Non-harmonic Fourier Analysis of A/D Conversion

Zoran Cvetković
Department of Electronic Engineering
King's College London
Strand, London WC2R 2LS, UK
Email:zoran.cvetkovic@kcl.ac.uk

***Abstract*— Oversampled A/D conversion is traditionally studied using statistical approach. In this paper we review results of deterministic analysis of oversampled A/D conversion which revealed some surprising facts about its actual accuracy and rate-distortion characteristics and lead to the construction of a single-bit oversampled A/D conversion scheme which attains an exponential rate-distortion characteristics.**

I. Introduction

Analog-to-digital (A/D) conversion involves discretization of an input analog signal in time, implemented by means of regular sampling with an interval τ, followed by amplitude discretization using uniform scalar quantization with a quantization step q. An A/D converter is illustrated in Figure 1. If the input signal is bandlimited, and τ is smaller than the Nyquist sampling interval, τ_N, then the signal can be perfectly reconstructed from its samples. On the other hand, the amplitude discretization introduces an irreversible loss of information. Precise mathematical characterization of this loss of information, *i.e.* the accuracy of A/D conversion, remained evasive for many decades. To cite R. M. Gray, "*Deceptively simple in its description and construction, the uniform quantizer has proved to be surprisingly difficult to analyze*" [1].

Traditionally, quantization error is modelled as an *iid* process, uniformly distributed in the $(-q/2, q/2)$ range [2]. This approach leads to the following result about the mean square error between an input bandlimited signal f and a signal f_r reconstructed from quantized samples of f by means of linear filtering:

$$E(|f_r(t) - f(t)|^2) = \frac{q^2}{12}\frac{\tau}{\tau_N} \ . \tag{1}$$

The conclusion drawn from this formula is that by means of oversampling the accuracy of A/D conversion can be increased beyond the limits imposed by the precision of the quantizer. From the rate distortion perspective, it was traditionally believed that this approach to attaining high conversion accuracy was suboptimal since the bit-rate R increases linearly with the oversampling ratio $r = \tau_N/\tau$, leading to a rate distortion characteristics of the form

$$E(|f_r(t) - f(t)|^2) = c_1\frac{q^2}{12}\frac{1}{R} \ , \tag{2}$$

whereas, if the accuracy is increased by refining quantization, the error decays exponentially in the bit-rate.

$$E(|f_r(t) - f(t)|^2) = c_2 2^{-R}\frac{\tau}{\tau_N} \ . \tag{3}$$

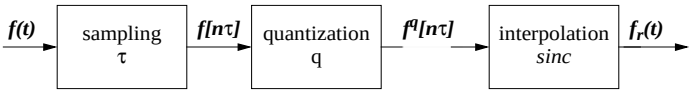

Fig. 1. *Block diagram of simple oversampled A/D conversion followed by classical linear reconstruction.*

Still, due to the costs involved in the production of high-resolution quantizers, high accuracy of modern techniques for A/D conversion is achieved through oversampling.

However, even an elementary deterministic analysis reveals that the statistical approach is very misleading about the actual accuracy of oversampled A/D conversion. In particular, the error between f and its reconstructed version f_r, obtained at some finite τ by means of sinc[1] interpolation (ideal low-pass filtering) is given by

$$e_\tau(t) = f(t) - f_r(t) = \tau\sum_n (f(n\tau) - f^q(n\tau))\mathrm{sinc}(t - n\tau) \ ,$$

where $f^q(n\tau)$ is the quantized version of $f(n\tau)$. When τ tends to zero, the error signal converges toward

$$e(t) = \lim_{\tau\to 0} e_\tau(t) = \int_{-\infty}^{\infty}(f(s) - f^q(s))\mathrm{sinc}(t - s)ds \tag{4}$$

which is not the zero function. The discrepancy between the expression in (1) and this observations about the asymptotic error behaviour is due to the fact that the statistical model is not valid for high oversampling ratios when correlations between quantization errors are more pronounced. And contrary to what could be expected from this negative result about the accuracy of oversampled A/D conversion, many classes of signals can actually be reconstructed with a higher accuracy than predicted by (1), but that requires more sophisticated methods than linear filtering.

In this paper, we review some relatively recent results of deterministic analysis of oversampled A/D conversion [4], [3], which established some surprising facts about its actual accuracy and rate-distortion characteristics and lead to a construction of a single-bit oversampled A/D conversion scheme which attains an $O(\tau^2)$ error behaviour and exponential rate-distortion characteristics [5], [6].

[1]We use the notation $\mathrm{sinc}(t) = \sin\pi t/\pi t$.

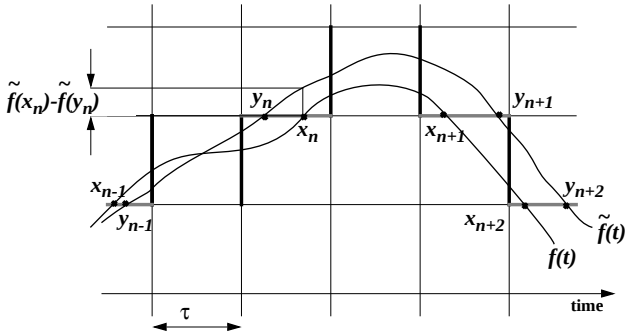

Fig. 2. *Quantization threshold crossings of an analog signal f and its consistent estimate $\tilde{f}$. If f goes through a certain quantization threshold at a point x_n, then $\tilde{f}$ has to cross the same threshold at a point y_n in the same sampling interval. The error amplitude at the point x_n is equal to $|\tilde{f}(x_n) - f(x_n)| = |\tilde{f}(y_n) - \tilde{f}(x_n)|$.*

II. Deterministic Analysis of Oversampled A/D Conversion

Consider oversampled A/D conversion of signals in the space of π-bandlimited functions in $L^2(\mathbb{R})$. We shall denote this space by $\mathcal{V}_\pi$. Let $f \in \mathcal{V}_\pi$ be at the input of an A/D converter with a sampling interval τ and a quantization step q, as shown in Figure 1. The question which we are trying to answer is: how accurately f can be reconstructed from the sequence of its quantized samples $(f_n^q[n\tau])_{n\in\mathbb{Z}}$. To this end, consider a signal $\tilde{f} \in \mathcal{V}_\pi$ which is a *consistent estimate* of f, that is a signal which at the output of the A/D converter produces the same sequence of quantized samples as f. Since $f^q[n\tau] = \tilde{f}^q[n\tau],\ \forall n \in \mathbb{Z}$, if f has a quantization threshold crossing at some time instant $t = x_n$, then $\tilde{f}$ must cross the same quantization threshold at some instant $t = y_n$ which lies in the same sampling interval with x_n, and vice versa, as illustrated in Figure 2. This implies that the difference between f and $\tilde{f}$ at $t = x_n$ is proportional to τ, in particular

$$|\tilde{f}(x_n) - f(x_n)| = |\tilde{f}(y_n) - \tilde{f}(x_n)| \leq \tilde{f}'(\xi_n)\tau \qquad (5)$$

where ξ_n is a point in the interval between x_n and y_n (see Figure 2). As $\tau \to 0$, for any consistent estimate $\tilde{f}$ of f, the difference $|\tilde{f}(x_n) - f(x_n)|$ is progressively smaller, and in the limit when $\tau = 0$, $\tilde{f}(x_n) = f(x_n)$ at every quantization threshold crossing x_n of f. Does this imply that in the limit $\tilde{f}(t) = f(t),\ \forall t \in \mathbb{R}$, and that any sequence of consistent estimates converges in L^2 sense to f as $\tau \to 0$? The answer depends on the sampling properties of the sequence of quantization threshold crossings $(x_n)_{n\in\mathbb{Z}}$ in $\mathcal{V}_\pi$.

There are three possible scenarios.

1. $(x_n)_{n\in\mathbb{Z}}$ is not a *sequence of uniqueness* [7] in $\mathcal{V}_\pi$. In this case, even in the limit, $\tau = 0$, when $\tilde{f}(x_n) = f(x_n),\ \forall x_n$, $\tilde{f}$ and f can be arbitrarily different at other time instants.

2. $(x_n)_{n\in\mathbb{Z}}$ is a *sequence of uniqueness* but not a sequence of stable sampling in $\mathcal{V}_\pi$. For $\tau = 0$, in this case $\tilde{f}(t) = f(t),\ \forall t \in \mathbb{R}$, however for any finite τ, even though $\tilde{f}$ and f are close at all $t = x_n$, there is no guarantee that they would be close at other time instants.

3. $(x_n)_{n\in\mathbb{Z}}$ is *a sequence of stable sampling in $\mathcal{V}_\pi$* [7]. When the stable sampling condition is satisfied, any sequence of consistent estimates converges in L^2 sense to f at a rate established by the following theorem [3].

Theorem 2.1: Let the sequence of quantization threshold crossings $(x_n)_{n\in\mathbb{Z}}$ of a signal f in $\mathcal{V}_\pi$, be a uniformly discrete sequence of stable sampling in $\mathcal{V}_\pi$. There exists a $\delta > 0$ such that for all $\tau \leq \delta$, any consistent estimate $\tilde{f}$ of f satisfies

$$\|\tilde{f} - f\|^2 \leq c_f \|f\|^2 \tau^2, \qquad (6)$$

where c_f is a constant which depends on $(x_n)_{n\in\mathbb{Z}}$ but does not depend on τ or $\tilde{f}$.

Hence, there exist signals for which the accuracy of oversampled A/D conversion cannot be improved by increasing the oversampling ratio, and there are also signals for which the error converges to zero as $\tau \to$, and if that is the case the error decays proportionally to τ^2 rather than τ. While some intuition about these two classes of signals was given in [3], which demonstrated that both classes comprise significant sets of bandlimited signals, their precise characterization is very intricate.

The stronger error bound of Theorem 2.1 has quite appealing practical implications, however the constant c_f in (6) depends on a particular signal, hence the τ^2 accuracy is not uniform on sufficiently general sets of bandlimited signals. Another important issue is finding practical reconstruction algorithms which attain the accuracy asserted by Theorem 2.1.

The reconstruction can be approached as the problem of interpolation of a bandlimited signal from an irregular sequence of samples $(f(x_n))_{n\in\mathbb{Z}}$ where each x_n is known with some finite precision, $\tau/2$, whereas $f(x_n)$'s are known with infinite precision since they are equal to corresponding quantization threshold. For good numerical properties of practical algorithms for reconstruction of f from this information, the density of $(x)_{n\in\mathbb{Z}}$ needs to be greater than 1 [5] (above the Nyquist rate), however, it turns out that the density of quantization threshold crossings of a signal in $\mathcal{V}_\pi$ cannot be greater than 1 [3].

Let us now reconsider the bit-rate of oversampled A/D conversion. Observe that the sequence of quantized samples $(f^q[n\tau])_{n\in\mathbb{Z}}$ of f is completely specified by the corresponding sequence of quantization threshold crossings [4]. To encode the sequence of quantization threshold crossings, we first partition the time axis into consecutive segments of length $T > \tau$. Then we represent the m-th quantization threshold crossing on a given segment T as (i_m, l_m), where i_m is the position of the sampling interval in which the crossing occurs within T, and l_m is the level of the corresponding threshold. Since the bit-rate needed for encoding levels l_m does not depend on τ, and the bit-rate needed to encode indices i_m increases as $\log_2(T/\tau)$, the overall bit-rate of this encoding scheme is

$$R = C_1 + C_2 \log_2(T/\tau) \quad , \qquad (7)$$

where C_1 and C_2 are constants which do not depend on τ [4]. Finally, if the error converges to zero as established by Theorem 2.1, with this encoding scheme the rate-distortion characteristics of oversampled A/D conversion becomes

$$\|\tilde{f} - f\|^2 \leq c_{1,f} 2^{-\alpha_f R} \tag{8}$$

where $c_{1,f}$ and α_f are constants which depend on f but do not depend on τ. Similarly, even with the linear reconstruction, for moderate oversampling ratios for which the statistical model of the quantization error is valid, it follows from (1) that the rate-distortion characteristics of oversampled A/D conversion with the quantization threshold crossings encoding is

$$E(|f_r(t) - f(t)|^2) \leq c_{2,f} 2^{-\alpha_f R/2} \ . \tag{9}$$

The considerations in this section reveal both the potential of oversampled A/D conversion to achieve a superior accuracy and rate distortion characteristics than suggested by the traditional point of view, as well as its limitations and problems that need to be solved in order to guarantee a uniform $O(\tau^2)$ error behaviour and exponential rate distortion characteristics on broad classes of bandlimited signals. These lead to the construction of a single-bit A/D conversion scheme [5], described in the following sections, which resolves all the problems involved by introducing a deterministic dither function.

III. A Dithered A/D Conversion Scheme

We shall consider signals in the set $\mathcal{C} = \{f : f \in \mathcal{V}_\pi, \ \|f\|_\infty \leq 1\}$, that is, the set of π-bandlimited signals with finite energy and amplitude bounded by 1.

The single-bit dithered A/D converter is defined by means of a dither function d and a parameter $\lambda > 1$. We shall assume that the dither function satisfies conditions which ensure that the composite signal $f + d$ changes sign on every interval $I_n = (n/\lambda - 1/(2\lambda), n/\lambda + 1/(2\lambda))$, for every f in $\mathcal{C}$. In particular, we shall require that d is a C^1-function that for all $n \in \mathbb{Z}$ satisfies

$$\left| d\left(\frac{n}{\lambda} + \frac{1}{2\lambda}\right) \right| \geq \gamma > 1 \ , \tag{10}$$

$$\operatorname{sgn}\left[d\left(\frac{n}{\lambda} - \frac{1}{2\lambda}\right)\right] = -\operatorname{sgn}\left[d\left(\frac{n}{\lambda} + \frac{1}{2\lambda}\right)\right] \ . \tag{11}$$

An example of an appropriate dither is the sine function, $d(t) = \gamma \sin(\lambda \pi t)$. The sequence $|(f + d)(n/\lambda + 1/(2\lambda))|$, $n \in \mathbb{Z}$, therefore alternates in sign, hence, there must be at least one zero-crossing of $f + d$ in every interval I_n. We can then select one zero-crossing t_n in every interval I_n, forming a sequence $(t_n)_{n\in\mathbb{Z}}$ which is sufficiently dense to form a sequence of stable sampling in $\mathcal{V}_\pi$. This motivates the following definition [5].

Definition 3.1: Let d be a bounded C^1-function satisfying (10) and (11), and let $\lambda > 1$ be a fixed parameter. The single-bit dithered oversampled analog-to-digital converter, $\mathbf{D}^d_{\lambda,\tau}$, is defined as the operator $\mathbf{D}^d_{\lambda,\tau} : \mathcal{C} \to \ell^\infty(\mathbf{Z})$ given by

$$(\mathbf{D}^d_{\lambda,\tau} f)[n] = \min\{m : m \in \mathbb{Z}, m\tau \in I_n, \\ \operatorname{sgn}[(f + d)(m\tau)] \neq \operatorname{sgn}[(f + d)(m\tau + \tau)]\} - \mu_n$$

where $\mu_n = \lfloor n/\lambda\tau \rfloor$.

The output of the converter is a sequence of indices of sampling intervals where zero-crossings of $f + d$ occur; one zero-crossing within each interval I_n. For simplicity, in the definition of the converter we choose this to be the first zero-crossing in I_n, but any other selection algorithm would work as well.

The bit-stream produced in this conversion allows for pointwise reconstruction of any signal in $\mathcal{C}$ with $O(\tau)$ accuracy, uniformly in time and on $\mathcal{C}$, as asserted by the following theorem [5].

Theorem 3.2: If $f, \tilde{f} \in \mathcal{C}$ satisfy $\mathbf{D}^d_{\lambda,\tau}\tilde{f} = \mathbf{D}^d_{\lambda,\tau} f$, then

$$|\tilde{f}(t) - f(t)| < c\tau \ ,$$

uniformly on t, where c is independent of τ, f, or $\tilde{f}$.

The bit-rate, R, of this conversion scheme is determined by the number of sampling intervals within each interval I_n of size $1/\lambda$. Here λ will be kept fixed, and τ will typically be significantly smaller than $1/\lambda$. Thus the bit-rate needed for specifying the location of one data change within I_n with precision τ satisfies

$$R \leq -\lambda \log_2(\tau\lambda) + 1 \ . \tag{12}$$

This result on the bit-rate along with the result of Theorem 3.2 then imply that there exist a positive constants c_1 such that for any conversion consistent estimate $\tilde{f}$ of f

$$|g(t) - f(t)| < c_1 2^{-R/\lambda} \quad ,$$

uniformly in t, and uniformly for f in $\mathcal{C}$.

Let us now focus on explicit local-reconstruction algorithms that attain the accuracy asserted by Theorem 3.2. Observe that for every interval I_n, the $\mathbf{D}^d_{\lambda,\tau}(f)$ gives us the value of a time instant $s_{n,k} = k\tau + n/\lambda$ such that $\operatorname{sgn}((f + d)(s_{n,k})) = -\operatorname{sgn}((f+d)(s_{n,k+1}))$,[2] implying that $(f+d)(t)$ must be zero for some $t \in (s_{n,k}, s_{n,k+1})$. Let us define,

$$t_n := s_{n,k} + \tau/2; \tag{13}$$

since $|f'(t)| \leq \pi\|f\| \leq \pi$ for all t (this follows from $f \in \mathcal{V}_\pi$, see [10]) and $|d'(t)| \leq \|d\|_{C^1} =: \Delta$, it follows that

$$f(t_n) = -d(t_n) + \epsilon_n \quad , \quad \text{where} \quad |\epsilon_n| \leq (\Delta + \pi)\tau/2 \quad . \tag{14}$$

The bit sequence $\mathbf{D}^d_{\lambda,\tau}(f)$ thus specifies the sampling sequence $(t_n)_{n\in\mathbb{Z}}$ and, within an error proportional to τ, the values of f at these sampling points. The sequence $(t_n)_{n\in\mathbb{Z}}$ is uniformly discrete, i.e. $\inf_{n,k\in\mathbb{Z}, n\neq k} |t_n - t_k| > 0$. Moreover, the lower uniform density of $(t_n)_{n\in\mathbb{Z}}$ equals $\lambda > 1$. Therefore, $(t_n)_{n\in\mathbb{Z}}$ constitutes a sequence of stable sampling for all the spaces $\mathcal{V}_{\mu\pi}$ for all $\mu < \lambda$ [7], [9], and thus for $\mathcal{V}_\pi$ and $\mathcal{C}$. Hence, any function f in $\mathcal{C}$ can be reconstructed

[2] We assume, for simplicity, that τ divides $1/\lambda$.

from its samples $(f(t_n))_{n\in\mathbb{Z}}$ in a numerically stable manner. In our case, these samples are, however, known only at a finite precision. The following theorem asserts that there exist C^∞ functions ψ_n such that

$$\psi_n(t) \le \frac{C_N}{(1+|t|)^{-N}} \quad \text{for all} \quad n \in Z,\ t \in I\!R, N \ge 1\ , \quad (15)$$

(where C_N are constants that do not depend on f), which can be linearly combined to reconstruct f from its quantized samples with $O(\tau)$ accuracy [5].

Theorem 3.3: Let $\mathbf{D}^d_{\lambda\tau}$ be a dithered A/D converter, as described in Definition 3.1, and let f be a function in $\mathcal{C}$. Define a sequence of time instants $(t_n)_{n\in\mathbb{Z}}$ as

$$t_n = \left(\mu_n + (\mathbf{D}^d_{\lambda,\tau} f)[n] + \frac{1}{2}\right)\tau \quad ,$$

where $\mu_n = \lfloor n/\lambda\tau \rfloor$. There exist functions ψ_n of fast decay, as specified by (15), such that the approximation $\tilde{f}$ of f, given by

$$\tilde{f}(t) = -\sum_{n\in\mathbb{Z}} d(t_n)\psi_n(t-t_n) \quad (16)$$

satisfies

$$|f(t) - \tilde{f}(t)| \le C(\Delta+\pi)\tau\ , \quad (17)$$

for all t, where C does not depend on f or τ.

The significance of the fast decay of functions ψ_n is that it guarantees local reconstruction with a small accuracy degradation. In particular, consider reconstructing f at some $t \in (k/\lambda - 1/(2\lambda), k/\lambda + 1/(2\lambda))$ as

$$f_{\mathrm{app}:L}(t) = -\sum_{n=k-L}^{k+L} d(t_n)\psi_n(t-t_n)\ .$$

Then the overall reconstruction error satisfies

$$|f(t) - f_{\mathrm{app}:L}(t)| \le C(\Delta+\pi)\tau + \frac{2\gamma\lambda C_N}{N-1}\left(\frac{\lambda}{L}\right)^{N-1} \ , \quad (18)$$

for all $t \in (k/\lambda - 1/(2\lambda), k/\lambda + 1/(2\lambda))$ and for all $N \ge 2$, where constants γ, λ and C_N are as defined in the above. This bound does not depend on k and is uniform in f on $\mathcal{C}$.

IV. Reconstruction Using Finite Interpolation

The results presented in the previous section prove the $O(\tau)$ accuracy of the proposed single-bit A/D conversion scheme. In this section we consider a practical reconstruction algorithm which comes very close to the $O(\tau)$ accuracy. The approach we take is to interpolate between the quantized samples $(f(t_k))_{k\in Z}$, as specified by (13) and (14), to estimate samples $(f(n/\lambda))_{n\in\mathbb{Z}}$, and then synthesize f using an appropriate function φ of fast decay as

$$f(t) = \sum_{n\in\mathbb{Z}} f\left(\frac{n}{\lambda}\right)\varphi\left(t-\frac{n}{\lambda}\right) \quad . \quad (19)$$

For that purpose one can use any function φ such that its Fourier transform $\hat{\varphi}$ is C^∞, and satisfies $|\hat{\varphi}(\omega)| = 0$ for $|\omega| > \lambda\pi, \hat{\varphi}(\omega) = \frac{1}{\sqrt{2\pi}}$ for $|\omega| \le \pi$, and $0 < \hat{\varphi}(\omega) < \frac{1}{\sqrt{2\pi}}$ for $\pi < |\omega| \le \lambda\pi$ (see Figure 3). If the samples $f(\frac{n}{\lambda})$ in (19) are replaced by perturbed values $f^*_n = f(\frac{n}{\lambda}) + \varepsilon_n$, with $|\varepsilon_n| \le \varepsilon$ for all n, then the resulting sum approximates f within an error proportional to ε, uniformly in t:

$$\left| f(t) - \frac{1}{\lambda}\sum_n f^*_n \varphi\left(t-\frac{n}{\lambda}\right)\right| \le \frac{\varepsilon}{\lambda}\sum_n \left|\varphi\left(t-\frac{n}{\lambda}\right)\right| \le C\varepsilon \quad . \quad (20)$$

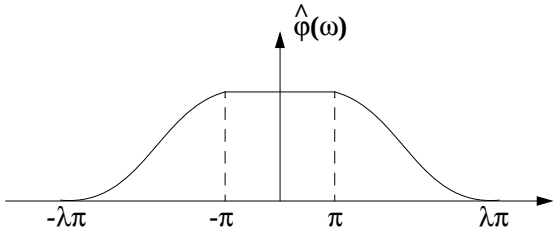

Fig. 3. *π-bandlimited signals can be reconstructed from their samples taken at $t_n = n/\lambda,\ n \in \mathbb{Z},\ \lambda > 1$ as $f(t) = \sum f(n/\lambda)\varphi(t-n/\lambda)$, where φ is any function such that $|\hat{\varphi}(\omega)| = 0$ for $|\omega| > \lambda\pi$ and $\hat{\varphi}(\omega) = 1/\sqrt{2\pi}$ for $|\omega| \le \pi$. That allows to design $\hat{\varphi}(\omega)$ so that it is C^∞, which makes φ decay faster than any inverse polynomial in the time domain.*

Consider now estimating samples $(f(n/\lambda))_{n\in\mathbb{Z}}$ from quantized samples $(f(t_n))_{n\in Z}$ using Lagrangian interpolation. Without loss of generality, we focus on interpolating f at $t = 0$ from $f(t_k),\ k = -L, -L+1, \ldots, L$. Observe that since the dither function has maximum amplitude $\gamma > 1$, C^1 norm $\Delta := \sup|d'(t)|$, and attains either γ or $-\gamma$ between each two zero crossings, then t_n's must be separated by at least $\delta := \frac{2(\gamma-1)}{\Delta+\pi} - \tau$. Let $f_{\mathrm{app}:L}(0)$ denote the approximation to $f(0)$ computed by Lagrange interpolation from $f(t_l)$ with $|l| \le L$. In general, the Lagrange interpolation $g_{\mathrm{approx},K}(x)$ of a function g that is K times continuously differentiable, based on the values values $g(x_1), \cdots, g(x_K)$, satisfies the bound

$$|g(x) - g_{\mathrm{app},K}(x)| \ \le\ \frac{1}{K!}\sup_y |g^{(K)}(y)| \prod_{k=1}^{K} |t-t_k| \quad .$$

In our case $|f^{(l)}(t)| \le \pi^l$, because f is in $\mathcal{V}_\pi$ and $|f(t)| \le 1$. It follows that [5]

$$|f(0) - f(0)_{\mathrm{app}:L}| \ \le\ \sqrt{L}\left(\frac{\pi}{2\lambda}\right)^{2L+1} \quad . \quad (21)$$

This error decreases exponentially as L increases. However, as explained earlier, our reconstruction does not use the exact $f(t_n)$, but approximate values that are within $(\Delta+\pi)\tau/2$ of the true $f(t_n)$. We therefore also need to estimate the error between $f_{\mathrm{app}:L}$ and $\tilde{f}_{\mathrm{app}:L}$, which is the Lagrange interpolation between the approximate values of $f(t_n)$. The explicit form of the Lagrange interpolation formula allows us to bound this as [5]

$$|f(0)_{\mathrm{app}:L} - \tilde{f}(0)_{\mathrm{app}:L}| \ \le C_1\ \frac{(\Delta+\pi)\tau}{\lambda\delta}L^2 \quad . \quad (22)$$

Combining the two estimates, we find

$$\left|f(0) - \tilde{f}(0)_{\mathrm{app}:L}\right| \ \le\ \sqrt{L}\left(\frac{\pi}{2\lambda}\right)^{2L+1} \ +\ C_1\frac{(\Delta+\pi)\tau L^2}{\lambda\delta} \quad .$$

Hence, the error has the form $|f(0) - \tilde{f}_{\mathrm{app}:L}(0)| \leq a_1 L^\gamma e^{-\alpha L} + a_2 \tau L^\beta$, where $\alpha = 2|\log(\pi/2\lambda)|$, $\gamma = 1/2$ and $\beta = 2$. We select the interpolation order L so that the interpolation error is kept below the quantization error; in particular, $a_1 e^{-\alpha L} \leq a_2 \tau^{1+\epsilon}$ for some $\epsilon > 0$. It follows that for

$$L \geq \frac{1}{\alpha} \log \frac{a_2}{a_1} + \frac{1+\epsilon}{\alpha} |\log \tau| \tag{23}$$

the approximation error can be bounded as

$$|f(0) - \tilde{f}(0)_{\mathrm{app}:L}| < c_2 \tau |\log c_3 \tau|^\beta \quad . \tag{24}$$

If f is now synthesized from interpolated samples according to (19), the error bound is provided by (20), with $\varepsilon = c_2 \tau |\log c_3 \tau|^\beta$. The error of local reconstruction, when f is synthesized using finitely many expansion terms around a point of interest, exhibits the same fast decay in the number of expansion terms as the truncation error given in (18).

Observe that for Lagrangian interpolation to be effective, in the sense that the interpolation error can be reduced to a level below the quantization error one needs $\lambda > \pi/2$. Lagrangian interpolation also causes accumulation of the quantization error according to a power law in the interpolation order that ultimately causes degradation of the overall conversion accuracy by a factor $O(|\log \tau|^\beta)$. Another practical reconstruction scheme which is valid for any $\lambda > 1$ and improves the overall accuracy over the Lagrangian approach from $O(\tau |\log c_4 \tau|^\beta)$ to $O(\tau)$ was proposed in [6]. A particularly elegant and accurate reconstruction scheme can be used if the dither is the cosine function $d(t) = \gamma \cos \lambda \pi t$ with $\lambda > 3$ [6]; that scheme is reviewed in the following section.

V. Reconstruction for Cosine Dither with $\lambda > 3$

In the case of cosine dither $d(t) = \gamma \cos(\lambda \pi t)$ with $\lambda > 3$, $\gamma > 1$, f can be reconstructed from its samples taken at zeros of $f + d$ as if they are regular samples. In particular, f can be represented as

$$f(t) = \frac{1}{\lambda} \sum_n f(\xi_n) K(t - \xi_n) \ , \tag{25}$$

where ξ_n are zero crossings of $f + d$, and K is an L^1 function such that $\hat{K}(\omega) = 1$ for $|\omega| < \pi$, and $\hat{K}(\omega) = 0$ for $|\omega| > \lambda\pi - 2\pi > \pi$. This result was established in [11]. Positions of zero crossings of $f(t) + \gamma \cos \lambda \pi t$ are in our A/D conversion scheme known only at a finite precision, hence we reconstruct f as

$$\tilde{f}(t) = \frac{1}{\lambda} \sum_n (f(\xi_n) + \delta_n) K(t - \xi_n - \theta_n) \quad , \tag{26}$$

where $|\delta_n| < c\tau$ and $|\theta_n| < \tau/2$. The error made by this approximation can be shown to be bounded as [6]

$$|\tilde{f}(t) - f(t)| < \frac{c_4}{\lambda} \tau \quad . \tag{27}$$

We now design K such that $\int_{|t|>L/\lambda} |K(t)| dt \leq e^{-\alpha L}$, which would make the error of local reconstruction decay exponentially in the length of the interpolation interval. A function K of this kind is given by

$$K(t) = \frac{L/\lambda}{\sinh \varepsilon L/\lambda} \frac{\sin((\pi + \varepsilon)t)}{\pi t} \frac{\sin \varepsilon \sqrt{t^2 - (L/\lambda)^2}}{\sqrt{t^2 - (L/\lambda)^2}} \tag{28}$$

where $\varepsilon = (\lambda\pi - 3\pi)/2$. Observe that the factor $\sin \varepsilon \sqrt{t^2 - (L/\lambda)^2} / \sqrt{t^2 - (L/\lambda)^2}$ in the expression for K is an entire function of exponential type at most ε, so it is an L^2 function bandlimited to ε. Therefore, K is bandlimited to $\lambda\pi - 2\pi > \pi$ and $\hat{K}(\omega) = 1$ for $|\omega| < \pi$. One can verify that K is in L^1, and that $\int_{|t|>L/\lambda} |K(t)| dt \leq e^{-\varepsilon L/\lambda}$.

If such a function K is used for reconstruction of f from quantized $2L + 1$ samples $f(t_n)$ around t according to (26), the resulting error can be bounded as

$$|f(t) - f_{\mathrm{app}:L}(t)| < \frac{c_4}{\lambda} \tau + \frac{c_5}{\lambda} e^{-\alpha L} \ . \tag{29}$$

Hence, when

$$L > \frac{1+\epsilon}{\alpha} |\log \tau| + \frac{1}{\alpha} \log \frac{c_4}{c_5} \quad ,$$

for some $\epsilon > 0$, we obtain

$$|f(t) - f_{\mathrm{app}:L}(t)| < \frac{2c_4}{\lambda} \tau \quad . \tag{30}$$

To summarize: in the case of cosine dither, $d(t) = \gamma \cos \lambda \pi t$, $\lambda > 3$, we are able to reconstruct f using the simple quadrature formula in (25) [11]. With interpolation of this kind using kernel K given by (28) , local reconstruction from $O(|\log \tau|)$ quantized samples attains $O(\tau)$ accuracy.

Acknowledgements

The author is grateful to I. Daubechies, B. F. Logan, and M. Vetterli for their collaboration on this topic.

References

[1] R. M. Gray. *Source Coding Theory*. Kluwer Academic Publishers, 1990.

[2] W. R. Bennett. Spectra of quantized signals. *Bell System Technical Journal.* Vol. 27, pp. 446-472, July 1948.

[3] Z. Cvetković and M. Vetterli. On simple oversampled A/D conversion in $L^2(\mathbf{R})$. *IEEE Trans. on Information Theory.* Vol. 47, No. 1, pp. 146–154, January 2001.

[4] Z. Cvetković and M. Vetterli. Error-rate characteristics of oversampled analog-to-digital conversion. *IEEE Trans. Information Theory.* Vol. 44, No. 5, pp. 1961–1964, September 1998.

[5] Z. Cvetković and I. Daubechies. Single-Bit Oversampled A/D conversion with Exponential Accuracy in the Bit-Rate. *Proc. Data Compression Conference*, DCC 2000, pp. 343–352, Snowbird, Utah, March 2000.

[6] Z. Cvetković, I. Daubechies, and B. F. Logan. Interpolation of Bandlimited Functions from Quantized Irregular Samples. *Proc. Data Compression Conference*, DCC 2002, pp. 412–421, Snowbird, Utah, April 2002.

[7] H. J. Landau. Sampling, Data Transmission, and the Nyquist Rate. *Proceedings of the IEEE.* Vol. 55, No. 10, pp. 1701–1706, October 1967.

[8] A. Beurling. *The Collected Works of Arne Beurling, Vol. 2, Harmonic Analysis,* (L. Carleson, P. Malliavin, J. Neuberger, and J. Wermer, Eds.), pp. 341–365, Birkhäuser, Boston, 1989.

[9] S. Jaffard. A Density Criterion for Frames of Complex Exponentials. *Michigan Math. Journal.* Vol. 38, pp. 339–348, 1991.

[10] R. M. Young. *An Introduction to Nonharmonic Fourier Series.* Academic Press, New York, 1980

[11] B. F. Logan. Signals Designed for Recovery after Clipping - I. Localization of Infinite Products. *AT&T Bell Laboratories Technical Journal*, Vol. 63, No. 2, pp. 261–285, February 1984.