Minecraft of System Modeling

Pavel Loskot
ZJU-UIUC Institute
Haining, China
pavelloskot@intl.zju.edu.cn

***Abstract*—The best selling computer game of all times, Minecraft, represents the world as discrete blocks. The Minecraft-like worlds may be unknowingly created by many mathematical models of the real-world systems, when their inputs and outputs are discretized. This paper investigates system modeling and identification with noisy, discretized, but otherwise static inputs and outputs. Such a scenario occurs, for example, when configuring and measuring the system is time-consuming and costly. The task is to infer the model parameters from a limited number of input-output measurements. It is shown that, in this setting, the traditional least-squares model fitting is ineffective. A better strategy is to first accurately estimate the static input and output values, and then obtain the model parameters by inverting the model numerically by solving an underlying set of equations for the same number of unknown model parameters. These results have direct implications on creating and interpreting mathematical models of systems, and even physical laws, when the noisy measurements are implicitly or explicitly discretized.**

***Keywords*—*Linear model; Mean-square error; Minecraft; Quantization; Parameter estimation; System identification*.**

I. INTRODUCTION

Mathematical models are used extensively in many applications. The models are usually represented by the sets of parameterized equations describing the model input-output relationships. The aim of model identification is to recover the model parameters from the noisy measurements of its inputs and outputs. These measurements may be explicitly or implicitly quantized. The former is used to reduce the storage and transmission requirements, and to speed-up computations at the expense of loosing some information and accuracy. The implicit quantization is more subtle, and it occurs when the resolution of measured samples is insufficient, for example, due to the use of inexpensive measuring equipment.

Simply inverting the model in order to recover the model parameters from the measurements of its inputs and outputs is often unacceptable. The model inversion tends to greatly amplify the measurement noises, which leads to large estimation errors [1]. The model-based parameter estimation methods are often used to obtain the optimum and numerically efficient estimators in the presence of strong measurement noises. However, for model identification [2] and supervised machine learning [3], an alternative strategy can be adopted. In particular, the input and output values can be estimated independently from their noisy, and possibly discretized measurements. For static values, this corresponds to estimating unknown constants in additive noises. If the estimators used are unbiased and consistent, the measurement noises can be sufficiently suppressed, so the model inversion is acceptable to accurately infer the model parameters.

The paper [4] is one of the earliest studies on estimating the state of dynamic linear systems from quantized measurements. The authors demonstrated that Kalman filtering is still effective even under these conditions. This problem was considered again in [5] as a joint design of the quantizer and the estimator. The classical textbook [2] covers a wide range of topics in adaptive filtering including system identification and adaptive filter design with quantized inputs. The paper [6] investigates the optimum techniques for signal detection and estimation, and evaluates the corresponding performance losses due to uniform signal quantization. The confidence intervals of the discretized likelihood-based estimators with quantized inputs were studied in [7]. The encoding and decoding schemes for quantized random processes were designed in [8] to enable their efficient transmissions under the age-of-information constraints. The Cramér-Rao bounds for estimating the parameters from quantized measurements were derived in [9].

In this paper, we consider the problem of identifying the model parameters from quantized noisy measurements of both the model inputs and outputs. The model inputs and outputs are assumed to be static, so their values can be inferred with a high accuracy from a sufficient number of measurements assuming the consistent and unbiased estimators. The model parameters are then obtained by solving a set of linear or non-linear equations. It is also shown that the traditional least squares fitting of the model to the input and output data is much less effective, when the input and output measurements are noisy and quantized. This is also an important issue, for example, in supervised machine learning.

The following notations are adopted in the paper: $\mathrm{Av}[\cdot] = (1/T)\int_{-T/2}^{T/2}(\cdot)\,\mathrm{d}t$, and, $\mathrm{Av}[\cdot] = (N+1)^{-1}\sum_{i=-N/2}^{N/2}(\cdot)$, are the time-averaging (arithmetic average) operators in continuous and discrete time, respectively, $\mathrm{E}[\cdot]$ is the statistical expectation, $\boldsymbol{x}$ denotes a column vector, whereas $\boldsymbol{X}$ denotes a matrix, $(\cdot)^T$ and $(\cdot)^{-1}$ denote the matrix transpose and inverse, respectively, $\langle\cdot,\cdot\rangle$ denotes the dot-product, $\dot{f}$ is the first derivative of function, f, $\lfloor\cdot\rfloor$, $\lceil\cdot\rceil$, and $\mathrm{sign}(\cdot)$ are the floor function, the ceiling function, and the sign function, respectively, and $\mathbb{R}$ and $\mathbb{Z}$ represent the sets of real numbers and integers, respectively.

The rest of the paper is organized as follows. Section II outlines system model with uniformly, and also binary quantized inputs and outputs. The estimation of model parameters is described in Section III. The estimator variances are studied in Section IV. Discussion and future work are in Section V.

II. SYSTEM MODEL

A general parameterized model with multiple inputs and outputs (MIMO) is shown in Figure 1. Such a model can be succinctly described by a single equation,

$$f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{a}) = 0 \tag{1}$$

relating the model inputs, $\boldsymbol{x}$, outputs, $\boldsymbol{y}$, and a given set of model parameters, $\boldsymbol{a}$. Importantly, it is assumed that the input as well as output measurements of model (1) are first quantized and de-noised, before estimating the parameters, $\boldsymbol{a}$.

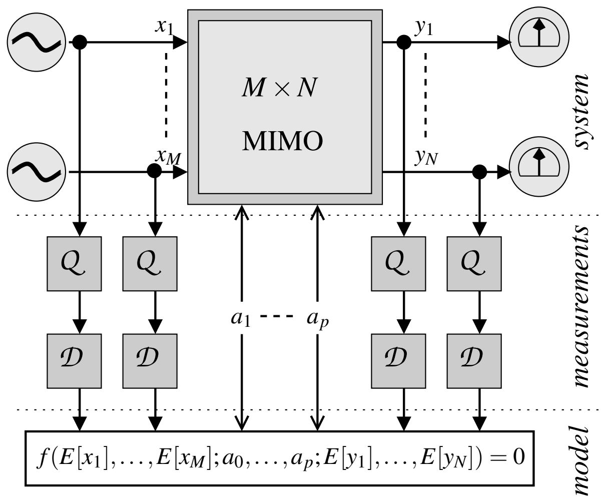

Figure 1. Modeling and measurements of a static $(M \times N)$ MIMO system.

Note that, here, the system modeling assumes the expected values of the inputs and outputs. In practice, measuring the statistical means can be problematic, when the random processes are non-stationary or non-ergodic [10]. The measuring instruments usually report the time-averaged values over a certain time-window. On the other hand, the expected values are more a theoretical concept, which is used, for example, when deriving the estimators of random signals to minimize the given risk. However, under the law of large numbers, the expectations can be replaced by the time averages. These differing views and assumptions can be reconciled by assuming the statistical and time averaging at the same time, i.e., by assuming, $\mathrm{Av}[\mathrm{E}[\cdot]] = \mathrm{E}[\mathrm{Av}[\cdot]]$. Depending on the type of a random process, $x(t)$, different averages are related as:

$$\begin{array}{lcc}
\mathrm{E}[x] = \mathrm{Av}[\mathrm{E}[x]] = \mathrm{Av}[x] & \Leftrightarrow & \begin{array}{c}\text{ergodic \&}\\ \text{stationary,}\end{array}\\
\mathrm{E}[x] \neq \mathrm{Av}[\mathrm{E}[x]] = \mathrm{Av}[x] & \Leftrightarrow & \begin{array}{c}\text{ergodic \&}\\ \text{non}-\text{stationary,}\end{array}\\
\mathrm{E}[x] = \mathrm{Av}[\mathrm{E}[x]] \neq \mathrm{Av}[x] & \Leftrightarrow & \begin{array}{c}\text{non}-\text{ergodic \&}\\ \text{stationary,}\end{array}\\
\mathrm{E}[x] \neq \mathrm{Av}[\mathrm{E}[x]] \neq \mathrm{Av}[x] & \Leftrightarrow & \begin{array}{c}\text{non}-\text{ergodic \&}\\ \text{non}-\text{stationary.}\end{array}
\end{array} \tag{2}$$

A. Linear SISO model

For the sake of notational simplicity, consider a single-input, single-output (SISO) model.

The linear SISO model is described by a linear combination of p basis functions, $\phi_i(x)$, i.e.,

$$y = a_0 + \sum_{i=1}^{p} a_i\,\phi_i(x). \tag{3}$$

If the functions, $\phi_i(x)$, are mutually orthogonal, i.e., the dot-product, $\langle \phi_i, \phi_j \rangle \neq 0$, for $\forall i \neq j$, then p is also the dimension (rank) of the linear model. The n output measurements, y_i, collected at n input values, x_i, are related as,

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & \phi_1(x_1) & \cdots & \phi_p(x_1) \\ \vdots & \vdots & & \vdots \\ 1 & \phi_1(x_n) & \cdots & \phi_p(x_n) \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} \tag{4}$$
$$\boldsymbol{y} = \boldsymbol{\Phi}(\boldsymbol{x}) \cdot \boldsymbol{a}.$$

The basis functions are generally non-linear, however, they can be linearized about a chosen value, x_0, as,

$$\phi_i(x) \doteq \phi_i(x_0) + \dot{\phi}_i(x_o)(x - x_0). \tag{5}$$

Such linear approximations can be also defined in multiple dimensions [11]. The caveat is that the approximation (5) is only valid in the vicinity of x_0, and choosing the suitable value can be problematic. For example, if linear model (3) represents a polynomial regression, then it can be rewritten assuming the linearized basis functions as,

$$y = a_0 + \sum_{i=1}^{p} a_i(A_i x + B_i) \tag{6}$$

where $A_i = \dot{\phi}_i(x_0)$, and, $B_i = \phi_i(x_0) - \dot{\phi}_i(x_0)x_0$.

B. Quantized measurements

The measurements are quantized for various reasons. For instance, the explicitly quantized values require less memory for storage, and the numerical computations become faster to perform. The implicit quantization occurs when the resolution of the measurements is insufficient with respect to a given modeling application. The most common is a uniform quantization having the equidistant quantization intervals of length, Δ, i.e.,

$$\check{x} = \mathcal{Q}(x) = \left\lfloor \frac{x - \Delta/2}{\Delta} \right\rfloor + 1 \in \mathbb{Z} \tag{7}$$

so that the quantization error, $\varepsilon_\Delta = x - \Delta\check{x}$, and,

$$\Delta(\check{x} - 1/2) \leq x < \Delta(\check{x} + 1/2). \tag{8}$$

Note also that, $\lfloor x \rfloor + 1 \neq \lceil x \rceil$, for the integer values of the argument. In addition, the quantized values are often bounded to a finite set of integers between, $-\check{x}_{\max}$, and, $\check{x}_{\max}$.

Alternatively, the binary quantization,

$$\check{x} = \mathcal{Q}_2(x) = \mathrm{sign}(x) \in \{-1, +1\} \tag{9}$$

can be sufficient in some applications.

The issue with implicit quantization due to insufficient resolution is illustrated in Figure 2, assuming a linear system, $y = 3x/2$, and the uniform quantization with $\Delta = 1/2$. It can be observed that, the model having only the quantized inputs, $y = aQ(x)$, is nearly identical to the unquantized model, $y = ax$. However, when both the input and the output are quantized, a formerly linear model becomes a staircase function (red dashed line), $Q(y) = aQ(x)$. In this case, only one noise-free measurement is necessary to determine the constant, a. If such a measurement is taken at points, A, B, or C, the proportionality constant is inferred to be equal to 1, 5/4, or 7/4, respectively. Consequently, the implicit or explicit quantization of the output values have a severe impact on identifying the model parameters.

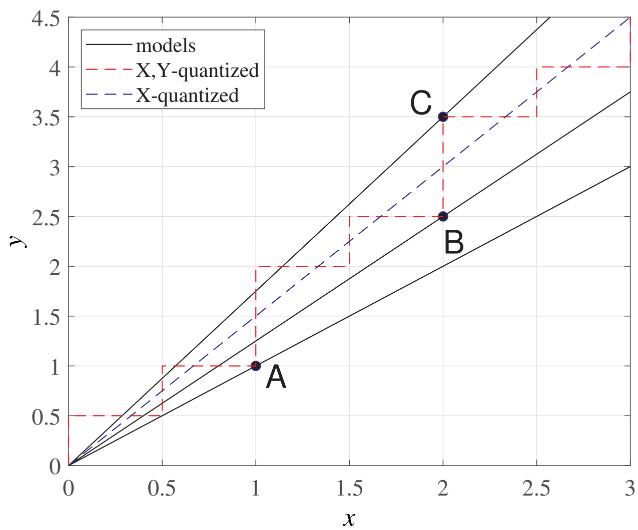

Figure 2. The consequences of the input-output uniform quantization on modeling linear SISO systems.

III. Estimating Model Parameters

Assume that n noisy measurements can be written as,

$$\begin{aligned} y_i &= \bar{y} + \varepsilon_{yi} \\ x_i &= \bar{x} + \varepsilon_{xi} \end{aligned} \tag{10}$$

where the additive noises, ε_{yi}, and, ε_{xi}, have zero means, i.e., $\mathrm{E}[y_i] = \bar{y}$, and, $\mathrm{E}[x_i] = \bar{x}$. Even when the measurements at different time instances can be assumed to be independent, the input-output correlations,

$$\mathrm{E}[x_i y_i] = \bar{x}\bar{y} + \mathrm{E}[\varepsilon_{xi}\varepsilon_{yi}] \tag{11}$$

are affected by the noise covariances, $\mathrm{E}[\varepsilon_{xi}\varepsilon_{yi}] \neq 0$.

Provided that the measurements are noisy, the input-output relationship (4) can only be satisfied approximately. The over-determined linear systems with $n \gg p$ measurements can be solved by considering the least-squares (LS) model fitting. The closed-form expression for the LS estimate of the model parameters is well-known, i.e., [12]

$$\hat{\boldsymbol{a}}_{\mathrm{LS}} = \left(\boldsymbol{\Phi}^T(\boldsymbol{x})\boldsymbol{\Phi}(\boldsymbol{x})\right)^{-1}\boldsymbol{\Phi}^T(\boldsymbol{x})\boldsymbol{y}. \tag{12}$$

Substituting the noisy measurements (10) into (12), while also assuming a linearization of the basis functions (5) about the mean, $\bar{x}$, the resulting linear model (4) can be written as,

$$\boldsymbol{y} = \left[\mathbf{1}_{(n,1)} \mid \bar{\boldsymbol{\Phi}}(\bar{x}) + \boldsymbol{\varepsilon}_x \cdot \dot{\boldsymbol{\phi}}^T(\bar{x})\right] \cdot \boldsymbol{a} \tag{13}$$

where $\mathbf{1}_{(n,1)}$ is the all-ones column vector, the constant matrix, $\bar{\boldsymbol{\Phi}}(\bar{x})$, has identical rows with the elements, $\phi_i(\bar{x})$, the column vector, $\boldsymbol{\varepsilon}_x$, contains additive noises, ε_{xi}, at the model input, and the constant column vector, $\dot{\boldsymbol{\phi}}(\bar{x})$, has the elements, $\dot{\phi}_i(\bar{x})$.

In order to obtain an insight into the LS solution of (13) for the model parameters, $\boldsymbol{a}$, consider the LS sum over the n measurements, i.e.,

$$\mathrm{LS}(a_0, \boldsymbol{a}) = \sum_{i=1}^{n}\left(y_i - a_0 - \left(\dot{\boldsymbol{\phi}}\varepsilon_{xi} + \boldsymbol{\phi}\right)^T \cdot \boldsymbol{a}\right)^2 \tag{14}$$

where the parameter, a_0, was taken out of the p-element vector, $\boldsymbol{a}$, $\boldsymbol{\phi}$ represents the row of the matrix, $\bar{\boldsymbol{\Phi}}(\bar{x})$, transposed to become a column vector, and let the vector of derivatives, $\dot{\boldsymbol{\phi}}(\bar{x}) \equiv \dot{\boldsymbol{\phi}}$. Note that both vectors, $\dot{\boldsymbol{\phi}}$, and, $\boldsymbol{\phi}$, are independent of the index, i. The model parameters minimizing the LS value are the solution of the set of linear equations, i.e.,

$$\begin{aligned} \frac{\partial}{\partial a_0}\mathrm{LS}(\hat{a}_0, \hat{\boldsymbol{a}}) &= 0 \\ \frac{\partial}{\partial \boldsymbol{a}}\mathrm{LS}(\hat{a}_0, \hat{\boldsymbol{a}}) &= \mathbf{0}. \end{aligned} \tag{15}$$

After some lengthy, but otherwise straightforward manipulations, we get,

$$\hat{a}_0 = \mathrm{Av}[y_i] - \left(\dot{\boldsymbol{\phi}}\mathrm{Av}[\varepsilon_{xi}] + \boldsymbol{\phi}\right)^T \hat{\boldsymbol{a}} \tag{16}$$

where $\mathrm{Av}[y_i] = (1/n)\sum_{i=1}^{n} y_i$, and, $\mathrm{Av}[\varepsilon_{xi}] = (1/n)\sum_{i=1}^{n}\varepsilon_{xi}$. Noticing that, $y_i - \mathrm{Av}[y_i] = \varepsilon_{yi}$, we obtain the solution for $\boldsymbol{a}$, which can be substituted into (16), i.e.,

$$\dot{\boldsymbol{\phi}}\,\mathrm{Av}[\varepsilon_{xi}\varepsilon_{yi}] + \boldsymbol{\phi}\mathrm{Av}[\varepsilon_{yi}] = \left(\mathrm{Av}\left[\left(\dot{\boldsymbol{\phi}}\,\varepsilon_{xi} + \boldsymbol{\phi}\right)\left(\dot{\boldsymbol{\phi}}\,\varepsilon_{xi} + \boldsymbol{\phi}\right)^T\right] - \mathrm{Av}\left[\dot{\boldsymbol{\phi}}\,\varepsilon_{xi} + \boldsymbol{\phi}\right]\mathrm{Av}\left[\dot{\boldsymbol{\phi}}\,\varepsilon_{xi} + \boldsymbol{\phi}\right]^T\right)\boldsymbol{a}. \tag{17}$$

The right-hand side of (17) can be further simplified as,

$$\dot{\boldsymbol{\phi}}\,\mathrm{Av}[\varepsilon_{xi}\varepsilon_{yi}] + \boldsymbol{\phi}\mathrm{Av}[\varepsilon_{yi}] = \dot{\boldsymbol{\phi}}\dot{\boldsymbol{\phi}}^T\mathrm{Av}\left[(\varepsilon_{xi} - \bar{\varepsilon}_x)^2\right]\boldsymbol{a} \tag{18}$$

where $\bar{\varepsilon}_x = \mathrm{Av}[\varepsilon_{xi}]$. Finally, the LS estimates of the model parameters are then computed as,

$$\hat{\boldsymbol{a}} = \left(\dot{\boldsymbol{\phi}}\dot{\boldsymbol{\phi}}^T\right)^{-1}\dot{\boldsymbol{\phi}}\frac{\mathrm{Av}[\varepsilon_{xi}\varepsilon_{yi}]}{\mathrm{Av}[(\varepsilon_{xi} - \bar{\varepsilon}_x)^2]} + \left(\dot{\boldsymbol{\phi}}\dot{\boldsymbol{\phi}}^T\right)^{-1}\boldsymbol{\phi}\frac{\mathrm{Av}[\varepsilon_{yi}]}{\mathrm{Av}[(\varepsilon_{xi} - \bar{\varepsilon}_x)^2]}. \tag{19}$$

For a large number of samples, $n \gg 1$, $\mathrm{Av}[\varepsilon_{yi}] \doteq 0$, and the final expression for estimating the model parameters becomes,

$$\hat{\boldsymbol{a}} = \left(\dot{\boldsymbol{\phi}}\dot{\boldsymbol{\phi}}^T\right)^{-1}\dot{\boldsymbol{\phi}}\frac{\mathrm{Av}[\varepsilon_{xi}\varepsilon_{yi}]}{\mathrm{Av}[(\varepsilon_{xi} - \bar{\varepsilon}_x)^2]}. \tag{20}$$

As an illustrative example, assume a simple linear SISO model, $y_i = a_1 x_i + a_0$, with $p = 2$ parameters. Assuming (20), the LS estimates of the model parameters are,

$$\begin{aligned} \hat{a}_0 &= \bar{y} - \hat{a}_1\bar{x} \\ \hat{a}_1 &= \frac{\mathrm{Av}[(y_i - \bar{y})(x_i - \bar{x})]}{\mathrm{Av}[(x_i - \bar{x})^2]} \end{aligned} \tag{21}$$

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-245-6

where $\bar{y} = \mathrm{Av}[y_i]$, and, $\bar{x} = \mathrm{Av}[x_i]$. The resulting mean-square error (MSE) is equal to,

$$\begin{aligned}\mathrm{MSE}(\hat{a}_0, \hat{a}_1) &= \sum_{i=1}^{n} (y_i - \hat{a}_1 x_i - \hat{a}_0)^2 \\ &= \sum_{i=1}^{n} ((y_i - \bar{y}) - \hat{a}_1 (x_i - \bar{x}))^2 \\ &= \mathrm{Av}\left[(y_i - \bar{y})^2\right] - \frac{\mathrm{Av}[(x_i - \bar{x})(y_i - \bar{y})]^2}{\mathrm{Av}[(x_i - \bar{x})^2]}.\end{aligned} \tag{22}$$

Moreover, for the specific model of measurements (10), and an asymptotically large number of measurements, $n \gg 1$, the LS estimate of a_1 can be rewritten as,

$$\hat{a}_1 = \frac{\mathrm{E}[\varepsilon_{xi}\varepsilon_{yi}]}{\mathrm{E}\left[\varepsilon_{xi}^2\right]} = \frac{\mathrm{cov}[\varepsilon_{xi}\varepsilon_{yi}]}{\mathrm{var}[\varepsilon_{xi}]}. \tag{23}$$

In this case, the resulting MSE is equal to,

$$\mathrm{MSE}(\hat{a}_0, \hat{a}_1) = \mathrm{E}\left[\varepsilon_{yi}^2\right] - \frac{\mathrm{E}[\varepsilon_{xi}\varepsilon_{yi}]^2}{\mathrm{E}\left[\varepsilon_{xi}^2\right]}. \tag{24}$$

Importantly, examining eqs. (22) and (24), it can be observed that the achievable MSE is greatly affected by the cross-covariance terms, $\mathrm{Av}[(x_i - \bar{x})(y_i - \bar{y})]^2$, and, $\mathrm{E}[\varepsilon_{xi}\varepsilon_{yi}]$, respectively. In practice, this cross-covariance can be expected to be much larger between the zero-mean processes representing the model inputs and outputs than between the measurement noises at the model inputs and outputs. Consequently, the LS estimation of the model parameters performs poorly when the input and output signals are noisy constants as assumed in (10). In such a case, some other strategy for identifying the model parameters has to be adopted.

A. Estimating the model inputs and outputs

In the absence of measurement noises, the $n = (p+1)$ measurements are sufficient to obtain the model parameters in (4) by inverting the matrix, $\boldsymbol{\Phi}$, i.e.,

$$\boldsymbol{a} = \boldsymbol{\Phi}^{-1}(Q(\boldsymbol{x}))Q(\boldsymbol{y}). \tag{25}$$

However, theoretical guarantees about the existence of the inverse, $\boldsymbol{\Phi}^{-1}$, are not considered further in this paper.

The noise in the measurements of the static model inputs and outputs can be suppressed statistically by taking repeated measurements. In particular, considering the input-output model (10), this leads to the problem of estimating the deterministic, but otherwise unknown constants in the zero-mean, stationary additive noises from multiple measurements.

Several strategies were proposed in the literature for estimating the deterministic (without any prior knowledge) parameters [12]. The minimum variance unbiased (MVUB), and among them, the best linear unbiased (BLUE) methods yield the estimators with the minimum variance, provided that they exist, and that they can be found. The LS estimator will perform poorly as argued in the previous subsection. The maximum-likelihood (ML) estimator is relatively easy to obtain for simple input-output signal models (10), and since it is asymptotically unbiased as well as consistent, this estimator is selected here. Furthermore, note that it is sufficient to only consider the estimators for one input-output signal, since all these input-output signals have the same model (10).

In particular, given n quantized measurements, x_i, $i = 1, 2, \ldots, n$, the task is to derive an ML estimator of the constant, $\bar{x}$, in an additive noise, ε_{xi}. In this paper, we assume that the additive noise is zero-mean, Gaussian, and stationary with the variance, σ^2. If the measurements are unquantized, it is straightforward to show that the ML estimator is the arithmetic mean, i.e., [12]

$$\hat{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} (\bar{x} + \varepsilon_{xi}) = \bar{x} + \bar{\varepsilon}_{xi}. \tag{26}$$

In the case the measurements are quantized into integer values using the mapping (7), the probability of the measurement, $\check{x}_i = k$, where $k \in \mathbb{Z}$ can be computed as,

$$\Pr(\check{x}_i = k) = Q\left(\frac{\Delta(k - 1/2) - \bar{x}}{\sigma}\right) - Q\left(\frac{\Delta(k + 1/2) - \bar{x}}{\sigma}\right) \tag{27}$$

where the Q-function for the standard Gaussian variable is defined as,

$$Q(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}\, \mathrm{d}t. \tag{28}$$

Provided that the additive noise is also white, the measurements are independent, and the ML estimator maximizes the joint probability density,

$$\Pr(\{\check{x}_i\}_i) = \prod_{i=1}^{n} Q\left(\frac{\Delta(\check{x}_i - 1/2) - \bar{x}}{\sigma}\right) - Q\left(\frac{\Delta(\check{x}_i + 1/2) - \bar{x}}{\sigma}\right). \tag{29}$$

Taking the logarithm, and then the derivative by $\bar{x}$ (i.e., the parameter to be estimated), we obtain,

$$\frac{\partial}{\partial \bar{x}} \log \Pr(\{\check{x}_i\}_i) = -\frac{1}{\sigma}\sum_{i=1}^{n} \frac{\dot{Q}\left(\frac{\Delta(\check{x}_i - 1/2) - \bar{x}}{\sigma}\right) - \dot{Q}\left(\frac{\Delta(\check{x}_i + 1/2) - \bar{x}}{\sigma}\right)}{Q\left(\frac{\Delta(\check{x}_i - 1/2) - \bar{x}}{\sigma}\right) - Q\left(\frac{\Delta(\check{x}_i + 1/2) - \bar{x}}{\sigma}\right)}. \tag{30}$$

In order to find, for which value of $\bar{x}$, the expression (30) becomes zero to maximize the log-likelihood, we can linearize the Q-function and its derivative about the point, x_0, i.e.,

$$\begin{aligned} Q(x) &\approx Q(x_0) - \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x_0^2/2}(x - x_0) \\ \dot{Q}(x) &\approx \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x_0^2/2}(x_0 x - x_0^2 - 1). \end{aligned} \tag{31}$$

The corresponding approximations are then,

$$\begin{aligned} Q(x_0 - b) - Q(x_0 + b) &\approx b\, \mathrm{e}^{-x_0^2/2}\sqrt{\frac{2}{\pi}} \\ \dot{Q}(x_0 - b) - \dot{Q}(x_0 + b) &\approx -x_0\, b\, \mathrm{e}^{-x_0^2/2}\sqrt{\frac{2}{\pi}}. \end{aligned} \tag{32}$$

Assuming $x_0 = (\Delta\check{x}_i - \bar{x})/\sigma$, and, $b = (\Delta/2)/\sigma$, in approximations (32), the derivative of the log-likelihood function (30) can be greatly simplified as,

$$\sum_{i=1}^{n} \frac{\Delta\check{x}_i - \bar{x}}{\sigma^2} \stackrel{!}{=} 0. \tag{33}$$

Consequently, we find that the ML estimator of $\bar{x}$, from the quantized noisy measurements, $\check{x}_i$, is again a simple arithmetic average, i.e.,

$$\hat{\bar{x}} = \Delta \frac{1}{n} \sum_{i=1}^{n} \check{x}_i. \tag{34}$$

However, and importantly, note that the ML estimator was derived under the assumption that, $b = (\Delta/2)/\sigma$, is relatively small (i.e., $b < 1$), so that the linearization is sufficiently accurate. The value, $\Delta/2$, also represents the maximum quantization error, and thus, we can conclude that the arithmetic average estimator can be expected to perform comparatively well as the arithmetic average estimator for the unquantized measurements, when $(\Delta/2) \ll \sigma$.

The similar derivation can be performed for the case of binary quantization (9) when the measurements are quantized to, -1, and, $+1$, values. Under the assumption that, $\bar{x} \ll \sigma$, the ML estimator (which, in this case, can be shown to be actually the MVUB estimator) becomes,

$$\hat{\bar{x}} = \sigma \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^{n} \check{x}_i, \quad \check{x}_i \in \{-1, +1\}. \tag{35}$$

Thus, the estimator for the binary quantized measurements requires knowledge of the noise standard deviation, σ.

IV. Estimator Variances

In this section, the goal is to compare the variances of the estimation errors for different estimators considered in the previous section. In particular, when the measurements are unquantized, the estimator (26) is unbiased, and its variance is simply,

$$\mathrm{E}\left[(\hat{\bar{x}} - \bar{x})^2\right] = \sigma^2/n. \tag{36}$$

When the measurements are uniformly quantized, the ML estimator (34) may be biased, i.e.,

$$\begin{aligned} \mathrm{E}\left[\hat{\bar{x}}\right] &= \frac{\Delta}{n} \sum_{i=1}^{n} \mathrm{E}[\check{x}_i] = \frac{\Delta}{n} \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} k \Pr(\check{x} = k) \\ &= \frac{\Delta^2}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} k \, \mathrm{e}^{-\frac{(\Delta k - \bar{x})^2}{2\sigma^2}} \end{aligned} \tag{37}$$

where we assumed linearization (32) of the Q-function.

Further insight can be obtained by analyzing the best case, and the worst case quantization scenarios. In particular, without loss of generality, the best case scenario occurs, when $\bar{x} = 0$ (more precisely, if $\bar{x}$ is an integer multiple of Δ); then, the mean, $\mathrm{E}\left[\hat{\bar{x}}\right] = 0$, and the ML estimator (34) is unbiased. On the other hand, the largest bias occurs for the values, $\bar{x} = \pm\Delta/2$ (more precisely, if $\bar{x}$ is an odd-integer multiple of $\Delta/2$). Hence, let, $\bar{x} = -c\Delta/2$, where $c = 0$, represents the best case, and $c = 1$, represents the worst case scenario, respectively.

The ML estimator (34) with quantized measurements has the variance,

$$\begin{aligned} \mathrm{E}\left[\left(\hat{\bar{x}} - \mathrm{E}\left[\hat{\bar{x}}\right]\right)^2\right] &= \frac{\Delta^2}{n^2} \mathrm{E}\left[\sum_{i,j=1}^{n} \check{x}_i \check{x}_j\right] - \Delta^2 \mathrm{E}[\check{x}_i]^2 \\ &= \frac{\Delta^2}{n} \left(\mathrm{E}\left[\check{x}_i^2\right] - \mathrm{E}[\check{x}_i]^2\right). \end{aligned} \tag{38}$$

To simplify the notation, define the moment [cf. (37)],

$$\begin{aligned} Z_m(\Delta/\sigma) &= \mathrm{E}[\check{x}_i^m \mid \bar{x} = -c\Delta/2] \\ &= \sum_{k=-\infty}^{\infty} k^m \Pr(\check{x} = k \mid \bar{x} = -c\Delta/2) \\ &= \frac{\Delta}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} k \, \mathrm{e}^{-\frac{(k+c/2)^2}{2} \frac{\Delta^2}{\sigma^2}}. \end{aligned} \tag{39}$$

After substituting $Z_m(\Delta/\sigma)$ into (38), the final expression for the estimator variance becomes,

$$\mathrm{E}\left[\left(\hat{\bar{x}} - \mathrm{E}\left[\hat{\bar{x}}\right]\right)^2\right] = \frac{\Delta^2}{n} \left(Z_2(\Delta/\sigma) - Z_1^2(\Delta/\sigma)\right). \tag{40}$$

The derived MSE expression (40) is compared with the computer simulations in Figure 3 assuming $n = 100$ measurements, and the quantization intervals with $\Delta = 1/2$. It can be observed that the derived expression is in a good agreement with simulations, provided that the condition, $\Delta \ll \sigma$, is satisfied. For larger values of Δ/σ, the derived expression represents a loose lower bound of the actual MSE. As expected, when the estimator with quantized inputs is unbiased (the best case scenario), the MSE continues to be reduced by reducing the amount of measurement noise. When the quantization error makes the estimator to be biased, the MSE eventually saturates, as might be expected.

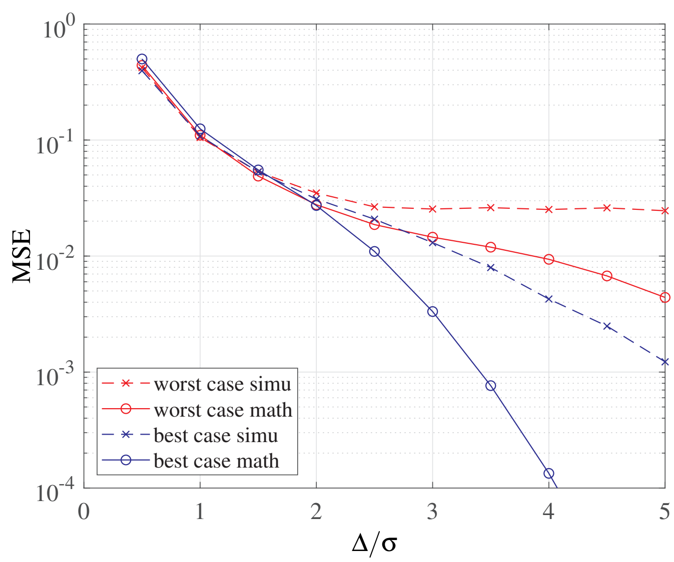

Figure 3. The MSE of the ML estimator with uniformly quantized measurements corresponding to the best case and the worst case quantization errors, respectively.

The variance of the MVUB estimator (35) with the binary quantized measurements can be shown to be,

$$\mathrm{E}\left[\left(\hat{\bar{x}} - \mathrm{E}\left[\hat{\bar{x}}\right]\right)^2\right] = \frac{\pi}{2} \frac{\sigma^2}{n}. \tag{41}$$

Thus, it is $(\pi/2)$ times larger than the variance (36) of the estimator from unquantized measurements, and importantly, provided that the condition, $\bar{x} \ll \sigma$ is satisfied.

The Cramér-Rao bound can be derived using again a linearization of the Q-function in the low signal-to-noise ratio (SNR) regime to obtain, [13]

$$\mathrm{E}\left[\left(\hat{\hat{x}}-\mathrm{E}\left[\hat{\hat{x}}\right]\right)^2\right] \geq J^{-1} = \frac{\sigma^2}{n}\frac{(1-Q(\bar{x}/\sigma))\,Q(\bar{x}/\sigma)}{\dot{Q}(\bar{x}/\sigma)} \qquad (42)$$

where J denotes the Fisher information matrix (a scalar value, here). The normalized Cramér-Rao bound, nJ^{-1}/σ^2, is shown in Figure 4 (black-line), together with the MSE of the estimator having the binary quantized measurements (41) (blue-line), and the MSE of the estimator with unquantized measurements (36) (red-line). It can be observed that the MSE raises quickly with improving SNR. In such a case, the binary quantization error starts dominating, and it cannot be reduced, for example, by simply increasing the number of measurements.

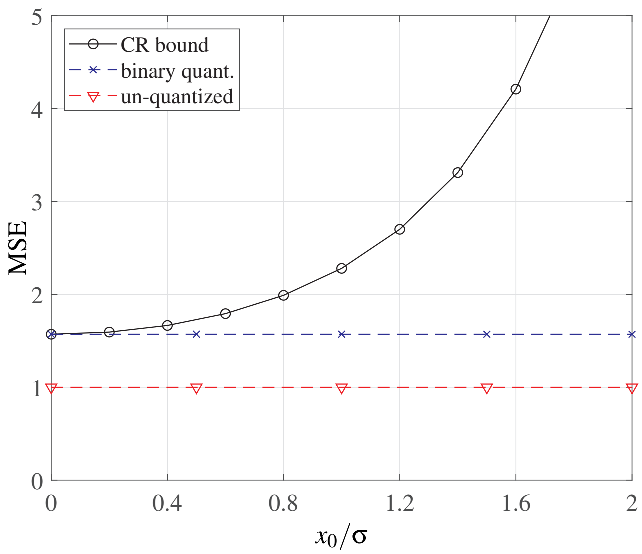

Figure 4. The Cramér-Rao bound of the estimator with the binary quantized measurements (black line), the actual MSE in the low-SNR regime (blue line), and the MSE of the estimator with the unquantized measurements (red-line).

V. Discussion and Future Work

Our investigations showed that the quantization noise can be neglected, provided that it is comparable with the measurement noise. If this condition is not satisfied, the estimators are only unbiased and consistent with respect to the additive measurement noise, and the estimation error is dominated by the residual quantization noise. The measurements obtained at both the system inputs and outputs represent a classical problem of system identification. When the inputs and outputs are static, i.e., they are constant values observed in an additive noise, the recommended strategy for estimating the model parameters is to first clean the input-output measurements by suppressing the measurement noises. This can be done independently for each input and output using different types of estimators. The noise-free input and output values can be then substituted into the model, and the model parameters are obtained by solving the same number of linear or non-linear equations representing the system model. This strategy is superior to classical least-squares model fitting (i.e., without suppressing the measurement noises first), provided that the inputs and outputs are noisy constant values. Furthermore, estimating the model parameters from input-output data pairs resembles a supervised machine learning. The main difference is that the data examples for machine learning are usually assumed to be noise-free, and the number of parameters assumed in machine learning models can be excessively large.

In this paper, our focus was on identifying relatively small linear models from their input-output measurements. Such models are common not only in engineering, but they also represent many physical laws. For example, Schrödinger and Maxwell's equations are both linear. It was shown in Figure 2 that the coarse-grained quantization can substantially affect the model, and also our perception of reality, if the model represents a physical law. This phenomenon is referred to here as Minecraft of system modeling, since the quantization makes the reality to appear as if it consisted of discrete blocks.

The future work can investigate the optimum representations of MIMO systems with discretized inputs and outputs. The non-linear systems can be modeled by recursive structures [14]. The fundamental question is how to suppress the quantization noise akin to suppressing the measurement noise. In this paper, the static input and output values were considered. Measuring the systems having the random processes as their inputs and outputs is more challenging, as it requires precise time-synchronization of the measurements at all the system inputs and outputs.

References

[1] P. Loskot, "Key ideas in parameter estimation," in *Proc. SIGNAL*, 2004, pp. 19–27.

[2] S. Haykin, *Adaptive Filter Theory*, 5th ed. Pearson Education, Harlow, England, 2014.

[3] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. Springer-Verlag New York, USA, 2000.

[4] J. C. Keenan and J. B. Lewis, "Estimation with quantized measurements," in *IEEE ICDC*, 1976, pp. 1284–1291.

[5] M. Fu and C. E. de Souza, "State estimation for linear discrete-time systems using quantized measurements," *Automatica*, vol. 45, no. 12, pp. 2937–2945, Dec. 2009.

[6] H. Poor, "Fine quantization in signal detection and estimation," *IEEE Trans. Info. Theory*, vol. 34, no. 5, pp. 960–972, Sep. 1988.

[7] S. Vardeman and C.-S. Lee, "Likelihood-based statistical estimation from quantized data," *IEEE Trans. Instrum. Measurements*, vol. 54, no. 1, pp. 409–414, Feb. 2005.

[8] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, "Sample, quantize, and encode: Timely estimation over noisy channels," *IEEE Trans. Communications*, vol. 69, no. 10, pp. 6485–6499, Oct. 2021.

[9] P. Stoica, X. Shang, and Y. Cheng, "The cramér-rao bound for signal parameter estimation from quantized data," *IEEE Sig. Proces. Magazine*, vol. 39, no. 1, pp. 118–125, Dec. 2021.

[10] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer-Verlag, New York, USA, 1994.

[11] P. Loskot, "Polynomial representations of high-dimensional observations of random processes," *Mathematics*, vol. 9, no. 123, pp. 1–24, Jan. 2021.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Upper Saddle River, NJ, USA, 1993, vol. I.

[13] Z. Hrdina, *Statistical Radio-Engineering*. Publishing Company of the Czech Technical University of Prague, 1996, in Czech.

[14] S. S.-T. Yau, X. Chen, X. Jiao, J. Kang, Z. Sun, and Y. Tao, *Principles of Nonlinear Filtering Theory*. Springer Nature Switzerland AG, 2024.