digital measurements of the system output. Their integration requires the implementation of high-resolution analogue-to-digital converters (ADCs) and, thus, results in longer design times, larger silicon areas and

increased costs. Our objective is then to develop a parameter estimation method that relies on very low-resolution (ideally binary) measurements, in order to keep the added cost as small as possible.

In the field of micro-electronics, this issue has been addressed by Negreiros (2003): the proposed method consists in using a white Gaussian input to excite the unknown (linear) system and to estimate the power spectral density (PSD) of the binary output, which is assumed to be stationary. From this estimated PSD, it is possible to analytically derive the modulus of the transfer function of the unknown system. However, one cannot use this approach to obtain any information concerning the phase of the transfer function. Moreover, the generation of a white Gaussian input requires a high-resolution digital-to-analogue converter (DAC), whose implementation can be costly. These issues are partially solved in (Juillard and Colinet, 2008), where a white Bernoulli input is used as an exciting signal (thus requiring only a 1-bit DAC). Provided the mixing properties of the linear system are good (i.e. its impulse response does not vanish too quickly), it is possible to establish an analytical relationship between the cross-covariance function of the binary inputs and outputs and the impulse response of the unknown system and, thus, to identify the system.

Outside the context of micro-electronics, the most significant contributions come from Wigren (1998) and Wang and his co-workers (Wang *et al.*, 2006; Wang *et al.*, 2003; Zhao Y. *et al.*, 2007). In (Wang *et al.*, 2003), Wang introduced a new method for estimating parameters from binary (or quantized) data. The unknown system is excited by a periodic signal and, as in (Rafaljowicz, 1996), the threshold of the quantizer is randomly specified by a partially known dithering signal. However, the hypotheses on the quantizer's input and on the random threshold are much less strict than in (Rafaljowicz, 1996): the main constraint is that the cumulative distribution function (cdf) of the threshold must be invertible. Although it is usually possible to generate continuous dithering signals, this constraint prevents their approach from being used with

discrete-valued dithering signals, which are very easy to generate in the context of microelectronic applications. Under these hypotheses, the parameter estimation problem reduces to solving a linear system. This approach is generalized in (Wang *et al.*, 2006), where it is shown that the cdf does not have to be known a priori: it can be estimated along with the parameters of the system. This work has been extended from finite impulse response (FIR) systems to infinite impulse response (IIR) systems and to nonlinear Wiener systems (Wang *et al.*, 2006; Zhao *et al.*, 2007). Finally, Wigren has developed a least mean squares (LMS) approach to the problem of online parameter estimation from binary observations (Wigren, 1998). This method is based on an approximation of the quantizer, which makes it possible to define an approximate gradient of the least-squares criterion (this is developed in section II). Under some hypotheses (Ljung, 1977), it is possible to guarantee the asymptotical convergence of this method to the nominal parameters (or at least to the boundary of the model set).

Each of these methods has its merits and weaknesses. It is our opinion that Wigren's approach is the one that fits better in the context of test of electronics system: as opposed to Wang's or Rafajlowicz's methods, it does not require a varying threshold (dithering signal), which can only be implemented with a high-resolution DAC or a surface-consuming analog component. Furthermore, it can be used online, without specific input design, as well as offline. However, it does rely on an approximation of the quantizer - and, thus, it is not an exact method, stricto sensu, as opposed to (Colinet and Juillard, 2008). Moreover, convergence is not guaranteed in the presence of measurement noise at the quantizer's input.

The method introduced in (Colinet and Juillard, 2008) bears some likeness to Wigren's, in the sense that it is based on the minimization of a least-squares criterion. Its main distinguishing feature is that no approximation of the quantizer is made (hence, it is not based on a "pseudo-gradient"): rather, it relies on a weighted least-squares (WLS) criterion where the parameter-dependent weights are chosen in order to smooth out the discontinuities. The asymptotical convergence of this approach to the nominal system parameters can be guaranteed, even in the presence of measurement noise, provided the signal at the quantizer's input is Gaussian and centred. Note that these hypotheses used in (Colinet and Juillard, 2008) only provide sufficient conditions for the asymptotical convergence and should not be considered as restrictive.

1.2 Outline of the article

In this paper, we present some new asymptotical results concerning WLS criteria in the presence of measurement noise at the quantizer's input. The structure of the article is the following. In section II, the notations are introduced and the WLS criteria presented in (Colinet and Juillard, 2008) is re-introduced, along with their asymptotical properties. In section III, we study an alternative WLS approach: we prove that this approach is as efficient as the one proposed in

(Colinet and Juillard, 2008) in the absence of measurement noise. On the other hand, we show that this other approach leads to a systematic error in the presence of measurement noise. Some examples are given in section IV. Section V contains some concluding remarks.

2. PRELIMINARIES

2.1 Framework and notations

Let us consider a discrete-time invariant linear system H . We assume H has a finite impulse response of length L , i.e. the impulse response can be represented by a column vector $\boldsymbol{\theta} = (\theta_k)_{k=1}^L$. Let u_l be the known scalar value of the system input at time l and y_l be the (scalar) value of the system output, so that:

$$y_l = \boldsymbol{\varphi}_l^T \boldsymbol{\theta}, \quad (1)$$

where $\boldsymbol{\varphi}_l = (u_k)_{k=l-L+1}^l$ is the (column) vector of observations at time l . Let d_l be an additive dithering signal and b_l an additive noise at the quantizer's input. Let m_x and σ_x^2 denote the first- and second-order moments of any signal x . The system output is measured via a 1-bit ADC so that only the sign $s_l = S(z_l)$ of the system output is known, where

$$\begin{cases} S(x) = 1, & \text{if } x \geq 0 \\ S(x) = -1, & \text{otherwise} \end{cases}, \quad (2)$$

and $z_l = y_l + d_l + b_l$.

We are interested in finding an estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, based on N observations of u_l , s_l and, if need be, d_l . It is assumed that u_l , d_l and b_l are stationary and uncorrelated and that z_l is Gaussian and centred. The estimated quantities are denoted by a hat (e.g. $\hat{z}_l = \hat{y}_l + \hat{d}_l$). In order to simplify some expressions, d_l and b_l are assumed to be white and centred and $\hat{\boldsymbol{\theta}}$ has length L .

These notations are summed up in Fig. 1.

2.2 WLS approach for parameter estimation based on binary data

In (Colinet and Juillard, 2008), we show that the problem of parameter estimation from binary measurement can be treated by minimizing WLS criteria of the form:

$$J_f^N = \frac{1}{4} \frac{\sum_{l=1}^N f(\hat{z}_l) (s_l - \hat{s}_l)^2}{\sum_{l=1}^N f(\hat{z}_l)}, \quad (3)$$

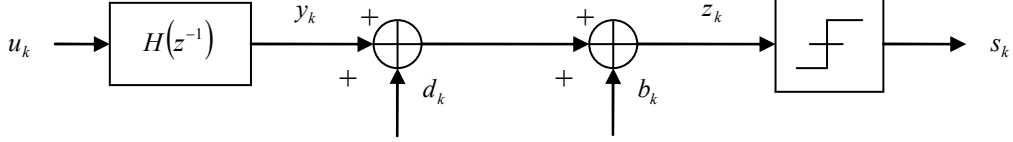


Fig. 1. Framework and notations

where $f(\cdot)$ is any continuously differentiable even function that satisfies $f(0)=0$ and $f'(x)>0, \forall x \in \mathfrak{R}^{**}$ (in practice, $f(x)=x^2$ is often used). It is simple to show that $J_f^N(\hat{\theta})$ is continuously differentiable everywhere, except where:

$$\sum_{l=1}^N f(\hat{z}_l) = 0. \quad (4)$$

Since $f(\cdot)$ is positive, (4) reduces to:

$$\begin{bmatrix} \Phi_1^T \\ \vdots \\ \Phi_N^T \end{bmatrix} \hat{\theta} + \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix} = \mathbf{0}, \quad (5)$$

which is a system of N equations with L ($L < N$) unknowns. Provided u_l has a density and $\exists l / d_l \neq 0$, it can be proved that (5) has no solution with probability 1. When $d_l = 0, 1 \leq l \leq N$, (4) has $\hat{\theta} = \mathbf{0}$ for solution, regardless of u_l .

Since $(\hat{s}_l - s_l)^2$ takes only two values, 0 and 4, it is clear that $0 \leq J_f^N \leq 1$ and that:

$$J_0^N = 0 \Rightarrow J_f^N = 0, \quad (6)$$

where J_0^N is the (discontinuous) non-weighted least-squares criterion. In (Colinet and Juillard, 2008), the case when $f(x)=x^2$ is studied in depth and the properties of the corresponding WLS criterion, J_1^N , are established in a probabilistic framework, when N goes to infinity. In particular, it is shown that J_1^N converges surely to:

$$J_1^\infty = \frac{1}{\pi} \left(\arccos(r) - r\sqrt{1-r^2} \right), \quad (7)$$

where r is the correlation coefficient of z and \hat{z} . Under the assumptions made in section II.1, the expression of r is:

$$r(\hat{\theta}) = \frac{\sigma_d^2 + \theta^T \mathbf{R}_u \hat{\theta}}{\sqrt{\sigma_b^2 + \sigma_d^2 + \theta^T \mathbf{R}_u \theta} \sqrt{\sigma_d^2 + \hat{\theta}^T \mathbf{R}_u \hat{\theta}}}, \quad (8)$$

where \mathbf{R}_u is the $L \times L$ covariance matrix of u . By straightforward differentiation, one can then prove the following properties.

Property 1 - Provided \mathbf{R}_u has full rank, the minimum of J_1^∞ is reached for:

- $\hat{\theta} = \theta$ if $\sigma_d^2 \neq 0$.
- $\hat{\theta} = \lambda \theta, \forall \lambda > 0$ if $\sigma_d^2 = 0$.

Property 2 - $J_1^\infty(\hat{\theta})$ is convex in a neighbourhood of its minimum.

Note that these properties hold regardless of the presence or absence of noise at the quantizer's input. Note also that it is not possible to identify the system's gain if $\sigma_d^2 = 0$. The dithering signal acts as reference amplitude, as in (Wang *et al.*, 2003).

3. AN ALTERNATIVE WLS APPROACH

An alternative approach to the problem of parameter estimation based on binary measurements is to define a criterion of the form:

$$G_f^N = \frac{1}{4N} \sum_{l=1}^N f(\hat{z}_l) (s_l - \hat{s}_l)^2, \quad (9)$$

where $f(\cdot)$ has the same properties as in section II.2, in order to ensure continuous differentiability of G_f^N . In the case when $f(x)=x^2$, the corresponding criterion, G_1^N , lends itself to analysis in much the same way as J_1^N . In particular, if z_l and \hat{z}_l are stationary:

$$G_1^\infty \stackrel{\Delta}{=} \lim_{N \rightarrow \infty} G_1^N = \frac{1}{4} E(\hat{z}_l^2 (s_l - \hat{s}_l)^2), \quad (10)$$

where $E(\cdot)$ denotes the expected value. Under the assumption that z_l and \hat{z}_l are Gaussian and centred, we have:

$$\begin{aligned} G_1^\infty &= \frac{1}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{z}^2 (S(\hat{z}) - S(z))^2 p(\hat{z}, z) d\hat{z} dz \\ &= \frac{1}{4} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{z}^2 p(\hat{z}, z) d\hat{z} dz + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{z}^2 p(\hat{z}, z) d\hat{z} dz \right), \end{aligned} \quad (11)$$

where

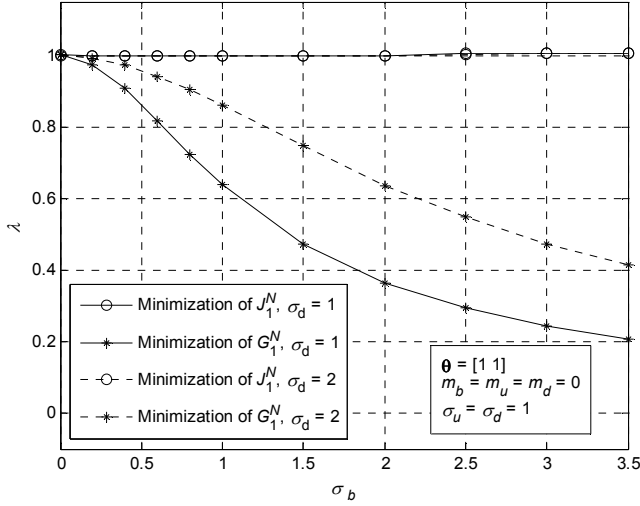


Fig. 2. $\hat{\lambda}$ vs. σ_b when u_i , b_i and d_i are Gaussian and centred.

$$p(\hat{z}, z) = \frac{\exp\left(\frac{-1}{2(1-r^2)}\left(\frac{\hat{z}^2}{\sigma_z^2} + \frac{z^2}{\sigma_z^2} - \frac{2r\hat{z}z}{\sigma_z\sigma_z}\right)\right)}{2\pi\sigma_z\sigma_z\sqrt{1-r^2}},$$

and thus:

$$G_1^\infty = \frac{\sigma_z^2}{\pi} \left(\text{acos}(r) - r\sqrt{1-r^2} \right) = \sigma_z^2 J_1^\infty. \quad (12)$$

The optimal $\hat{\theta}$ is found by differentiating (12). This yields:

$$\frac{\partial G_1^\infty}{\partial \hat{\theta}} = \mathbf{0} \Leftrightarrow 2J_1^\infty \mathbf{R}_u \hat{\theta} + \sigma_z^2 \frac{dJ_1^\infty}{dr} \frac{\partial r}{\partial \hat{\theta}} = \mathbf{0}, \quad (13)$$

where

$$\frac{dJ_1^\infty}{dr} = -\frac{2}{\pi} \sqrt{1-r^2}, \quad (14)$$

and

$$\sigma_z^2 \frac{\partial r}{\partial \hat{\theta}} = \frac{\sigma_z}{\sigma_z} \mathbf{R}_u \hat{\theta} - r \mathbf{R}_u \hat{\theta}. \quad (15)$$

Thus, (13) is equivalent to:

$$\text{acos}(r) \mathbf{R}_u \hat{\theta} - \sqrt{1-r^2} \frac{\sigma_z}{\sigma_z} \mathbf{R}_u \hat{\theta} = \mathbf{0}. \quad (16)$$

If \mathbf{R}_u has full rank, this reduces to:

$$\text{acos}(r) \hat{\theta} = \sqrt{1-r^2} \frac{\sigma_z}{\sigma_z} \hat{\theta}. \quad (17)$$

Thus the optimal $\hat{\theta}$ is (at least) collinear to θ . Two cases must now be distinguished, depending on whether a dithering signal is present at the quantizer's input or not.

3.1 Case 1: $\sigma_d^2 \neq 0$

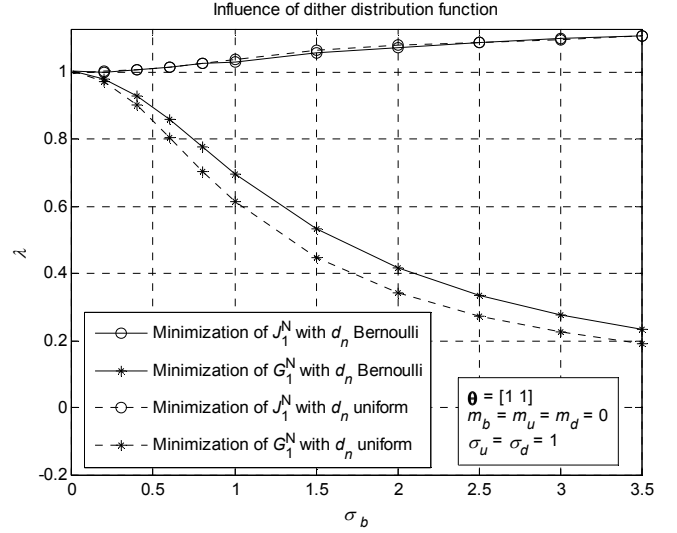


Fig. 3. $\hat{\lambda}$ vs. σ_b when u_i and b_i are Gaussian and centred, whereas d_i is uniform or Bernoulli.

Suppose now that $r=1$: in this case, $\text{acos}(r) = \sqrt{1-r^2} = 0$ and (16) is verified. In other words, any vector $\hat{\theta}$ such that $r=1$ minimizes G_1^∞ . Now, it is clear that r can be equal to 1 if and only if:

$$\begin{cases} \sigma_b^2 = 0 \\ \hat{\theta} = \theta \end{cases}. \quad (18)$$

Thus the following property holds.

Property 3 - Provided \mathbf{R}_u has full rank, the minimum of G_1^∞ is reached for $\hat{\theta} = \theta$ if $\sigma_d^2 \neq 0$ and $\sigma_b^2 = 0$.

Consider now that $\sigma_b^2 \neq 0$ and, as a consequence, that $r \neq 1$. (17) can be rewritten as:

$$\hat{\theta} = \frac{\sqrt{1-r^2}}{\text{acos}(r)} \frac{\sigma_z}{\sigma_z} \theta. \quad (19)$$

Let

$$\lambda = \frac{\sqrt{1-r^2}}{\text{acos}(r)} \frac{\sigma_z}{\sigma_z} = g(r) \frac{\sigma_z}{\sigma_z}. \quad (20)$$

Taking the square of (20) and multiplying by σ_z^2 yields:

$$\begin{aligned} \lambda^2 (\sigma_b^2 + \sigma_d^2 + \theta^T \mathbf{R}_u \theta) &= g(r)^2 (\sigma_d^2 + \lambda^2 \theta^T \mathbf{R}_u \theta) \\ \Leftrightarrow \lambda^2 (\sigma_b^2 + \sigma_d^2 + (1-g(r)^2) \theta^T \mathbf{R}_u \theta) &= g(r)^2 \sigma_d^2 \end{aligned} \quad (21)$$

and therefore:

$$\lambda = \frac{g(r) \sigma_d}{\sqrt{\sigma_b^2 + \sigma_d^2 + (1-g(r)^2) \theta^T \mathbf{R}_u \theta}} \quad (22)$$

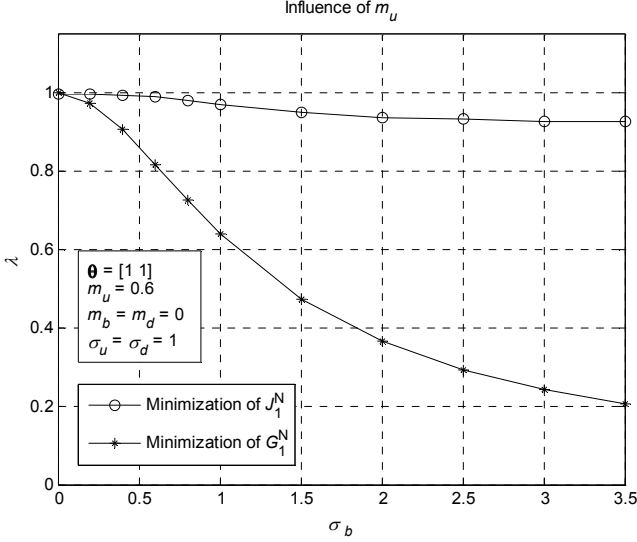


Fig. 4. $\hat{\lambda}$ vs. σ_b when d_i and b_i are Gaussian and centred and u_i is Gaussian with nonzero mean.

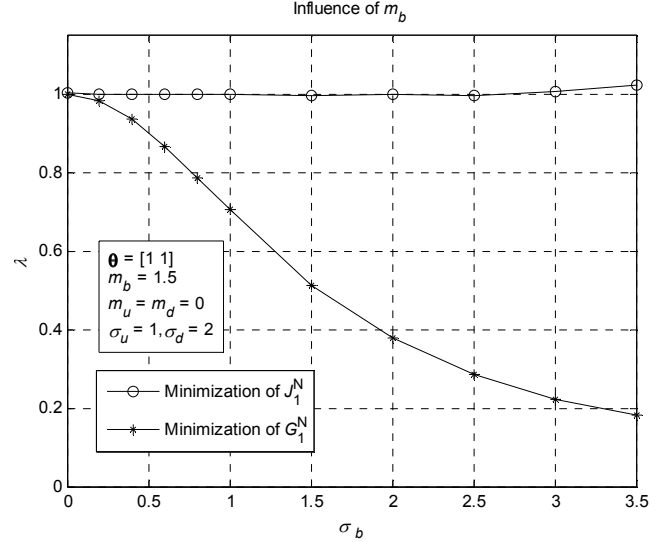


Fig. 5. $\hat{\lambda}$ vs. σ_b when u_i and d_i are Gaussian and centred and b_i is Gaussian with nonzero mean.

We know that $0 \leq g(r) < 1, \forall r < 1$ and that $\sigma_b^2 \neq 0$. From (22), it is straightforward that:

$$\lambda < 1. \quad (23)$$

Thus, in the presence of measurement noise at the quantizer's input, one cannot hope to achieve a consistent estimation of the nominal parameters θ by minimizing G_1^∞ .

Property 4 - Provided \mathbf{R}_u has full rank, the minimum of G_1^∞ is reached for $\hat{\theta} = \lambda\theta$, $\lambda < 1$ if $\sigma_d^2 \neq 0$ and $\sigma_b^2 \neq 0$.

3.2 Case 2: $\sigma_d^2 = 0$

In the simpler case when $\sigma_d^2 = 0$, it is simple to show the following two properties.

Property 5 - Provided \mathbf{R}_u has full rank, the minimum of G_1^∞ is reached for:

- $\hat{\theta} = \lambda\theta, \forall \lambda \geq 0$, if $\sigma_d^2 = 0$ and $\sigma_b^2 = 0$.
- $\hat{\theta} = \mathbf{0}$, if $\sigma_d^2 = 0$ and $\sigma_b^2 \neq 0$.

4. SIMULATION RESULTS

In this section, we compare the results obtained with WLS criteria G_1^N and J_1^N defined in the previous sub-sections. We choose $\theta = [1 \ 1]$ (i.e. $L = 2$). The signals u_i , d_i and b_i have Gaussian distributions with zero mean: σ_u is set equal to 1, whereas σ_d and σ_b may vary. In order to satisfy the asymptotic conditions, one must impose $N \gg L$ (Colinet and Juillard, 2008). For each experiment, $\hat{\theta}$ is estimated with

G_1^N and J_1^N , thanks to a BFGS quasi-Newton method, implemented by the *fminunc* function of Matlab. Each experiment is repeated a large number of times, in order to estimate $\tilde{\lambda}$, the expected value of λ , for different values of σ_d and σ_b . More precisely, for given σ_d and σ_b , $\tilde{\lambda}$ is defined as:

$$\tilde{\lambda} = \frac{1}{P} \sum_{p=1}^P \frac{\|\hat{\theta}_p\|}{\|\theta\|}, \quad (24)$$

where $\hat{\theta}_p = \arg \min(G_1^N)$ or $\arg \min(J_1^N)$, depending on which criterion is tested, and P is the number of experiments. P is chosen so that the standard deviation of $\tilde{\lambda}$ is small. Note that the angle between $\hat{\theta}_p$ and θ must be small for (24) to give a meaningful estimation of λ . Since $N \gg L$, this should indeed be the case: as a measure of precaution, one can verify that the correlation coefficient between $\hat{\theta}_p$ and θ is very close to 1, for all experiments.

The results are shown in Fig. 2. The value of λ estimated with J_1^N is always very close to 1, regardless of the value of σ_d or σ_b : this is as expected from the asymptotic analysis (Property 1) and shows that it is possible to properly estimate the system parameters regardless of the level of noise at the quantizer input, when J_1^N is selected as the criterion.

Conversely, when G_1^N is used, λ is equal to 1 only when $\sigma_b = 0$. When σ_b increases, the estimated value of λ decreases, which illustrates Property 4. This shows that this criterion is not suitable for parameter estimation in a noisy context. Moreover, λ is not only a function of σ_b but also

of σ_d : as can be expected, the larger σ_d is with respect to σ_b , the closer λ is to 1.

Several other scenarios can be tested in which the constraints on the distributions of u_i , d_i and b_i are made less strict. In Fig. 3, u_i and b_i are Gaussian and centred, whereas d_i is uniform or Bernoulli. This test shows that, when σ_b increases, λ becomes larger than 1 when J_1^N is used. In Fig. 4, d_i and b_i are Gaussian and centred, whereas u_i is Gaussian with nonzero mean: simulations show that λ becomes smaller than 1 when J_1^N is used. Finally, in Fig. 5, u_i and d_i are Gaussian and centred, whereas b_i is Gaussian with nonzero mean. However, this does not seem to have any influence on λ , which stays very close to 1 when J_1^N is used. In these last three scenarios, the qualitative behaviour of λ when G_1^N is used does not differ much from the case when all signals are Gaussian and centred. One can also notice that the bias is always larger when G_1^N is used.

These simulations tend to show that the hypotheses made for obtaining Properties 1-6 may somehow be relaxed when $\sigma_b = 0$: J_1^N and G_1^N give equally good results regardless of the distribution of d_i or of the mean value of u_i or b_i . When $\sigma_b \neq 0$, J_1^N and G_1^N fail in correctly estimating $\|\theta\|$ when d_i is Bernoulli or uniform. This is also the case when u_i has nonzero mean. This goes to show that, if $\sigma_b \neq 0$, the only hypothesis that may be relaxed concerns the measurement noise b_i .

5. CONCLUSION

In this paper, we present a new WLS approach for parameter estimation based on binary data: this approach is an alternative to the one presented in (Colinet and Juillard, 2008). However, we have shown that these two criteria, although they are closely related, do not have the same asymptotical behavior. In particular, we have established that, in the presence of noise, using G_1^N for determining the system parameters leads to an underestimation of $\|\theta\|$, even though the direction of the vector is properly determined. These asymptotical results are confirmed by simulations performed in a Gaussian context. Other scenarios, where not all signals are Gaussian or centred, have also been tested: these show that the hypotheses made for obtaining the theoretical results can somehow be relaxed.

REFERENCES

Charlot, B. *et al.* (2001). Electrically induced stimuli for MEMS self-test. Proceedings IEEE VLSI Test Symposium, pp. 60-66.

Colinet, E. and J. Juillard (2008). A weighted least-squares approach to parameter estimation problems based on

binary measurements. Submitted to IEEE Transactions on Automatic Control.

Juillard J and E. Colinet (2008). Initialization of the BIMBO self-test method using binary inputs and outputs. Proceedings of the 46th IEEE Conference on Decision and Control, pp. 161-166.

Ljung L. (1977). Analysis of recursive stochastic algorithms. IEEE Transactions on Automatic Control, Vol. AC-22, pp. 551-575.

Ljung L. (1999). System identification - theory for the user, Prentice Hall, Upper Saddle River.

Negreiros M. *et al.* (2003). Ultimate low cost analog BIST. Proceedings 2003 Design Automation Conference, pp. 570-573.

Rafaljowicz E. (1996). Linear systems identification from random threshold binary data. IEEE Transactions on Signal Processing, Vol. 44, pp. 2064-2070.

Walter E. and L. Pronzato (1997). Identification of parametric models from experimental data, Springer-Verlag, Heidelberg.

Wang Y. L. *et al.* (2006). Joint identification of plant rational models and noise distribution functions using binary-valued observations. Automatica, Vol. 42, pp. 535-547.

Wang Y. L. *et al.* (2003). System identification using binary sensors. IEEE Transactions on Automatic Control, Vol.48, pp. 1892-1907.

Wigren T. (1998). Adaptive filtering using quantized output measurements. IEEE Transactions on Automatic Control, Vol.46, pp. 975-978.

Zhao Y. *et al.* (2007). Identification of Wiener systems with binary-valued output observations. Automatica, Vol.43, pp. 1752-1765.