Time Series Reconstruction from Quantized Measurements

M. Wang
Centre for Process Systems Engineering
Imperial College of Science Technology and Medicine
Prince Consort Road, London SW7 2BY

S. Saleem and N. F. Thornhill*
Department of Electronic and Electrical Engineering
UCL (University College London)
Torrington Place, London WC1E 7JE

Abstract

This paper describes a Quantization Regression (QR) algorithm which generates a nonlinear estimate of an autoregressive time series from quantized measurements. Its purpose is to retrieve the underlying information from quantized signals such as those from the analogue to digital converter of a plant instrument. The reconstructed signals have uses in data-centric applications such as controller performance assessment and system identification. The algorithm based upon Ziskand and Hertz is a combination of the 'Gaussian Fit' scheme of Curry with expectation-maximization (EM) algorithm of Dempster et al. The performance of Quantization Regression algorithm is compared with two other methods in fitting of an autoregressive time series for the reconstruction of a quantized signal.

Keywords

EM algorithm, Linear regression, Quantization, Performance analysis, Time series analysis

Introduction

A great deal of engineering data occurs in the form of time series where observations are dependent and where the nature of this dependence is of interest. For example, Desborough and Harris (1992) used routine closed loop process data to estimate the normalized control loop performance index. In that case, an AR model was implemented. Thornhill et al. (1999) observed that quantization of the measurements influenced the normalized performance index and thus motivated the need for an algorithm to accurately recover a signal from a quantized time series. Quantization is often observed in the outputs of process instruments even though 10 bit A/D conversion provides 1024 quantization levels because a measurement controlled to a steady value is likely to sample just a few of the available quantizer levels.

Ziskand and Hertz (1993) proposed an algorithm to estimate coefficients of a quantized autoregressive (qAR) process. Their implementation and examples illustrated the case of one-bit (two level) quantization. The algorithm was based on the following two developments: Dempster et al. (1977) presented a method for computing maximum likelihood estimates from "incomplete" data, i.e. data having a many-to-one mapping in the measurement function. Quantized process data are incomplete because many values of the underlying signal map to each quantizer level. The algorithm comprised two steps: the first expectation, the second maximization, and was called the EM algorithm. Shumway and Stoffer (1982) proposed an approach to smoothing and forecasting for time series with incomplete observations. The EM algorithm was used in conjunction with Kalman smoothed estimators to derive a recursive procedure for estimating the qAR parameters by maximum likelihood. The algorithm is a general technique for finding maximum likelihood estimates from incomplete data (Little and Rubin, 1986).

Expectation calculations for the maximum likelihood step are provided by the Gaussian Fit algorithm. The method was proposed by Curry (1970) for a discrete-time nonlinear filter that recursively fits a Gaussian distribution to the first two moments of the conditional distribution of a system state vector. The Gaussian Fit algorithm is easy to compute and can handle non-stationary data and its operation is independent of the quantization scheme used. However, it requires more computation than a linear filter and can be applied only to Gauss-Markov processes. It is applicable to quantized process data because such data can be expressed as a Gauss-Markov process in state-space form.

The algorithm presented by Ziskand and Hertz (1993) estimated the coefficients of several superimposed qAR signals of known model order using a two-level quantizer. The signals represented sinusoids at different frequencies. The contribution of this paper is the extension of quantized regression to multiple quantizer levels and to qAR signals of order 2 and higher. The Akaike Information Criterion was used to determine the order of the AR model. Further, the QR algorithm was compared with two other methods to illustrate why it can recover the underlying signal better. Simulations and experimental data support the studies. The conclusion is that the QR algorithm can optimally estimate the model parameters and recover the underlying signal at the same time.

Methods

Problem Description

Quantized autoregressive (AR) time series are the subject of the work. The process may be written as follows:

$$S(n) = \phi \cdot S(n-1) + w(n)$$
$$x(n) = h \cdot S(n) + v(n)$$
$$z(n) = g(x(n))$$

*Author to whom all correspondence should be addressed. email: n.thornhill@ee.ucl.ac.uk, Tel: +44 20 7679 3983, Fax: +44 20 7388 9325

where $S(n) = (s(n-m+1), \ldots, s(n))^T$ is a state vector, $\{x(n)\}_{n=1}^N$ the N samples from an autoregressive process of order m and $\{z(n)\}_{n=1}^N$ the quantized measurements of the process. $w(n)$ and $v(n)$ are white noise samples with mean zero and variances σ_w^2 and σ_0^2 respectively and independent from each other. The state transition matrix, ϕ, has the autoregressive coefficients $a_1, a_2, \ldots, a_m$ in the last row:

$$\phi = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ a_m & \cdots & \cdots & a_2 & a_1 \end{pmatrix}$$

The observation vector $h = (0, 0, \ldots, 0, 1)$ is of length m, and g is the non-linear quantizer function whose input is the AR signal plus noise. The algorithm is suitable for both uniform and non-uniform quantizers, although in this paper uniform quantization intervals were used. The quantization interval is the distance between quantizer levels. The problem is to determine the model order, m, to estimate the AR coefficients $a_1, a_2, \ldots, a_m$ and to recover the underlying signal $\{s(n)\}_{n=1}^N$.

Implementation of QR Algorithm

The Quantization Regression (QR) algorithm is iterative and includes two main conceptual steps:

Step 1: A modified Kalman smoothing algorithm finds smoothed estimates and their covariances. At step n the key calculations are:

$$S_n^n = S_n^{n-1} + K_n \left((E(x(n))|z_n) - hS_n^{n-1} \right)$$
$$P_n^n = P_n^{n-1} - K_n h P_n^{n-1} + K_n \mathrm{cov}\left(x(n)|z_n\right) K_n^T$$

where S_n^{n-1} and P_n^{n-1} are one step ahead predictions of the state vector and its variance, z_n the sequence of quantized measurements and K_n is the Kalman gain. A conventional Kalman filter would use the quantized measurement z_n directly, but the modified algorithm calculates the expected value of x_n given that the quantized measurement falls within the observed quantization interval. The expectation is computed using the Gaussian Fit approximation (Curry, 1970). The last term in the calculation for P_n^n captures the inflation in the variance of the estimate caused by quantization.

Step 2: The likelihood function is maximised by iterative adjustment of the AR coefficients $a_1, a_2, \ldots, a_m$, the estimated signal and σ_w^2 and σ_0^2.

Typing errors found in Ziskand and Hertz were corrected as follows where the underline indicates the altered terms,

$$K_n = P_n^{n-1} h^T \underline{(\sigma_x^2)^{-1}}$$
$$P_{n-1}^N = \underline{P_{n-1}^{n-1}} + J_{n-1}\left(P_n^N - P_n^{n-1}\right)\underline{J_{n-1}^T}$$
$$P_{n-1,n-2}^N = P_{n-1}^{n-1} J_{n-2}^T + \underline{J_{n-1}}\left(P_{n,n-1}^N - \phi(r) P_{n-1}^{n-1}\right)\underline{J_{n-2}^T}$$

and the loop in the lag-one covariance calculation was for $n = N, N-1, \ldots, \underline{2}$.

The following improvements were made: (1) σ_0^2 was updated during the iteration rather than being taken as known; (2) the innovative form of the log liklihood function (Shumway and Stoffer, 1982) was used as a stopping criterion.

$$\Delta(\log L) = -\frac{1}{2}\sum_{n=1}^{N} \log(\sigma_x^2) - \frac{1}{2\sigma_x^2}\sum_{n=1}^{N}\left(z(n) - hS_n^{n-1}\right)^T\left(z(n) - hS_n^{n-1}\right)$$

Iterations stopped when the innovation became small; (3) the initial variance in the iteration was fixed and the state vector was initialised thus: $S_0^0(r+1) = S_0^N(r)$.

Determination of Model Order and Quality of Fit

Ziskand and Hertz assumed the model order m was known. Where the model order is unknown one approach to determining m is to fit AR process of progressively high order, to calculate the sum of squared errors (SSE) for each value of m and to plot SSE against m (Chatfield, 1989). One chooses the value of m where the addition of extra parameters gives little improvement in fit. The method does not work here since the aim is to recover the underlying signal not the quantized observation. An approach suited to maximum likelihood estimation is to use Akaike Information criterion (AIC) to determine model order. The required complete-data log likelihood function ($\log L$) is presented in Ziskand and Hertz (1993), and Shumway and Stoffer (1982). AIC can be calculated from $\log L$ and the model order m as $AIC = -2\log L + 2m$. The model order is the value of m where the largest decrease occurs.

The multiple coefficient of determination R^2 (Scheaffer and McClave, 1995) was used to determine the quality of the reconstruction:

$$R^2 = 1 - \frac{\sum_{n=1}^{N}\left(y(n) - \hat{y}(n)\right)^2}{\sum_{n=1}^{N}\left(y(n) - \bar{y}\right)^2}$$

where $\hat{y}(n)$ is the reconstruction of a true underlying signal $y(n)$ and $\bar{y}$ is its mean. $R^2 = 0$ implies the variance of the reconstruction error is as large as the variance of the measurements, and thus that there is a complete lack of fit of the model, while $R^2 = 1$ implies a perfect reconstruction.

Other Methods for Comparison

Two other methods for estimation of the AR coefficients were implemented for comparison.

One method was a one-step linear least squares estimate (LLS) of the AR coefficients. The quantized data $\{z(n)\}_{n=1}^{N}$ themselves were modelled as an AR series:

$$\begin{pmatrix} z(n) \\ z(n-1) \\ \vdots \\ z(m+1) \end{pmatrix} = \begin{pmatrix} z(n-1) & \cdots & z(n-m) \\ z(n-2) & \cdots & z(n-m-1) \\ \vdots & & \vdots \\ z(m) & \cdots & z(1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} + \begin{pmatrix} e(n) \\ e(n-1) \\ \vdots \\ e(m+1) \end{pmatrix} \quad (1)$$

The above matrix equation can be written as:

$$Y = X \cdot a + e$$

and least squares estimates of the AR parameters determined from:

$$\hat{a} = \left(X^T X\right)^{-1} X^T Y$$

The above calculation coincides with the linear regression approach in Desborough and Harris (1992) using a one-step prediction. It was also used in Thornhill et al. (1999) where the adverse influence of quantization was noted.

The other method, termed Kalman smoothing, used the same formulation as the QR algorithm but without the Gaussian Fit step. The Kalman smoother equations became:

$$S_n^n = S_n^{n-1} + K_n \left(z(n) - hS_n^{n-1}\right)$$
$$P_n^n = P_n^{n-1} - K_n h P_n^{n-1}$$

Results

Simulation Examples

A unit amplitude sinusoid wave was used as the quantizer input with quantization interval 0.5. Therefore the signal used just five quantizer levels and was coarsely quantized. The sine wave had a period of 51.2 samples, the value being selected so that 10 cycles used 512 samples (a power of 2). Using the method of backward differencing the sine wave $\sin(2\pi t/51.2)$ can be expressed as a two term AR series:

$$y(n) = 1.990y(n-1) - 0.998y(n-2)$$

Therefore we require the QR algorithm to give a model with $m = 2$ and coefficients $a_1 = 1.990$ and $a_2 = -0.998$. The AIC indicated that the optimum model order for the qAR time series was indeed $m = 2$. The QR algorithm was used to reconstruct an AR model with two terms and the results were compared with the reconstruction of $m = 2$ AR models using Kalman smoothing alone and the one-step linear least squares estimate. The recovering ability is compared in the left hand panel of Figure 1, where it can be seen visually that the QR algorithm provided the best reconstruction of the underlying sine wave. The results with the LLS method were the least satisfactory.

The estimated AR coefficients and R^2 values were:

QR:	$a_1 = 1.977$,	$a_2 = -0.993$,	$R^2 = 0.998$
Kalman:	$a_1 = 1.276$,	$a_2 = -0.304$,	$R^2 = 0.996$
LLS:	$a_1 = 0.982$,	$a_2 = -0.018$,	$R^2 = 0.968$

The QR algorithm recovered both coefficients with less than 1% error from the true values, while the LLS coefficients had large errors and could only achieve a model that said the next sample would be almost the same as the previous sample. Kalman smoothing without the Gaussian Fit step gave an intermediate result with the model coefficients in error by 36% and 70%. It is concluded that the major benefit of the QR algorithm with coarsely quantized data is the expectation step using the Gaussian Fit approximation.

A noisy sine wave with a signal to noise ratio of 1:1 was used as a second test signal into the quantizer input. Application of AIC indicated a model order of 5 or 6 for the QR algorithm and 6 for Kalman smoothing. The recovering ability of the three methods for $m = 6$ AR models is compared in the right hand panel of Figure 1. The QR algorithm and Kalman smoothing provided good reconstruction of the underlying sine wave and both were superior to the LLS reconstruction which provided less filtering of the noise. The R^2 values were:

QR:	$R^2 = 0.981$
Kalman:	$R^2 = 0.978$
LLS:	$R^2 = 0.764$

Influence of Quantization Interval

Figure 2 compares the R^2 measure across a range of quantization intervals. The left panel shows the results for the quantized sine wave signal using AR models with model order $m = 2$. The uppermost trend is for the QR algorithm, the lowest is the LLS algorithm and the Kalman smoothing algorithm is in-between. As was suggested by Figure 1, the Gaussian Fit element of the QR algorithm gave benefits over and above its Kalman smoothing component. The benefit increased as the quantization interval increased.

The right hand panel of Figure 2 shows the reconstruction performance of the quantized noisy sine wave using AR models with $m = 6$. The R^2 values for Kalman smoothing and QR were almost identical in this case, and both gave improvements over the LLS method. Thus when the unquantized signal included white noise, no matter what the quantization interval was, the recovery was contributed by Kalman smoothing. It is concluded that the benefit of the Gaussian Fit to the QR algorithm reduces as the influence of quantization reduces relative to the noise.

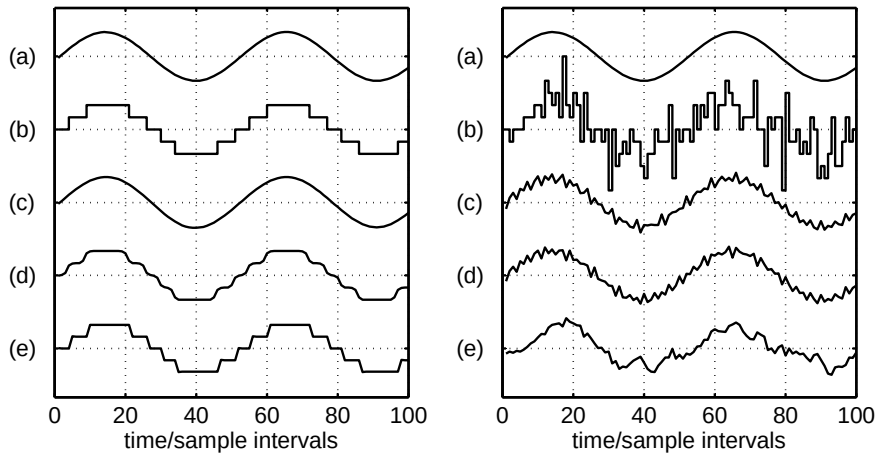

Figure 1: Reconstruction of a sine wave from quantized samples (left) and noisy quantized samples (right). (a): the sine wave signal; (b) quantized signal; (c) reconstructed with QR; (d) reconstructed with Kalman smoothing; (e) reconstructed with LLS.

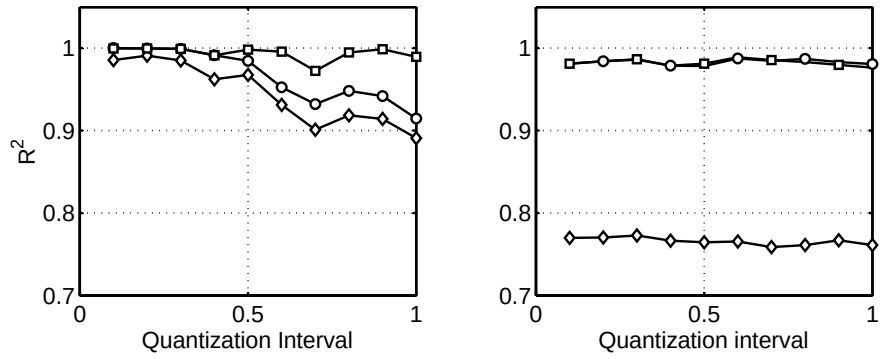

Figure 2: Reconstruction performance for a sine wave from quantized samples (left) and noisy quantized samples (right). QR: squares; Kalman smoothing: circles; LLS: diamonds.

Ensemble Length	a_1	a_2	R^2
N = 1440	1.977	-0.993	0.9981
N = 1280	1.979	-0.994	0.9981
N = 1024	1.977	-0.992	0.9980
N = 768	1.974	-0.989	0.9979
N = 512	1.968	-0.984	0.9977
N = 256	1.952	-0.968	0.9976

Table 1: The influence of ensemble length.

Influence of Ensemble Length

Table 1 shows the influence of data ensemble length, which was examined using the sine wave signal with quantization interval of 0.5.

The table shows that the values of the AR coefficients recovered by the QR algorithm diverged from the correct values as the data ensemble length became smaller. In this example a data ensemble length of 800 gave errors compared to the true values of about 1% in each of the AR coefficients, therefore it is recommended that the data ensemble length be at least 800 samples.

Performance with Plant Data

Figure 3 shows the behaviour of the QR, Kalman smoothing and LLS algorithms with pilot plant measurements. The measurements were from the transient response of the pH control in a buffered fed-batch yeast fermentation process. They were quantized by the A/D converter of the pH probe. The aim of the analysis was

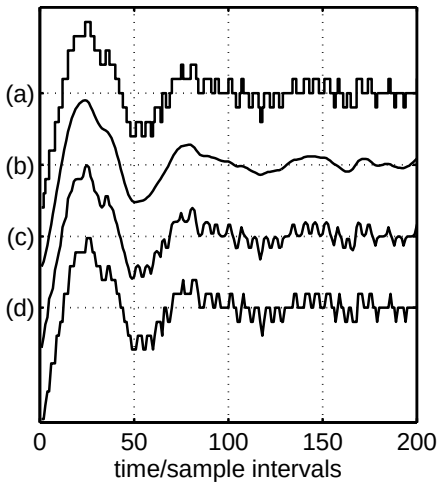

Figure 3: Reconstruction of quantized plant data. (a) quantized trend from plant; (b) reconstructed using QR; (c) using Kalman smoothing; (d) using LLS.

to recover the underlying smooth transient response of the fermenter pH. The closed loop under proportional-plus-integral control of the process has dominant under-damped complex conjugate poles and therefore the AR time series should have $m = 2$. The s-plane poles are $s = -0.025 \pm 0.108j$ corresponding to damped oscillations with a period of 58 samples and an exponential time constant of 40 samples. These s-plane poles map to z-plane poles at $z = 0.969 \pm 0.105j$.

Application of AIC to the QR algorithm indicated that the model order should indeed be $m = 2$. The reconstructions of $m = 2$ AR models using QR, Kalman smoothing and the LLS methods are shown in Figure 3. It was not possible to determine the R^2 values because the true underlying signal was not known. A quantitative comparison of the z-plane poles shows the QR algorithm has achieved the correct reconstruction because the AR model was:

$$y(n) = 1.914y(n-1) - 0.934y(n-2)$$

which has z-plane poles at $z = 0.957 \pm 0.090j$. These poles are near those of the closed loop process and it is concluded that the AR coefficients determined by the QR algorithm captured the under-damped oscillatory behavior of the pH response. The other models were:

$$\begin{array}{ll} \text{Kalman:} & y(n) = 1.210y(n-1) - 0.231y(n-2) \\ \text{LLS:} & y(n) = 0.972y(n-1) + 8 \cdot 10^{-3}y(n-2) \end{array}$$

both of whose z-plane poles are real. These recovered AR time series have no oscillatory behavior of their own and the reconstructions appear oscillatory only because they are driven by the experimental data.

Conclusion

For coarsely quantized signals, the QR (quantized regression) algorithm can recover the underlying signal from quantized observations better than the LLS (linear least squares) algorithm, as measured by the R^2 values. The fundamental reason behind the phenomenon is that QR algorithm based on the Gaussian Fit algorithm and Kalman smoothing can recover some nonlinear parts from the quantized observation, while the LLS algorithm assumes the quantizer has a fixed input-output relationship.

Kalman smoothing was also used without the Gaussian Fit scheme. Its performance was also superior to that of the LLS method. When the effects of quantization were dominant the performance of the QR algorithm was significantly better than that of Kalman smoothing alone. When the underlying signal was noisy, however, the performances of the QR and Kalman smoothing algorithms were similar. Therefore it is concluded that the Gaussian Fit scheme offers the most improvement when quantization is severe, but that when noise is predominant the majority of the benefit is due to Kalman smoothing.

QR reconstruction of a transient response signal from experimental plant data gave an AR series of the correct order and with z-plane poles close to the true values. It can be concluded that QR reconstruction has the capacity to be useful in the accurate reconstruction of autoregressive time series from quantized process data.

Acknowledgments

Meihong Wang gratefully acknowledges the financial support of the Centre for Process Systems Engineering, Imperial College of Science, Technology and Medicine.

References

Chatfield, C., *Time Series Analysis: an introduction*. Chapman and Hall Ltd., London (1989).

Curry, R. E., *Estimation and control with quantized measurements*. MIT Press, Cambridge, MA (1970).

Dempster, A. P., N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, **B-39**, 1–37 (1977).

Desborough, L. and T. Harris, "Performance Assessment Measures for Univariate Feedback Control," *Can. J. Chem. Eng.*, **70**, 1186–1197 (1992).

Little, R. and D. B. Rubin, *Statistical Analysis with Missing Data*. Wiley, New York, NY (1986).

Scheaffer, R. L. and J. T. McClave, *Probability and Statistics for Engineers*. Duxbury Press, Belmont (1995).

Shumway, R. H. and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Analysis*, **3**, 253–263 (1982).

Thornhill, N. F., M. Oettinger, and P. Fedenczuk, "Refinery-wide control loop performance assessment," *J. Proc. Cont.*, **9**, 109–124 (1999).

Ziskand, I. and D. Hertz, "Multiple frequencies and AR parameters estimation from one bit quantized signal via the EM algorithm," *IEEE Trans. Signal Proces.*, **41**, 3202–3206 (1993).