ON QUANTIZATION AND ITS IMPACT ON THE EXACT RECOVERY OF HIGH ORDER MOMENTS

L Cheded
Department of Systems Engineering
King Fahd University of Petroleum and Minerals
Dhahran 31261
Saudi Arabia
e-mail: FACG002@SAUPM00.BITNET

Abstract

This paper addresses the problem of the exact recovery of unquantized moments from their quantized counterparts. A brief review of amplitude quantization and its impact on the Exact Moment Recovery (EMR) problem is given. In particular, a special class of order p, called L_p, for which EMR is always achieved regardless of the quantization fineness used, is introduced together with some new results on its properties. Due to the tremendous practical gains that can accrue from the use of 1-bit quantized members of L_1, it is shown how to force any signal to become a member of this class, hence naturally re-discovering the dithered quantization process. Two approaches to the EMR problem and some simulation results which are in very good agreement with the theory, are presented.

1 Introduction

Analog-to-Digital (A/D) conversion of a signal consists of two processes: sampling and amplitude quantization. The first process is known to involve no loss of information as long as the band-limited signal is sampled according to Shannon's sampling theorem. However, because of the approximating nature of quantization, the second process involves a signal degradation whose severity depends on the quantization fineness used [1], [2]. For the clarity of the exposition, the impact of this quantization-induced signal degradation will be discussed in the context of the 1-D EMR problem only. A new general class of signals of order p and called L_p will be introduced and new results on some of its properties stated. Each member of L_1, exhibits the attractive property of linearizing, in the mean sense, the quantizer's Input//Output (I/O) characteristics. Of paramount importance is the fact that for members of L_p, the EMR problem is solved regardless of the quantization fineness used. It then follows that if the most practically-attractive choice of 1-bit quantization scheme is made, then several 1-bit DSP systems (correlators, Fourier Analyzers and Power-Spectral Analyzers) can be designed with attractive attributes such as structural simplicity, low cost, high input bandwidth and real-time processing capability.

2 Classical Quantization and Moments Recovery

We shall introduce here a general uniform quantizer Q characterized by a shift factor a $\varepsilon[-½, ½)$ and a uniform quantization step q. This quantizer, with input X and output X_Q, is defined by the following nonlinear operator $Q(.)$:

$$X \xrightarrow{Q} X_Q=Q(X)=(a+n+½)q \text{ if } (a+n)q \le X < (a+n+1)q; n\varepsilon Z \quad (1)$$

For a = 0 and a = -½, the well-known mid-stepper and mid-treader quantizers are obtained respectively. Note here that elements from the 2 sequences $\{t_n\}=\{(a+n)q\}$ and $\{y_n=(a+n+½)q\}$ are called the transition points and the representation levels of the quantizer Q, respectively. If the p-th order unquantized and quantized moments, denoted by μ_p and μ_{Qp} respectively, are defined by:

$$\mu_p = E[X^p] \quad (2)$$

$$\mu_{Qp} = E[X_Q^p] \quad (3)$$

then using the Characteristic Function approach, it was shown in [1] and [3] that the quantized-unquantized moments relationship is given by:

$$\mu_{Qp} = A_p + B_p \quad (4)$$

and A_p and B_p are here called the principal term and the bias term and defined respectively by:

$$A_p = \frac{1}{p+1}\sum_{r=0}^{p} C_r^{p+1}\left(\frac{q}{2}\right)^{p-r}\mu_r \quad (p\oplus r\oplus 1) \quad (5)$$

$$B_p = \sum_{n\neq 0} e^{-i2\pi an}\sum_{r=0}^{p-1}\left(\frac{q}{2}\right)^{p-r} i^{-r}\sum_{\lambda=0}^{(p-1)-r}\frac{p!\,i^{\lambda-1}}{r!(p-r-\lambda)!}\frac{p\oplus r\oplus\lambda}{(n\pi)^{\lambda+1}}W^{(r)}\left(\frac{2n\pi}{q}\right) \quad (6)$$

where $C_r^{p+1}=\binom{p+1}{r}$, $i=\sqrt{-1}$ and $\oplus$ stands for modulo-2 addition.

0-7803-2431-5/95 $4.00 © 1995 IEEE

An interesting property of the general uniform classical quantizer is now stated in the following new lemma whose proof can either be directly found in [4] or easily derived from [5].

Lemma: For a given p-th order quantized moment μ_{Qp}, a general uniform classical quantizer Q, of step q and shift factor a $\varepsilon[-½,½]$, is equivalent to a transformation $T_p(x)$ that:

1. Depends only on the first-order distribution function $P_t(x) = U(x)$ of the transition points $\{t_n\}$ and
2. Satisfies

$$\mu_{Qp} = E[T_p(x)] \quad \forall p \geq 1 \tag{7}$$

where

$$T_p(x) = \sum_n [(a+n+½)q]^p \{P_t[(a+n+1)q - x] - P_t[(a+n)q-x]\} \tag{8}$$

and U (x) is the familiar unit step function.

This lemma clearly brings out the key role played by the distribution function $P_t(x)$ of the transition points in controlling the shape of $T_p(x)$ which represents the classical quantizer's p-th order moment-sense Input/Output characteristics.

3 Exact Moments Recovery and L_p

As mentioned earlier, the solution of the EMR problem is achieved for members of L_p. We will therefore first introduce the class L_p and state two new theorems on its properties which are the key to the solution of the EMR problem and whose proofs can be found in [4] or [5].

3.1 Definition of L_p

An ergodic and stationary signal X is called a member of the p-th order class L_p if its Characteristic Function (CF), $W_X(u)$, verifies the following:

$$x^{(r)}\left(\frac{2n\pi}{q}\right) = 0 \quad \forall r \quad [0,p-1] \text{ and } n \neq 0 \tag{9}$$

It is very clear from this definition that all members of L_p share the inherent and attractive property of automatically cancelling the unwanted bias B_p given in (6).

3.2 Two key properties of L_p

Theorem 1

In a general uniform classical quantization, a signal is a member of L_p if and only if its corresponding transformation $T_p(x)$ is a polynomial of degree p

Theorem 2: If $T_p(x)$ is a polynomial of degree p, i.e:

$$T_p(x) = \sum_{\lambda=0}^{p} c_\lambda x^\lambda \tag{10}$$

then:

The coefficients c_λ are independent of the shift factor a and are given by:

$$c_\lambda = \frac{p!}{(p-\lambda+1)!\lambda!}\left(\frac{q}{2}\right)^{p-\lambda}[p \oplus \lambda \oplus 1] \tag{11}$$

Corollary to Theorem 2:

The coefficients c_λ have the following characteristics:

(a) When p and λ have the same parity, $c_\lambda > 0$ $\forall X \varepsilon L_p$ and $\forall \lambda \in [0,p]$

(b) When p and λ have different parities and X is such that $\mu_t = 0$ for all odd $t \leq p$, then $c_\lambda = 0$ $\forall \lambda \in [0,p]$

Combining equations (7), (10) and (11) yields the following Input/Output equations which are listed below for p = 1, 2, 3, and 4 only.

$$\mu_1 = \mu_{Q1} \tag{12}$$

$$\mu_2 = \mu_{Q2} + \left(-\frac{q^2}{12}\right) \tag{13}$$

$$\mu_3 = \mu_{Q3} + \left(-\frac{q^2}{4}\mu_1\right) \tag{14}$$

$$\mu_4 = \mu_{Q4} + \left(-\frac{q^2}{2}\mu_2 - \frac{q^4}{80}\right) \tag{15}$$

The bracketed terms on the RHS of the last 4 equations are the well-known Sheppard's corrections and, as shown above, are exact if and only if the signal in question is a member of L_p.

4. Two Approaches to Exact Moments Recovery

It is clear from Equations (12)-(15) that a p - $th(p \geq 1)$ unquantized moment can be exactly recovered from its quantized counterpart in a recursive manner and for any desired quantization fineness. However, this first approach can be parallelized using vector processing. To achieve this, re-write Equations (12) - (15) in a matrix form as follows:

$$\underline{\mu} = A \cdot \underline{\mu_Q} + \underline{b} \qquad (16)$$

where

$$\underline{\mu} = [\mu_1 \ \mu_2 \ \mu_3 \ \mu_4]^T$$

$$\underline{\mu_Q} = [\mu_{Q1} \mu_{Q2} \ \mu_{Q3} \ \mu_{Q4}]^T$$

$$\underline{b} = \left[0 \ -\frac{q^2}{12} \ 0 \ \frac{7q^4}{240}\right]^T$$

and

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{q^2}{4} & 0 & 1 & 0 \\ 0 & -\frac{q^2}{2} & 0 & 1 \end{bmatrix}$$

The computation of the vector $\underline{\mu_Q}$ can be rendered parallel by using, in this case, 4 separate channels, all fed with X_Q, with the first one consisting of an accumulator $E[.]$ only that would generate μ_{Q1} and each of the remaining 3 comprising a multiplier (to generate the corresponding signal to be accumulated, i.e. X_Q^2 or X_Q^3 or X_Q^4) and an accumulator (to produce the corresponding μ_{Qi} for $i = 2$ or 3 or 4). In the attractive case of a 1-bit quantization with bipolar output, the multipliers will be binary ones and hence implementable with EXNOR gates and the accumulators will simply become 1-bit up/down counters. Also, given q, both A and $\underline{b}$ can be pre-computed and stored in memory to be finally used, in either a purely software or hardware scheme, in the recovery of the desired vector $\underline{\mu}$. Note that this approach assumes that $X \varepsilon L_1$. However, if it is not, it can always be made so as shown next.

The second approach is parallel in nature, recovers μ_P directly from μ_Q , $\forall\, p \geq 1$ and relies on a simple technique of forcing any signal, not already a member of L_1, to become so. This technique was natually derived from our theoretical studies (see for example [1] or [5]) and consists of adding a reference signal R that is itself a type-1 member of L_1, to the signal X prior to quantization.

This is, in fact, alternatively known as dithered quantization and was brought to our attention, in a private communication, by Prof. Gray who recently jointly published [6] some results on this technique using Fourier series techniques rather than arguments from the sampling theory as we have done. A type-1 member signal of L_1 is one that is zero-mean and uniformly distributed over the amplitude range of the quantizer's input. Three other types of signals, called type 2, 3 and 4, can also be obtained from the basic type-1 signal as shown in [7]. It can easily be shown that if $R \ \varepsilon L_1$ and $X \varepsilon L_1$, then $(X + R) \ \varepsilon L_1$ [5]. In this approach to estimating μ_{Qp}, p quantizers are needed, with each one being fed with $(X + R_i)$, $i = 1,...,p$, where R_i, is statistically independent of R_j for $i \neq j$. Here, the quantizers' outputs, X_{Qi}, $i = 1, ..., p$, are all different from each other because each quantizer uses a different reference (or dither) signal. Furthermore, in this case, it is shown in [5] that the transformation $T_p(x)$ will take on the form of its p-D equivalent, i.e. $T_{1,...,1}(x,...,x)$ which, because of the statistical independence imposed on the R_i's, simply reduces to

$$T_{1,\ldots,1}(x,\ldots,x) = \prod_{i=1}^{p} T_1(x) \qquad (17)$$

Now, using Theorem 2, $T_1(x)$ can easily be shown to be:

$$T_1(x) = x \qquad (18)$$

It is now clear from (18) that all member signals of L_1 linearize the mean-sense Input/Output characteristics of the classical quantizer.

The moments recovery equation then becomes:

$$\mu_{Qp} = E\left[\prod_{i=1}^{p} X_{Qi}\right] = E\left[\prod_{i=1}^{p} T_1(x)\right] \qquad (19)$$

$$= E\left[X^p\right] = \mu_p$$

It is clear from (19) that, unlike in the first approach, no corrections whatsoever are involved here and only one, as opposed to (p-1), accumulator is needed. However, the requirements for (p-1) two-input multipliers is common to both approaches. As in the first approach, if 1-bit quantization is used, it would then result in a direct exact moment recovery scheme enjoying all the aforementioned practical attributes.

5 Simulation Results

The second approach was used to carry out some simulation work on estimating the autocorrelation function of both a noise-free and a noisy sinewave using 3 types of correlators: the Sampled-Data (SDC), the Modified Relay (MRC) and the Modified Polarity Coincidence (MPCC) Correlators. These sinewave signals have all been rendered type-1 members of L_1 by adding to each of them a type-1 reference signal simulated by either a 10- or 11-bit Pseudo Random Binary Signal (PRBS). Although not an exact member of L_1 as its Probability Density Function (PDF) is not exactly uniform, this type of reference signal was chosen for its combined advantage of providing an excellent

approximation to an exact member of L_1 and having an easy digital hardware implementation. Some representative results from our simulation work are shown in the opposite figures. All figures clearly show that excellent autocorrelation estimation accuracy is achieved in both the noise-free and noisy cases and is in a very good agreement with the theoretical expectations.

Finally, due to the relationship between moments and cumulants, an important application of this work is in the area of exact recovery of cumulants from their 1-bit quantized counterparts. This is being currently investigated.

Acknowledgement

The support of King Fahd University of Petroleum and Minerals is acknowledged.

References

1. L. Cheded, "Stochastic Quantization: theory and application to moments recovery," Ph.D thesis, UMIST (University of Manchester), Manchester, U.K., August, 1988.

2. B. Widrow, "Statistical Analysis of Amplitude-Quantized Sampled-Data Systems," Trans. Amer. Inst. Elec. Eng., Part II: Applications and Industry, Vol. 79, pp. 555-568, January 1961.

3. L. Cheded, P. A. Payne, "The Exact Impact of Amplitude Quantization on Multi-Dimensional, High-Order Moments Estimation," (accepted for publication in the Signal Processing Journal).

4. L. Cheded, "Some New Results on the Exact Moments Recovery using Deterministic Amplitude Quantization" (in progress, to be submitted to IEE Proceedings F).

5. L. Cheded, "Exact Moments Recovery using Stochastic Amplitude Quantization: Some New Results," (under revision, to be re-submitted to IEEE Trans. Information Theory).

6. R. M. Gray, T. G. Stockham, "Dithered Quantizers," IEEE Trans. IT, Vol. 39, No. 3, May 1993.

7. P. W. Wong, "Quantization Noise, Fixed-Point Multiplication, Round-off Noise and Dithering," IEEE Trans.ASSP, Vol. 38, No. 2, February, 1990.

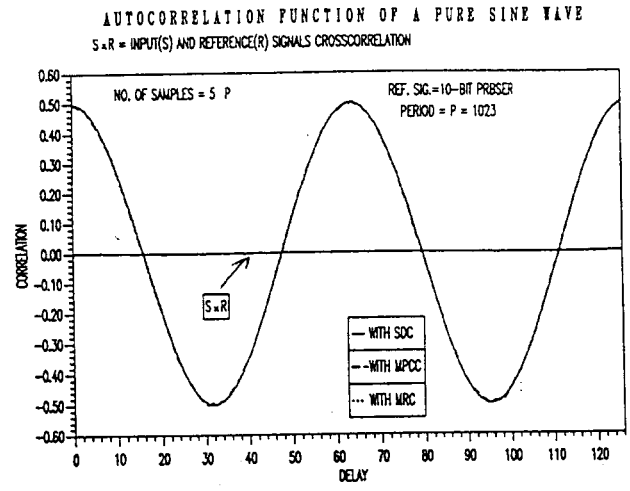

Figure 1

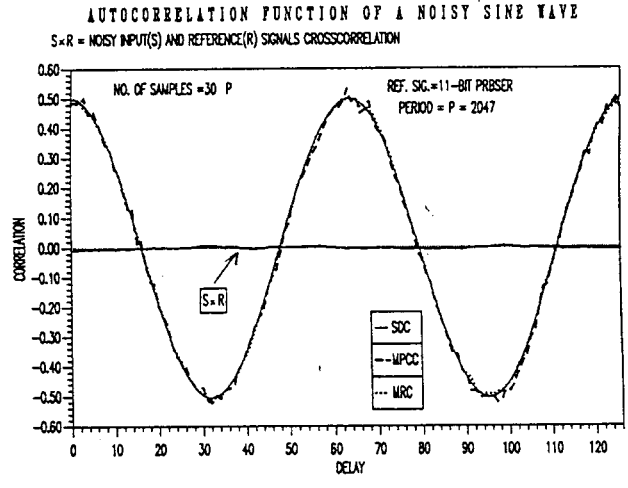

Figure 2

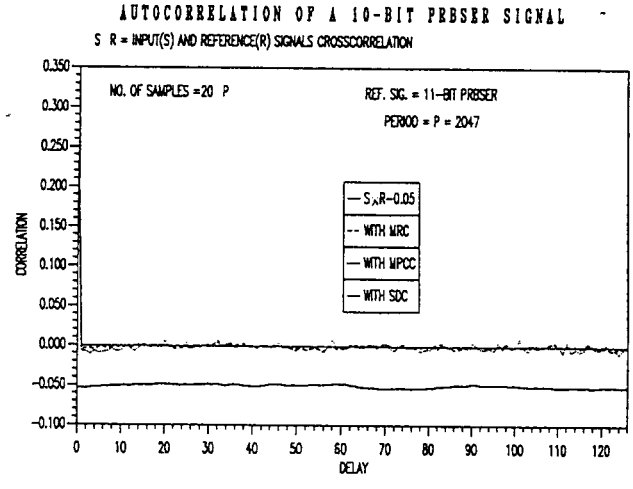

Figure 3