MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Asymptotic Property of Optimal Quantization for System Identification

Koji TSUMURA

(Communicated by Kazuo Murota)

METR 2004–10 February 2004

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Asymptotic Property of Optimal Quantization for System Identification

Koji Tsumura

Department of Information Physics and Computing, The University of Tokyo, Hongo 7–3–1, Bunkyo-ku, Tokyo 113–8656, tel: +81–3–5841–6891, fax: +81–3–5841–6886, Koji_Tsumura@ipc.i.u-tokyo.ac.jp

Abstract: In this paper, we analyse the asymptotic property of the optimal quantization of signals used for system identification in high resolution case. We show an optimal quantization scheme for minimizing estimation errors under a constraint on the number of subsections of the quantized signals or the expectation of the optimal code length. The optimal quantization schemes can be given by solving Eular–Lagrange's equations and the solutions are functions of the distribution density of the regressor vector. We show examples of solutions for several cases of the regressor vectors and discuss their meanings with respect to the possibility of parameter estimations. In the case of the constraint of code length, the necessary information to attain the optimal identification errors is given as a function of the entropy of the regressor vector.

Keywords: system identification, quantization, least squares method, MA model, entropy

1 Introduction

The recent rapid improvement in the transmission capacity of computer networks makes long-distance automatic control to be more realistic and the necessity of understanding the effects of transmission limitations on information in control systems has become more widely accepted. In particular, quantization problem of signals has been discussed actively by several research groups in the last few years and interesting results have been achieved.

The problem of quantization of signals in control systems has a long history. The early results on the quantization in control theory are introduced in the book [3] of the 70s and at first, the quantization error is assumed to be a simple noise. The result by [4], [5] is recognized as a break through, in whose papers the behaviour of control systems, and their stability or state estimation, are analysed in detail. In the last few years, stabilization problems of quantized systems have been actively considered, e.g., [11], [12], [1], [8], [6].

Compared to this activity in the stabilization or estimation problem, the quantization problem for system identification [7] has not been adequately considered. When a controlled plant with networks is unknown or

its system parameters may change during the operation, we need a form of adaptation for the control system. It is also necessary to know the effect of quantization of the I/O data used for the system identification.

From such point of view, this problem was considered in [10] and an optimal quantization scheme for minimizing estimation errors under a constraint on the number of levels of the quantized signals was given. The optimal quantization is not uniform and the profile of the resolution was shown. This result is strictly applicable for any resolution of quantization, however, it has two problems which should be improved. The first one is a strong assumption on the probability distribution of input signals and the second one is that data coding is out of view of the paper.

In this paper, we consider this problem in an asymptotic situation of high resolution case and give an optimal quantization of signals. The key idea is a notion of density of the number of quantized subsections and by using calculus of variations, analytic solutions are derived. The solutions are simple functions of the distribution density of input signals and we can easily figure out the profile of the density of the umber of quantized subsections. Moreover, these results suggest several important insights on system identification under the condition of finite information. We illustrate such facts for some cases and discuss on the complexity of the problem of system identification.

2 Formulation

The objective of this paper is to show the effect of quantizers of I/O signals for system identification on its performance in analytic and intuitive form as possible. In general, the quantization error behaves as a random signal when the quantizer has enough high resolution, therefore, such condition has been often assumed in the area of signal processing. However, of course, the quantization error has strong correlation with the original quantized signal and in particular for system identification, several kinds of correlation are used for calculating the estimation. Therefore, such assumptions should be examined carefully. The strict analysis is desirable for system identification with the case of general model and mild assumption on I/O signals, however, it may be difficult to derive intuitive understanding results. Under these views, in [10], it is shown that an intuitive understanding optimal quantizer can be given analytically under strong assumption on input signal for a simple MA model. Moreover, it is shown that there exists a trade-off between quantization error and noise error under a condition of constant information used for system identification. In this paper, we basically follow this idea and consider system identification by least square criterion for a simple MA model. The plant is:

$$y_o(i) = q(y(i)) + w(i),\ y(i) = \phi(i)\theta \tag{1}$$

$$\phi(i) := [\, u(i) \quad u(i-1) \quad \cdots \quad u(i-n+1) \,],\ \theta := [\, \theta_1 \quad \theta_2 \quad \cdots \quad \theta_n \,]^{\mathrm{T}}, \tag{2}$$

where w is a noise and q is a quantizer of the original analogue output y. In [10], the boundedness of y or u is assumed, however, it is not necessarily assumed in this paper except for some cases. The quantizer q is defined by

$$q(y) := \mathrm{sgn}(y)\bar{y}_j,\ y \in S_j,\ \bar{y}_j \geq 0 \tag{3}$$

$$S_0 := \{y = 0\},\ S_j := \{y : d_{j-1} < y \le d_j\},\ j > 0,\ S_j := \{y : d_{j-1} \le y < d_j\},\ j < 0 \quad (4)$$

$$d_0 = 0 < d_1 < d_2 \cdots,\ d_{-1} = -d_1,\ d_{-2} = -d_2,\ \ldots\ , \quad (5)$$

where $\mathrm{sgn}(y)\bar{y}_j$ is the assigned quantized value to the subsection S_j. The quantizer is symmetrical with respect to the origin, and hereafter we may omit references on the negative section if they are obvious from the context.

The estimated parameter $\hat{\theta}$ using the least squares method with I/O data $u(i)$ and $y_o(i)$ is given by

$$\hat{\theta} = (U^{\mathrm{T}}U)^{-1}U^{\mathrm{T}}\left(\tilde{Y} + W\right), \quad (6)$$

where

$$U := [\,\phi(1)^{\mathrm{T}} \quad \phi(2)^{\mathrm{T}} \quad \cdots \quad \phi(N)^{\mathrm{T}}\,]^{\mathrm{T}},\ W := [\,w(1) \quad w(2) \quad \cdots \quad w(N)\,]^{\mathrm{T}},$$
$$\tilde{Y} := [\,\tilde{y}(1) \quad \tilde{y}(2) \quad \cdots \quad \tilde{y}(N)\,]^{\mathrm{T}},\ \tilde{y}(i) := q(y(i)). \quad (7)$$

Define the quantization error between $\tilde{y}$ and y by

$$e(i) := \tilde{y}(i) - y(i), \quad (8)$$

and the estimated parameter $\hat{\theta}$ can be written as

$$\begin{aligned} \hat{\theta} &= (U^{\mathrm{T}}U)^{-1}U^{\mathrm{T}}(U\theta + E + W) \\ &= \theta + \Delta E + \Delta W \end{aligned} \quad (9)$$

$$E := [\,e(1) \quad e(2) \quad \cdots \quad e(N)\,]^{\mathrm{T}},\ \Delta E := (U^{\mathrm{T}}U)^{-1}U^{\mathrm{T}}E,\ \Delta W := (U^{\mathrm{T}}U)^{-1}U^{\mathrm{T}}W. \quad (10)$$

This shows that the estimation error $\hat{\theta} - \theta$ can be evaluated from the magnitudes of ΔE and ΔW. The conventional, and reasonable, method to evaluate ΔW is to show the convergence rate of

$$N(U^{\mathrm{T}}U)^{-1} \overset{N\to\infty}{\longrightarrow} \frac{1}{\sigma_o^2}I,\ \frac{1}{N}U^{\mathrm{T}}W \overset{N\to\infty}{\longrightarrow} O, \quad (11)$$

where σ_o is the covariance of u, by using the mutual independence of the input signal u and the noise w. This methodology is also basically applicable to the case of ΔE, however, we should note that u and e are not independent in general, and the situation is much more complicated. The magnitude of the cross term $U^{\mathrm{T}}E$ is essential for reducing that of ΔE when $N \to \infty$ by using the fact:

$$(U^{\mathrm{T}}U)^{-1}U^{\mathrm{T}}E \overset{N\to\infty}{\longrightarrow} \frac{1}{\sigma_o^2}\frac{1}{N}U^{\mathrm{T}}E. \quad (12)$$

Therefore, the square error of ΔE also converges to

$$\mathrm{E}\left[\|\Delta E\|_2^2\right] = \mathrm{E}\left[\Delta E^{\mathrm{T}}\Delta E\right] \overset{N\to\infty}{\longrightarrow} \mathrm{E}\left[\frac{1}{\sigma_o^4}\sum_{k=0}^{n-1}\left(\frac{1}{N}\sum_{i=1}^{N}u(i-k)e(i)\right)^2\right]. \quad (13)$$

The right hand side of (13) is rewritten by

$$\begin{aligned} \mathrm{E}\left[\frac{1}{\sigma_o^4}\sum_{k=0}^{n-1}\left(\frac{1}{N}\sum_{i=1}^{N}u(i-k)e(i)\right)^2\right] &= \frac{1}{\sigma_o^4}\mathrm{E}\left[\sum_{k=0}^{n-1}\left(\frac{1}{N}\sum_{i=1}^{N}u(i-k)e(i)\right)^2\right] \\ &= \frac{1}{\sigma_o^4}\frac{1}{N}\mathrm{E}\left[\sum_{k=1}^{n}\phi_k^2(i)e^2(i)\right]. \end{aligned} \quad (14)$$

Similarly, an element of the quantization error term is given by

$$\mathrm{E}\left[\Delta E_k^2\right] \;=\; \frac{1}{\sigma_o^4}\frac{1}{N}\mathrm{E}\left[\phi_k^2(i)e^2(i)\right]. \tag{15}$$

The expectation in the right hand side of (14) or (15) is written by

$$\mathrm{E}\left[\sum_{k=1}^{n}\phi_k^2(i)e^2(i)\right] \;=\; \int\left(\sum_{k=1}^{n}\phi_k^2(i)\right)e^2(i)f_o(\phi_1(i),\ldots,\phi_n(i))d\phi_1(i)\cdots d\phi_n(i). \tag{16}$$

or

$$\mathrm{E}\left[\phi_k^2(i)e^2(i)\right] \;=\; \int\phi_k^2(i)e^2(i)f_o(\phi_1(i),\ldots,\phi_n(i))d\phi_1(i)\cdots d\phi_n(i), \tag{17}$$

where f_o is the joint density of ϕ_1, ϕ_2, $\ldots$, ϕ_n. In the following, we consider to find minimizing quantizer for (16) or (17).

In order to calculate the expectation (16) or (17), we define subsets Φ_j of the regression vector ϕ associated with the subsection S_j by

$$\Phi_j := \{\phi : y = \phi\theta \in S_j\}. \tag{18}$$

Associated with Φ_j, we consider the following variable transformation [10]:

$$y \;=\; \phi\theta = \phi T\cdot T^{-1}\theta =: \tilde{\phi}\tilde{\theta},\; T^{-1}\theta = \begin{bmatrix}\tilde{\theta}_1\\ O\end{bmatrix}, \tag{19}$$

where T is an orthogonal matrix. Then, Φ_j is represented as

$$\Phi_j = \left\{\phi : \tilde{\phi}_1\tilde{\theta}_1 \in (d_{j-1},\; d_j]\right\},\; j>0. \tag{20}$$

We also define subsections on the space of $\tilde{\phi}_1$:

$$I_j := \left\{\tilde{\phi}_1 : \tilde{\phi}_1\tilde{\theta}_1 \in (d_{j-1},\; d_j]\right\},\; j>0, \tag{21}$$

then, the subsections S_j, Φ_j, and I_j correspond to each other, and the probability distribution of y depends only on that of $\tilde{\phi}_1$. Therefore, in order to analyse the probability distribution of y, the variable $\tilde{\phi}_1$ and its subsection I_j are convenient to deal with. Moreover, by using the orthogonal transformation of ϕ, (16) is also given by

$$\begin{aligned}&\int\left(\sum_{k=1}^{n}\phi_k^2(i)\right)e^2(i)f_o(\phi_1(i),\ldots,\phi_n(i))d\phi_1(i)\cdots d\phi_n(i)\\ &\quad=\;\int\left(\sum_{k=1}^{n}\tilde{\phi}_k^2(i)\right)e^2(i)f_o(\tilde{\phi}_1(i),\ldots,\tilde{\phi}_n(i))d\tilde{\phi}_1(i)\cdots d\tilde{\phi}_n(i),\end{aligned} \tag{22}$$

where $f_o(\tilde{\phi}_1,\tilde{\phi}_2,\ldots,\tilde{\phi}_n)$ is the joint density of $\tilde{\phi}_1$, $\tilde{\phi}_2$, $\ldots$, $\tilde{\phi}_n$. Similarly, the quantization error for $\tilde{\theta}_1$ is calculated by

$$\int\tilde{\phi}_1^2(i)e^2(i)f_o(\tilde{\phi}_1(i),\ldots,\tilde{\phi}_n(i))d\tilde{\phi}_1(i)\cdots d\tilde{\phi}_n(i). \tag{23}$$

Here let $\overline{\phi}$ denote

$$\overline{\phi} := [\,\tilde{\phi}_2 \quad \tilde{\phi}_3 \quad \cdots \quad \tilde{\phi}_n\,]^{\mathrm{T}},$$

then, the marginal distribution density $f(\tilde{\phi}_1)$ on the space of $\tilde{\phi}_1$ is defined by

$$f(\tilde{\phi}_1) := \int f_o([\ \tilde{\phi}_1\ \ \overline{\phi}^{\mathrm{T}}\ ]^{\mathrm{T}}) d\overline{\phi}.$$

The important point is that if the distribution of the regressor vector ϕ is given, it is possible to derive the distribution density $f(\tilde{\phi}_1)$ analytically or numerically in an enough accuracy if necessary. With the fact that the distribution of e is only given by that of $\tilde{\phi}_1$, then (22) is represented by

$$\int e^2 \left(\sum_{k=1}^{n} \tilde{\phi}_k^2 \right) f_o(\tilde{\phi}_1, \ldots, \tilde{\phi}_n) d\tilde{\phi}_1 \cdots d\tilde{\phi}_n = \int e^2 \cdot \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) d\tilde{\phi}_1,$$

where $\sigma(\tilde{\phi}_1)$ is a variance of f_o at $\tilde{\phi}_1$ defined by

$$\int \left(\sum_{k=1}^{n} \tilde{\phi}_k^2 \right) f_o(\tilde{\phi}_1, \ldots, \tilde{\phi}_n) d\tilde{\phi}_2 \cdots d\tilde{\phi}_n =: \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) \tag{24}$$

Similarly, (23) is calculated by

$$\int \tilde{\phi}_1^2(i) e^2(i) f_o(\tilde{\phi}_1(i), \ldots, \tilde{\phi}_n(i)) d\tilde{\phi}_1(i) \cdots d\tilde{\phi}_n(i) = \int \tilde{\phi}_1^2(i) e^2(i) f(\tilde{\phi}_1(i)) d\tilde{\phi}_1(i). \tag{25}$$

On the other hand, the expectation of ΔE should be zero, therefore,

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{N} (\phi_k(i) \cdot e(i)) \right] &= N \sum_j \left(\int_{\Phi_j} \phi_k(i) \cdot e(i) f_o(\phi_1(i), \ldots, \phi_n(i)) d\phi_1(i) \cdots d\phi_n(i) \right) \\ &=: N \sum_j \mathrm{E}_{\Phi_j}(\phi_k(i) \cdot e(i)) = 0. \end{aligned} \tag{26}$$

With this in mind, we consider the next optimal quantization problem for the signals for system identification.

Problem 2.1 *For the system (1) with Assumption 3.1, give a quantizer q that minimizes the variance of (16) such that $\mathrm{E}_{I_j}(\phi_k(i) \cdot e(i)) = 0$ $(\forall j)$ under the constraint of quantization number.*

3 Special Case [10]

In [10], a special case of the problem is considered. At first assume $\tilde{\theta}_1 = 1$ just for the simplification of expressions and consider to minimize (25). The next is assumed.

Assumption 3.1 *$\tilde{\phi}_1$ obeys a uniform distribution in $[-\kappa,\ \kappa]$.*

Then, the following simplified problem is considered.

Problem 3.1 *For the system (1) with Assumption 3.1, give a quantizer q that minimizes the variance of (16) such that $\mathrm{E}_{I_j}(\tilde{\phi}_1(i) \cdot e(i)) = 0$ $(\forall j)$ under the constraint of quantization number of $[-\kappa,\ \kappa]$.*

As described in Section 2, the quantization scheme of $[-\kappa,\ \kappa]$ on y is essentially equal to that on $\tilde{\phi}_1$ and it is composed of the setting of the subsections $I_{-M}, \ldots, I_{-2}, I_{-1}, I_0, I_1, I_2, \ldots, I_M$, and the quantized values

$$\begin{aligned} & q(y),\ y \in S_j \\ = \ & q(\tilde{\phi}_1),\ \tilde{\phi}_1 \in I_j \\ = \ & \tilde{y}_j \end{aligned} \tag{27}$$

for each subsection I_j. We define a ratio r_j for I_j $(= (d_{j-1},\ d_j])$, I_{j+1} $(= (d_j,\ d_{j+1}])$ where their boundaries d_j, d_{j+1} have a relation:

$$d_j = r_j d_{j+1},\ r_j \in [0, 1]. \tag{28}$$

From its definition, $r_1, r_2, \ldots, r_M$ completely decide the quantization scheme and the optimization problem can be reduced to find the optimal ratios r_j.

We obtain the following result.

Proposition 3.1 [10] *The optimal ratios r_j that minimize the sum of the variances of $I_1, I_2, \ldots, I_M$, and also $I_{-1}, I_{-2}, \ldots, I_{-M}$, are given by solving the following optimization problem iteratively.*

$$r_j := \arg \min_{r \in [0,\ 1]} \Omega_j(r) \tag{45}$$

$$\Omega_j(r) := \Omega_{j-1}^{\min} r^5 - 18(1-r)^5 + 45(1+r)^2(1-r)^3 + 5(1-r)^7(1+r)^{-2} \tag{46}$$

$$\Omega_j^{\min} := \Omega_j(r_j),\ \Omega_o^{\min} = 32 \tag{47}$$

The optimal value of the variance is given by

$$V_M(\tilde{\phi}_1 \cdot e) := \sum_{j=-M}^{M} V_{I_j} = \frac{1}{2160} \kappa^4 \Omega_M^{\min}. \tag{48}$$

Every ratio r_j can be explicitly given by (45) $\sim$ (47) iteratively. We can also derive the following series of lemmas.

Lemma 3.1 [10]

$$r_j < r_{j+1},\ \forall j > 0 \tag{49}$$

$$r_j \to 1,\ j \to \infty \tag{50}$$

Lemma 3.2 [10]

$$|I_j| > |I_{j+1}|,\ \forall j > 0 \tag{51}$$

Lemma 3.3 [10]

$$\prod_{j=1}^{\infty} \frac{1}{r_j} = \infty \tag{52}$$

Lemma 3.4 [10]

$$\Omega_M^{\min} \to \Psi_a^b(M),\ M \to \infty \tag{53}$$

where $a = -5 \cdot 3^{-\frac{5}{2}}$ *and* $b = \frac{3}{2}$*, and* $\Psi_a^b(m)$ *is a function of* m *defined as the solution of the following recurrence formula with an appropriate initial number* $\psi(0) = K$*.*

$$\psi(m) - \psi(m-1) = a\psi^b(m-1) \tag{54}$$

From (48), the variance $V_M(u \cdot e)$ at sufficiently large M satisfies

$$\begin{aligned} V_M(u \cdot e) &\leq 2\frac{1}{2160}\frac{1}{2\kappa}\kappa^5((-3/2+1)(-5 \cdot 3^{-\frac{5}{2}} + \nu)(M-1))^{\frac{1}{-3/2+1}} \\ &= A\kappa^4(M-1)^{-2},\ A := \frac{1}{2160}\left(5 \cdot 2^{-1} \cdot 3^{-\frac{5}{2}} - 2^{-1}\nu\right)^{-2}. \end{aligned} \tag{55}$$

4 Main Results

The important point of the results of [10] introduced in Section 3 is giving a quantizer which is strictly optimal regardless of the resolution of the quantization. On the other hand, the following problems or possible extensions can be pointed out:

1) extension of the minimization of $\|\Delta E_1\|_2^2$ to that of $\|\Delta E\|_2^2$

2) the assumption on the distribution of input signals is strong, that is, it is a uniform distribution with a special basis in a space of the regressor vector.

3) it is possible to reduce the necessary information by applying compression with data coding, therefore, we can extend the similar optimization problem of quantizers for the case of a constraint on that code length. However, in [10], such extension is not considered.

In this paper, we intend to solve these subjects and show that they are possible for an asymptotic situation of high resolution of the quantizer.

4.1 Constraint on the Number of Quantization Sections

The key idea to solve these problems is introducing the following concept on the distribution of quantization subsections.

Definition 4.1 *The quantity* $g(\tilde{\phi}_1)$ *which satisfies the following is called distribution density of the number of quantized subsections.*

$$g(\tilde{\phi}_1)d\tilde{\phi}_1 = \text{number of quantized subsections in } d\tilde{\phi}_1$$

From this definition, $g^{-1}(\tilde{\phi}_1)$ represents the width of the quantization step at $\tilde{\phi}_1$.

In [10], the quantized value for each subsection is strictly assigned to satisfy that the quantization error is zero in each subsection. Such consideration has significance in low resolution case of the quantization, however, the efficiency becomes trivial in high resolution case. In particular, at the asymptotic situation of $|I_i| \to 0$, the middle point is reasonable to be assigned as the quantized value. Therefore, we fix such quantized values in the following of this paper.

Then, we assume the following.

Assumption 4.1 *The marginal distribution density $f(\tilde{\phi}_1)$ and $g(\tilde{\phi}_1)$ are enough smooth.*

The "smoothness" means that it satisfies the next approximations. With the "smoothness" of the density $g(\tilde{\phi}_1)$, we assume that we can select the representative value $g_i \sim g(\tilde{\phi}_1)$ for the subsection I_i and then, we define the next.

$$p_i := \int_{I_i} f(\tilde{\phi}_1) d\tilde{\phi}_1 =: f_i g_i^{-1}$$

Moreover, by using the variance $\sigma(\tilde{\phi}_1)$ of f_o at $\tilde{\phi}_1$ defined in (24), we assume that an approximation of (16) can be derived as follows.

$$\begin{aligned}
& \int e^2 \left(\sum_{k=1}^{n} \tilde{\phi}_k^2 \right) f_o(\tilde{\phi}_1, \ldots, \tilde{\phi}_n) d\tilde{\phi}_1 \cdots d\tilde{\phi}_n = \int e^2 \cdot \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) d\tilde{\phi}_1 \\
= & \sum_i \int_{I_i} e^2 \cdot \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) d\tilde{\phi}_1 \\
\sim & \sum_i \int_{\tilde{\phi}_{1i} - \frac{1}{2} g_i^{-1}}^{\tilde{\phi}_{1i} + \frac{1}{2} g_i^{-1}} (\tilde{\phi}_{1i} - x)^2 \cdot \sigma^2(x) f(x) dx \sim \sum_i \int_{\tilde{\phi}_1 - \frac{1}{2} g_i^{-1}}^{\tilde{\phi}_1 + \frac{1}{2} g_i^{-1}} (\tilde{\phi}_{1i} - x)^2 \sigma^2(\tilde{\phi}_{1i}) f_i dx \\
\sim & \sum_i \frac{1}{12} g_i^{-3} \sigma^2(\tilde{\phi}_{1i}) f_i \sim \sum_i \int_{\tilde{\phi}_1 - \frac{1}{2} g_i^{-1}}^{\tilde{\phi}_1 + \frac{1}{2} g_i^{-1}} \frac{1}{12} g_i^{-2} \sigma^2(\tilde{\phi}_{1i}) f_i dx \\
\sim & \sum_i \int_{\tilde{\phi}_1 - \frac{1}{2} g_i^{-1}}^{\tilde{\phi}_1 + \frac{1}{2} g_i^{-1}} \frac{1}{12} g^{-2}(\tilde{\phi}_1) \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) d\tilde{\phi}_1 \\
= & \int \frac{1}{12} g^{-2}(\tilde{\phi}_1) \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) d\tilde{\phi}_1
\end{aligned}$$

From the above approximation, the original optimization problem:

$$\min \mathrm{E} \left[\|\Delta E\|_2^2 \right]$$

can be replaced by the following when $N \to \infty$ and high resolution case:

$$g_{\text{opt-n}}(\tilde{\phi}_1) := \arg\min_g \int \mathcal{F}(g(\tilde{\phi}_1),\ G(\tilde{\phi}_1)) d\tilde{\phi}_1 \tag{56}$$

$$\text{such that} \quad G(\tilde{\phi}_1^{\min}) = 0,\ G(\tilde{\phi}_1^{\max}) = M, \tag{57}$$

where

$$\mathcal{F}(g(\tilde{\phi}_1),\ G(\tilde{\phi}_1)) := \left(\frac{1}{g(\tilde{\phi}_1)} \right)^2 \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) \tag{58}$$

$$\frac{d}{d\tilde{\phi}_1} G(\tilde{\phi}_1) = g(\tilde{\phi}_1). \tag{59}$$

We can derive the next result.

Proposition 4.1 *The solution of (56) is:*

$$
\begin{aligned}
g_{\text{opt-n}}(\tilde{\phi}_1) &= K\sigma^{\frac{2}{3}}(\tilde{\phi}_1)f^{\frac{1}{3}}(\tilde{\phi}_1) && (60)\\
K &= D^{-1}M && (61)\\
D &= \int \sigma^{\frac{2}{3}}(\tilde{\phi}_1)f^{\frac{1}{3}}d\tilde{\phi}_1. && (62)
\end{aligned}
$$

Moreover, the optimized value is given by

$$
\int \mathcal{F}(g_{\text{opt-n}}(\tilde{\phi}_1),\ G_{\text{opt-n}}(\tilde{\phi}_1))d\tilde{\phi}_1 = D^3M^{-2}. \tag{63}
$$

Proof By using calculus of variations, the optimal solution can be given. That is, from the following Euler–Lagrange's equation:

$$
\frac{d}{d\tilde{\phi}_1}\left(\frac{\partial \mathcal{F}}{\partial g}\right) - \frac{\partial \mathcal{F}}{\partial G} = 0,
$$

we can get a differential equation:

$$
\frac{d}{d\tilde{\phi}_1}\left(-2g^{-3}(\tilde{\phi}_1)\sigma^2(\tilde{\phi}_1)f(\tilde{\phi}_1)\right) = 0,
$$

and by solving it, we get

$$
g(\tilde{\phi}_1) = K\sigma^{\frac{2}{3}}(\tilde{\phi}_1)f^{\frac{1}{3}}(\tilde{\phi}_1),\quad K : \text{constant}. \tag{64}
$$

The constant number K is directly calculated by the condition (57), on the hand, the value of the objective function is derived as follows.

$$
\begin{aligned}
\int \mathcal{F}(g_{\text{opt-n}}(\tilde{\phi}_1),\ G_{\text{opt-n}}(\tilde{\phi}_1))d\tilde{\phi}_1 &= \int (K\sigma^{\frac{2}{3}}(\tilde{\phi}_1)f^{\frac{1}{3}}(\tilde{\phi}_1))^{-2}\sigma^2(\tilde{\phi}_1)f(\tilde{\phi}_1)d\tilde{\phi}_1\\
&= \int K^{-2}\sigma^{\frac{2}{3}}(\tilde{\phi}_1)f(\tilde{\phi}_1)^{\frac{1}{3}}d\tilde{\phi}_1 = K^{-2}D\\
&= D^3M^{-2} && (65)
\end{aligned}
$$

□

From this result, the asymptotic optimal quantization at high resolution case, is easily calculated if the marginal distribution of the regressor vector $f(\tilde{\phi}_1)$ is known.

Note 4.1 When f_o is a multidimensional normal distribution:

$$
f_o(\tilde{\phi}_1, \tilde{\phi}_2, \ldots, \tilde{\phi}_n) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Gamma)^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\tilde{\phi}^{\mathrm{T}}\Gamma^{-1}\tilde{\phi}\right),\ \Gamma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n),
$$

then

$$
\sigma^2(\tilde{\phi}_1) = \tilde{\phi}_1^2 + \sum_{i=2}^{n}\sigma_i^2.
$$

This means that when n, the order of the MA model, is large, $\sigma(\tilde{\phi}_1)$ can be approximated as a constant. If $\sigma_i = \sigma_o$, $\forall i$, then,

$$
\sigma(\tilde{\phi}_1) \sim n^{\frac{1}{2}}\sigma_o, \tag{66}
$$

therefore,

$$
D = n^{\frac{1}{3}}\sigma_o^{\frac{2}{3}}\int f^{\frac{1}{3}}d\tilde{\phi}_1
$$

and

$$g_{\mathrm{opt-n}}(\tilde{\phi}_1) = M\left(\int f^{\frac{1}{3}} d\tilde{\phi}_1\right)^{-1} f^{\frac{1}{3}},$$

$$\begin{aligned}\int \mathcal{F}(g_{\text{opt-n}}(\tilde{\phi}_1),\, G_{\text{opt-n}}(\tilde{\phi}_1)) d\tilde{\phi}_1 &= \left(\int f^{\frac{1}{3}} d\tilde{\phi}_1\right)^3 n\sigma_o^2 M^{-2} \\ &= 6\sqrt{3}\pi n\sigma_o^4 M^{-2} \sim 10.39\pi n\sigma_o^4 M^{-2}. \end{aligned} \tag{67}$$

We illustrate $g_{\text{opt-e}}$ for the cases that $f(\tilde{\phi}_1)$ is normal distribution, uniform distribution and power law as follows.

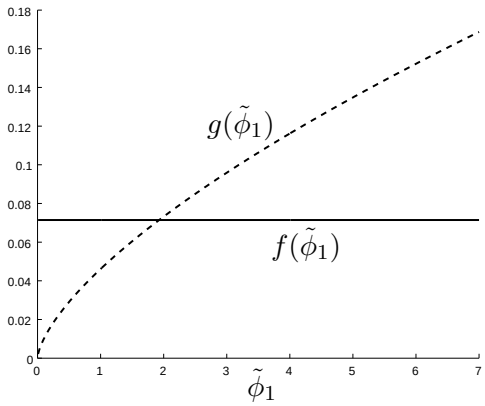

Fig. 1: Uniform distribution $f(\tilde{\phi}_1)$ (solid line) and the optimal distribution $g(\tilde{\phi}_1)$ (dashed line) in the case $\sigma(\tilde{\phi}_1) = \tilde{\phi}_1$

In Section 3, we introduce the strictly optimal quantization for any case of resolution when $f(\tilde{\phi}_1)$ is uniform distribution. Lemma 3.2 shows that the optimal quantization is coarse around the origin of $\tilde{\phi}_1$, on the other hand, it is high resolution near the bound of $\tilde{\phi}_1$. Such feature of the optimal quantization can be seen in this proposition which is for high resolution case (see Fig. 1). Fig. 1 is an example of a simple case $\sigma(\tilde{\phi}_1) = \tilde{\phi}_1$, and the proposition shows that the growing rate of the resolution against $\tilde{\phi}_1$ is known when $\sigma(\tilde{\phi}_1)$ is given analytically. In this case, the order of the growing rate is $\tilde{\phi}_1^{\frac{2}{3}}$, which is unknown from the results of the previous section.

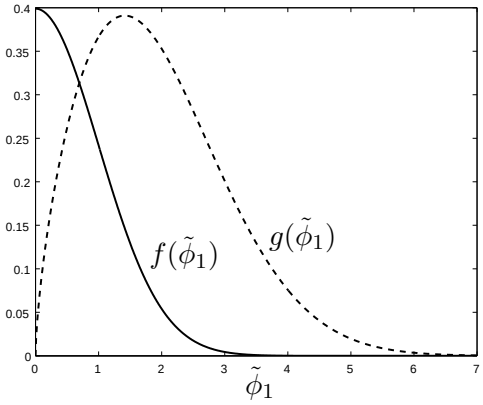

Fig. 2: Normal distribution $f(\tilde{\phi}_1)$ (solid line) and the optimal distribution $g(\tilde{\phi}_1)$ (dashed line) in the case $\sigma(\tilde{\phi}_1) = \tilde{\phi}_1$

In the case that $f(\tilde{\phi}_1)$ is normal distribution, the profile of the density $f(\tilde{\phi}_1)$ around the origin is flat, therefore, the optimal quantizer must have the similar feature of uniform distribution around the origin which case is discussed in Section 3 and the previous example. That is, the resolution should grow around it, and we can see such feature of $g_{\text{opt-e}}$. On the other hand, in the area of the tail of $f(\tilde{\phi}_1)$, $g(\tilde{\phi}_1)$ also goes down, however, against our intuition, the resolution remains high such as $g(3) \sim 0.201 \sim 51\%$ of $\max g(\tilde{\phi}_1)$ (~ 0.391) or $g(4) \sim 0.0758 \sim 19\%$ of $\max g(\tilde{\phi}_1)$, where $f(\tilde{\phi}_1)$ is enough small.

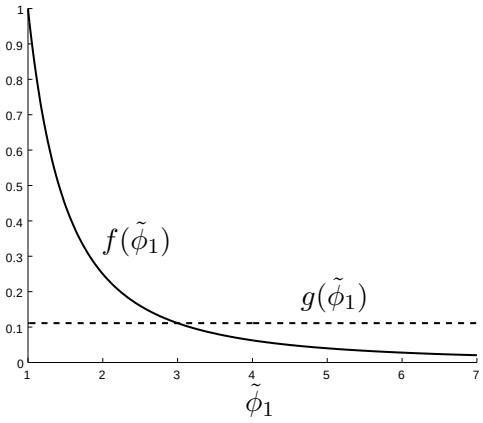

Fig. 3: Power law ($O(\tilde{\phi}_1^{-2})$) $f(\tilde{\phi}_1)$ (solid line) and the optimal distribution $g(\tilde{\phi}_1)$ (dashed line) in the case $\sigma(\tilde{\phi}_1) = \tilde{\phi}_1$

Next, we show the case of $f(\tilde{\phi}_1) \sim \tilde{\phi}_1^{-2}$ at enough large $\tilde{\phi}_1$ as an example of power law. In this case, $g_{\text{opt-n}}$ is constant and it is marginal for the existence of the solution. This result shows the difficulty of the system identification in an enough accuracy by using finite information on the system when the tail of the distribution density $f(\tilde{\phi}_1)$ is heavier than $O(\tilde{\phi}_1^{-2})$. This result explains the complexity of power law from the view point of parameter estimation of system identification.

4.2 Constraints on the Code Length of Signals

In the previous subsection, we derive an optimal quantizer to minimize the identification error under constraint of the number of quantization steps in the case of high resolution. The result is itself meaningful, however, for the purpose to reduce the information of the observed data from the identified system, it is more reasonable to apply coding for the quantized signals and measure the code length as the quantity of the information. According to this observation, we consider the minimization problem of identification error under constraint of the expectation of the optimal code length in high resolution case.

Let $E(\cdot)$ be an encoder which is a mapping from source alphabets to code alphabets and $l(\cdot)$ the code length. We regard $q(\tilde{\phi}_1)$ as the corresponding source alphabets, then, $l(E(q(\tilde{\phi}_1)))$ represents the code length of the code alphabets. The expectation of the code length for a quantized signal has relation with entropy from the following well-known proposition.

Proposition 4.2 [9, 2] *Let x be source alphabets, then,*

$$\mathrm{E}[l(E(x))] \geq H(x). \tag{68}$$

Moreover, there exists an optimal quantizer E which attains the lower bound.

By applying this proposition, the optimization problem of the quantizer for the code length is reduced to the same problem under constraint of entropy of the quantized signals.

On the other hand, the basic idea and tools to represent the quantizer in high resolution case are the same of the previous subsection. That is, under Assumption 4.1, we use g_i as the approximation of g in the subsection I_i. In this case, we can get the asymptotic approximation of the entropy of the quantized signal.

$$\begin{aligned}
H(f,g) &:= \sum_i -p_i \log p_i && (69)\\
&= \sum_i -\int_{I_i} f(\tilde{\phi}_1) d\tilde{\phi}_1 \log f_i g_i^{-1} && (70)\\
&= \int -f(\tilde{\phi}_1) \log f_i g_i^{-1} d\tilde{\phi}_1 && (71)\\
&\sim \int -f(\tilde{\phi}_1) \log \left(f(\tilde{\phi}_1) g^{-1}(\tilde{\phi}_1) \right) d\tilde{\phi}_1 && (72)\\
&= H(f) + \int -f(\tilde{\phi}_1) \log \left(g^{-1}(\tilde{\phi}_1) \right) d\tilde{\phi}_1 && (73)
\end{aligned}$$

By using this asymptotic approximation of the entropy (73), we consider the following problem for high resolution case.

$$\begin{aligned}
& g_{\text{opt-e}}(\tilde{\phi}_1) := \arg\min_g \int \mathcal{F}(g(\tilde{\phi}_1),\ G(\tilde{\phi}_1)) d\tilde{\phi}_1 && (74)\\
\text{such that} \quad & H(f,g) = \log M && (75)
\end{aligned}$$

Note that M is an expected number of quantization steps in the sense of (75) itself.

We can derive the following proposition.

Proposition 4.3 *The solution of (74) is:*

$$g_{\text{opt-e}}(\tilde{\phi}_1) = KM\sigma(\tilde{\phi}_1) \tag{76}$$

$$K = \exp L \tag{77}$$

$$L := -H(f) - \int f \log \sigma(\tilde{\phi}_1) d\tilde{\phi}_1 = \int f(\tilde{\phi}_1) \log \frac{f(\tilde{\phi}_1)}{\sigma(\tilde{\phi}_1)} d\tilde{\phi}_1 \tag{78}$$

Moreover, the optimized value is given by

$$\int \mathcal{F}(g_{\text{opt-e}}(\tilde{\phi}_1),\ G_{\text{opt-e}}(\tilde{\phi}_1)) d\tilde{\phi}_1 = K^{-2}M^{-2}. \tag{79}$$

Proof Let λ be a Lagrange multiplier and consider the minimization problem of the following quantity.

$$\begin{aligned}
&\int \mathcal{F}(g(\tilde{\phi}_1),\ G(\tilde{\phi}_1)) d\tilde{\phi}_1 + \lambda H(f, g) \\
= &\int \left(\frac{1}{g(\tilde{\phi}_1)}\right)^2 \sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) - \lambda f(\tilde{\phi}_1) \log\left(g^{-1}(\tilde{\phi}_1)\right) d\tilde{\phi}_1 + \lambda H(f) \\
= &\int f(\tilde{\phi}_1) \left(g^{-2}(\tilde{\phi}_1)\sigma^2(\tilde{\phi}_1) + \lambda \log g(\tilde{\phi}_1)\right) d\tilde{\phi}_1 + \lambda H(f)
\end{aligned}$$

From Eular–Lagrange's differential equation, we get

$$\frac{\partial}{\partial g}\left(g^{-2}\sigma^2(\tilde{\phi}_1) + \lambda \log g\right) = -2g^{-3}\sigma^2(\tilde{\phi}_1) + \lambda g^{-1} = \text{ constant }.$$

Fix the constant to be zero, then,

$$g = \left(\frac{2}{\lambda}\right)^{\frac{1}{2}} \sigma(\tilde{\phi}_1),$$

and by substituting it for $H(f,\ g)$, we get

$$H(f,\ g) = \int -f \log g^{-1} f d\tilde{\phi}_1 \tag{80}$$

$$= \log\left(\frac{2}{\lambda}\right)^{\frac{1}{2}} + \int -f \log \frac{f}{\sigma(\tilde{\phi}_1)} d\tilde{\phi}_1 = \log M. \tag{81}$$

Therefore,

$$\left(\frac{2}{\lambda}\right)^{\frac{1}{2}} = \exp\left(\int f \log \frac{f}{\sigma(\tilde{\phi}_1)} d\tilde{\phi}_1 + \log M\right), \tag{82}$$

and (76) is derived. By substituting $g_{\text{opt-e}}$ for the objective integral, the following is calculated.

$$\begin{aligned}
\int g_{\text{opt}-\text{e}}^{-2}(\tilde{\phi}_1)\sigma^2(\tilde{\phi}_1) f(\tilde{\phi}_1) d\tilde{\phi}_1 &= \frac{\lambda}{2} \\
&= K^{-2}M^{-2}
\end{aligned}$$

□

Note 4.2 The optimal $g_{\text{opt-e}}$ is a linear function of $\sigma(\tilde{\phi}_1)$ and its ratio is given by f and the expected number of the quantization steps M. On the other hand, the convergence rate of the minimized variance of

the quantization error term is M^{-2}. When f_o is a multidimensional normal distribution and n is large as considered in Note 4.1, by using (66),

$$L = -H(f) - \log(\sigma_o n^{\frac{1}{2}}),\ K = \exp(-H(f)) \cdot (\sigma_o n^{\frac{1}{2}})^{-1},$$

and

$$\begin{aligned} g_{\mathrm{opt-e}}(\tilde{\phi}_1) &= KM\sigma(\tilde{\phi}_1) \\ &= M \cdot \exp(-H(f)) \cdot (\sigma_o n^{\frac{1}{2}})^{-1} \cdot \sigma_o n^{\frac{1}{2}} \\ &= M \cdot \exp(-H(f)) \end{aligned}$$

$$\begin{aligned} \int \mathcal{F}(g_{\text{opt-e}}(\tilde{\phi}_1),\ G_{\text{opt-e}}(\tilde{\phi}_1))d\tilde{\phi}_1 &= \exp(2H(f))n\sigma_o^2 M^{-2} \\ &= 2e\pi n\sigma_o^4 M^{-2} \sim 5.44\pi n\sigma_o^4 M^{-2}. \end{aligned} \tag{83}$$

The comparison of (67) and (83) tells us that the case of the optimal coding attains about a half magnitude of the variance of the quantization error compared with $g_{\mathrm{opt-n}}$. Of course the definitions of M in Note 4.1 and here are slightly different, however, we can estimate the effect of coding on the total amount of information necessary for parameter estimation.

5 Conclusion

In this paper, we extended the results of optimal quantization problem for system identification in [10]. We consider two cases of the optimization: constraints on the number of quantization such as [10] and the code length. In [10], a strong condition is assumed, that is, the regressor vector has a special distribution density. On the other hand, in this paper, by employing a concept of quantization density for high resolution case, such assumption can be removed. We explicitly derived the optimal quantizations and the minimized quantization errors for these two cases. We also discussed on the general difficulties of parameter estimation with respect to the complexity of the stochastic processes.

References

[1] R. W. Brockett and D. Leberzon. Quantized feedback stabilization of linear systems. *IEEE Trans. Automat. Control*, AC-45-7:1279–1289, 2000.

[2] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley series in telecommunications. John Wiley & Sons, Inc., New York, 1991.

[3] R. E. Curry. *Estimation and control with quantized measurements*. M.I.T. Press, Cambridge, MA, 1970.

[4] D. F. Delchamps. Extracting state information from a quantized output record. *Systems & Control Letters*, 13:365–372, 1989.

[5] D. F. Delchamps. Stabilizing a linear system with quantized state feedback. *IEEE Trans. Automat. Control*, AC-35-8:916–924, 1990.

[6] N. Elia and S. K. Mitter. Stabilization of linear systems with limited information. *IEEE Trans. on Automatic Control*, AC-46-9:1384–1400, 2001.

[7] M. Gevers and G. Li. *Parametrization in control, estimation and filtering problems: Accuracy aspects*. Communications and control engineering series. Springer-Verlag, Berlin, 1993.

[8] G. N. Nair and R. J. Evans. Stabilization with data-rate-limited feedback: Tightest attainable bounds. *Systems and Control Letters*, 41:49–56, 2000.

[9] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423, 1948.

[10] K. Tsumura and J. Maciejowski. Optimal quantization of signals for system identification. *Technical Report of The Univ. Cambridge*, CUED/F-INFENG/TR445, 2002.

[11] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints – part I: State estimation problems. *IEEE Trans. Automat. Control*, AC-42-9:1294–1299, 1997.

[12] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints – II: Stabilization with limited information feedback. *IEEE Trans. Automat. Control*, AC-44-5:1049–1053, 1999.