Journal of Process Control 40 (2016) 13–23

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Quantisation and data quality: Implications for system identification☆

Yuri A.W. Shardt[a,*], Xu Yang[b], Steven X. Ding[a]

[a] Institute of Control and Complex Systems, University of Duisburg-Essen, Duisburg, North Rhine-Westphalia, Germany
[b] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form 11 January 2016
Accepted 18 January 2016
Available online 12 February 2016

Keywords:
Data quantisation
Data quality assessment
System identification
Historical data

ABSTRACT

In the pursuit of online, data-driven process control, there is a need to determine the quality of the data being processed before actually using it. One area that needs to be considered is data quantisation. Although in many applications it has been assumed that the impact of quantisation is to solely increase the variance of the signal, in certain cases this may not hold. This is especially the case when dealing with signals from poorly quantised sources, such as temperature sensors. In this case, the effect of quantisation cannot be solely considered by the impact of the increase in the variance. Therefore, this paper will examine the effects of small scale quantisation with the view of determining an appropriate metric for measuring the effect on data quality of quantisation. It will be shown that if the ratio of the unquantised signal variance and the distance between quantisation step sizes are below a given threshold, then the identification of the process parameters will be problematic. Detailed numerical simulations as well as an example drawn from a real system are presented to validate the proposed metrics and approach.

1. Introduction

To achieve complete data-driven process identification and control, there is a need to automatically assess the data quality of the incoming signals and take appropriate action before the signals inappropriately affect the computation of the process parameters and controller. In order to achieve this, a data quality assessment framework needs to be developed. Currently, there exist two similar approaches to developing an overall data quality framework [1,2]. Both these approaches focus on developing methods that can determine whether a given data set is sufficiently excited for process systems identification. However, it is possible that, although the signal is sufficiently excited for identification, due to various preprocessing that has been performed, the value obtained may not be accurate. One rather common method of preprocessing is quantisation or the conversion of continuous signal values to a limited set of possible signal values. In most cases, the effects of quantisation on the signal properties are minimal. In some cases, due to improper initial calibration or subsequent implementation, the quantisation may be inaccurate. Such behaviour is often seen in temperature signals, where the thermocouple used is calibrated for a wide range of temperatures leading to a rather coarse resolution. If the temperature measurements do not vary greatly, then it is possible that there will be a loss of precision due to the large quantisation that has been adopted. This can occur often in a closed-loop process operating under tight control, that is, with minimal excitation. Such data are often the foundation for subsequent use in process monitoring, fault detection and isolation, and adaptive controller design. Therefore, understanding the impact of quantisation will have an impact on these fields.

The analysis of the impact of quantisation on the signal properties and hence on process identification has been approached from many different perspectives. In the most common method, it has been assumed that the quantisation levels have a fine enough resolution so that the overall impact is that the quantised error is uncorrelated with the original signal. Hence, the only effect is an increase in the signal variance. In such cases, it is possible to use Sheppard's correction, that is, assume that quantisation increases the variance [3]. Work analysing the implications of this approach has been widely published [4–7] with the general conclusion that the overall impact on identification is minimal. However, although this approach works for many signals, it can fail when an improper quantisation has been performed. In such cases, there is a need to consider the full theory and examine its implications on the signal properties. The fundamentals of this rigorous approach can be found in [3,8], where a detailed model and formulation of quantisation and its effects on the signal can be seen. Finally, given the large use of system identification in many applications, the impact of quantisation will be important in such areas as fault detection,

☆ Part of this work was presented at the 17th IFAC Symposium on System Identification, SYSID 2015 in Peking, China.

* Corresponding author. Tel.: +49 203 379 3380.
E-mail addresses: yuri.shardt@uni-due.de (Y.A.W. Shardt), yangxu@ustb.edu.cn (X. Yang), steven.ding@uni-due.de (S.X. Ding).

http://dx.doi.org/10.1016/j.jprocont.2016.01.007
0959-1524/

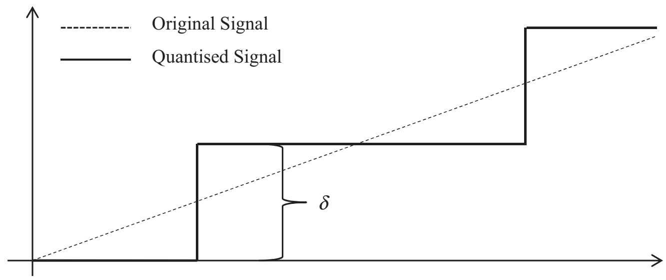

Fig. 1. Graphical representation of a quantiser.

data-driven controller design, and control performance monitoring.

Therefore, the objectives of this paper are (1) to examine in detail the effect of quantisation on system identification through a detailed review of quantisation theory; (2) to propose an appropriate metric that can be used to determine the suitability of the data given the quantisation present for use in system identification and process monitoring; and (3) to validate the results using both Monte Carlo simulations and experimental data from system identification.

2. The theory of quantisation

Consider a quantiser that takes a continuous, analogue signal and returns a quantised, analogue signal that only allows certain values for the signal. The behaviour of a quantiser is shown in Fig. 1. Mathematically, this can be modelled as [6]

$$\tilde{x}_t = \delta\left\lfloor\frac{(x_t+\lambda)}{\delta}\right\rfloor \tag{1}$$

where $x^{\sim}{}_t$ is the quantised signal, δ is the spacing between quantisation levels, λ is the centre point of the quantiser, x_t is the original unquantised signal, and $\lfloor\cdot\rfloor$ is the floor or round down function. The midpoint riser quantiser, where $\lambda = 0.5$, is the most common type of quantiser [6]. It can be noted that quantisation is a nonlinear process, which can lead to complications in the analysis of results.

For the purposes of this discussion, it will be assumed that the number of levels is infinite, that is, saturation does not occur, and that the true process is a Gaussian signal centred about the midpoint of one of the quantisation levels. Consider the case where the input to the quantiser can be modelled as a Gaussian distribution with a probability density function, f_x,

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}} \tag{2}$$

where σ is the standard deviation of the Gaussian noise, which is assumed to have a mean of zero. Let the error between the true value and the quantised value be given as

$$\varepsilon = x - \tilde{x} \tag{3}$$

It is desired to determine the properties of the error signal and hence its implications for identification. Thus, Theorem 1 determines the quantisation error density function, while Theorems 2 and 3 provide the mean and variance of the quantisation error. Finally, using the properties of the quantisation error, Theorem 4 determines the variance of the quantised signal, while Theorem 5 provides the correlation between the quantised and original signals. The uncentred case is briefly examined in Theorem 6 and it is shown that the overall results are no different from the centred case. Together Theorems 4–6 allow for the impact of quantisation on system identification to be clarified.

Theorem 1. *Quantisation Error Density Function: The probability density function of the quantisation error can be written as*

$$f_\varepsilon(x) = \begin{cases} \dfrac{1}{\delta}+\dfrac{2}{\delta}\displaystyle\sum_{n=1}^{\infty}\cos\left(\frac{2\pi n x}{\delta}\right)\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right) & -\dfrac{\delta}{2}\le x<\dfrac{\delta}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

Proof. A proof can be found in [3].□

Theorem 2. *Mean Value of the Error: The mean value of the error is zero.*

Proof. From the definition of the error, the mean value can be computed as

$$\begin{aligned} E(\varepsilon) &= \int xf(x)dx \\ &= \frac{1}{\delta}\int_{-\delta/2}^{\delta/2} x\left(1+2\sum_{n=1}^{\infty}\cos\left(\frac{2\pi n x}{\delta}\right)\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right)\right)dx = 0 \end{aligned} \tag{5}$$

□

Theorem 3. *Variance of the Error: The variance of the error can be written as*

$$E(\varepsilon^2) = \operatorname{var}(\varepsilon) = \frac{\delta^2}{12}\left[1+\frac{12}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right)\right] \tag{6}$$

Proof. Following the same approach as in Theorem 2, it can be shown that

$$\begin{aligned} E(\varepsilon^2) &= \int x^2 f(x)dx = \frac{1}{\delta}\int_{-\delta/2}^{\delta/2} x^2\left(1+2\sum_{n=1}^{\infty}\cos\left(\frac{2\pi n x}{\delta}\right)\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right)\right)dx \\ &= \frac{\delta^2}{12}\left[1+\frac{12}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right)\right] \end{aligned} \tag{7}$$

□

Theorem 4. *Variance of the Quantised Signal: The variance of the quantised signal is given as*

$$\begin{aligned} E(\tilde{x}^2) &= \operatorname{var}(\tilde{x}) = E((x-\varepsilon)^2) = E(x^2)+4\sigma^2\sum_{n=1}^{\infty}(-1)^n\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right) \\ &+\frac{\delta^2}{12}\left[1+\frac{12}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right)\right] \end{aligned} \tag{8}$$

Proof. A detailed proof, which involves the creation of a joint probability distribution for two Gaussian signals and their quantised analogues, is available in [3].□

Theorem 5. *Covariance of the Original Signal and its Quantisation Error: The covariance between the original Gaussian signal and the quantised signal is given as*

$$\begin{aligned} E(x\varepsilon) &= -2\sigma^2\sum_{n=1}^{\infty}(-1)^n\exp\left(-\frac{2\pi^2n^2\sigma^2}{\delta^2}\right) \\ &= -\sigma^2\left(\Theta_4\left(0,\exp\left(-\frac{2\pi^2\sigma^2}{\delta^2}\right)\right)-1\right) \end{aligned} \tag{9}$$

where Θ_4 is the fourth Jacobi theta function.