Fifty-third Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 29 - October 2, 2015

Optimal Trade-off Between Sampling Rate and Quantization Precision in A/D conversion

Alon Kipnis, Yonina C. Eldar and Andrea J. Goldsmith

***Abstract*—The jointly optimized sampling rate and quantization precision in A/D conversion is studied. In particular, we consider a basic pulse code modulation A/D scheme in which a stationary process is sampled and quantized by a scalar quantizer. We derive an expression for the minimal mean squared error under linear estimation of the analog input from the digital output, which is also valid under sub-Nyquist sampling. This expression allows for the computation of the sampling rate that minimizes the error under a fixed bitrate at the output, which is the result of an interplay between the number of bits allocated to each sample and the distortion resulting from sampling. We illustrate the results for several examples, which demonstrate the optimality of sub-Nyquist sampling in certain cases.**

I. INTRODUCTION

Representing an analog signal by a sequence of bits leads to a fundamental trade-off between the minimal distortion in the reconstruction of the signal from this sequence and its bitrate. This is described by the distortion-rate function (DRF) of the analog source. While the DRF gives the minimal distortion only as a function of the bitrate of the digital representation, in practice, A/D conversion schemes involve sampling and quantization. Therefore, hardware limitations in sampling and quantization determine current A/D technology. For instance, a key idea in determining the analog DRF is to map the continuous-time process into a discrete-time process based on sampling at or above the Nyquist frequency [1, Sec. 4.5.3]. However, since wideband signaling and A/D technology limitations can preclude sampling signals at their Nyquist rate [2], an optimal source code based on such a discrete-time representation may be impractical in certain scenarios.

An approach combining sampling and source coding, in which an analog Gaussian process is described from a rate-limited version of its samples, was considered in [3]. The main finding of [3] is that sampling at or above the Nyquist rate may not be necessary in order to achieve the DRF $D(R)$. Specifically, for each bitrate R, there exists another fundamental rate f_{RD} which may be smaller than the Nyqusit rate, such that sampling at the rate f_{RD} is enough to achieve $D(R)$. In addition, [3] proves the existence of a range of sampling frequencies for which distortion due to sampling can be traded with distortion due to lossy compression, without affecting the overall distortion sum. This general result serves as the motivation for the present work, where we seek to derive the optimal trade-off between sampling rate and quantization precision to minimize distortion in A/D converters with fixed bitrate outputs.

The optimal rate-distortion performance derived in [4] can be achieved by a vector quantizer applied to an estimate of the source from the samples. However, it was shown in [5] that as the sampling rate goes to infinity, the rate-distortion performance of a scalar quantizer may be dramatically inferior compared to that of a vector quantizer. Indeed, under a fixed bitrate at the output, a faster sampling frequency results in a lower quantization precision, and vice versa. Hence, there may be some A/D converters for which sub-Nyquist sampling minimizes distortion. Our results will thus indicate when sampling at Nyquist or sub-Nyquist rates yields minimum A/D distortion.

A very basic A/D conversion scheme is obtained by sampling and quantizing each sample using a scalar quantizer, which is referred to as pulse code modulation (PCM) [6]. Under this scheme, the overall bitrate R in the resulting digital representation is the product of the sampling rate f_s and the quantizer bit precision q. In this work we analyze A/D conversion via PCM as a source coding scheme: we consider the minimal error as a function of the bitrate R by assuming a statistical model on the input process and mean squared error (MSE) as our performance metric. The quantization distortion is modeled as an additive white noise whose magnitude decreases exponentially with the bits-per-sample q, where $q = R/f_s$. While this model was found to be accurate when the quantizer resolution is relatively high [7], the white noise assumption may not hold under a very coarse quantizer. Nevertheless, an analysis of the minimal MSE from single-bit measurements which does not use the white noise model provided MSE improvement of only up to 3db per octave in the bitrate compared to analysis using the white noise model [8]. This implies that the results in this work would suffer only a minor change under an exact model of the low-resolution quantizer. This approximation does not affect our conclusions which are based only on the scaling behavior of the MSE as a function of the sampling rate and the quantizer resolution. We elaborate more on this approximation in Section II.

When bitrate considerations are ignored, the PCM A/D conversion scheme considered here and higher order schemes such as Sigma-Delta modulation ($\Sigma\Delta$M) benefit from oversampling (sampling above the Nyquist rate of the input signal), which reduces in-band quantization noise. This increases the effective resolution of the quantizer, which is usually taken to be very coarse (typically 1-bit). While these modulators are attractive

978-1-5090-1824-6/15/$31.00 ©2015 IEEE

due to their relatively cheap hardware implementation [9], high correlation between consecutive time samples taken at high sampling rates implies that conventional oversampled modulations cannot lead to an efficient memory utilization without further coding of the samples [10]. Even with the additional coding suggested in [11], the sampling rate required to approach the DRF is still very high compared to sampling at the Nyquist rate. Other oversampled A/D conversion approaches which achieve exponential error reduction with the bitrate were proposed in [12], but so far have not been realized in practice. Since A/D technology limits sampling rates, oversampled A/D may not be practical for applications with signals of wide bandwidth, such as white-space estimation in cognitive radio systems [13], [2]. Moreover, high sampling rates increase the memory requirements of the A/D and the system power consumption.

These challenges in A/D technology motivate us to understand how to sample in a memory-efficient manner. In order to do so, we impose a constraint on the bitrate at the output of the system and examine the trade-off between sampling rate and distortion. We show that under certain assumptions on the signal, the rate-distortion function can be approached by sampling below the Nyquist rate.

The main result of this paper is an expression for the minimal MSE (MMSE) in A/D conversion using PCM under a fixed bitrate R at the output of the modulator. The result is valid for *any* sampling rate, regardless if the input is band-limited or not. This result allows us to compute the sampling rate $f_s^\star$ that minimizes the MMSE for this bitrate. To our knowledge, this is the first analysis of A/D conversion under a fixed bitrate in the sub-Nyquist regime. We show that in the case where the input signal is band-limited, $f_s^\star$ is obtained at the Nyquist rate or below it. This result proves the intuition that super-Nyquist sampling is never optimal. The value of $f_s^\star$ depends on the power spectrum distribution (PSD). The more uniform this distribution, the closer $f_s^\star$ is to the Nyquist rate. By considering several example input signals, we compare the behavior of $f_s^\star$ as a function of R to the minimal sampling rate f_{RD}, as defined in [3], that is needed to achieve the quadratic Gaussian DRF without any constraints on the quantizer.

The rest of this paper is organized as follows: in Section II we provide the relevant background on PCM, MSE estimation in sampling and the distortion-rate function of sampled processes. Our main results and discussion are given in Section III. Concluding remarks are provided in Section IV.

II. Background and Problem Formulation

A. Distortion-Rate Theory of Sampled Processes

An information theoretic bound on the MMSE in any A/D conversion scheme whose output bitrate is constrained to R bits per time unit is given by the DRF of the analog source. For a Gaussian stationary process $X(\cdot)$, this DRF is obtained in terms of $S_X(f)$, the PSD of $X(\cdot)$, by a parametric expression derived

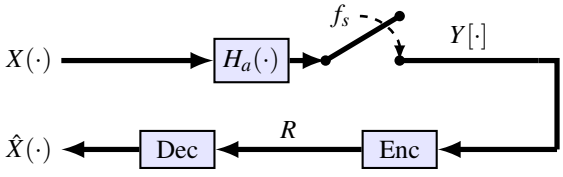


Fig. 1: Combined sampling and source coding model.

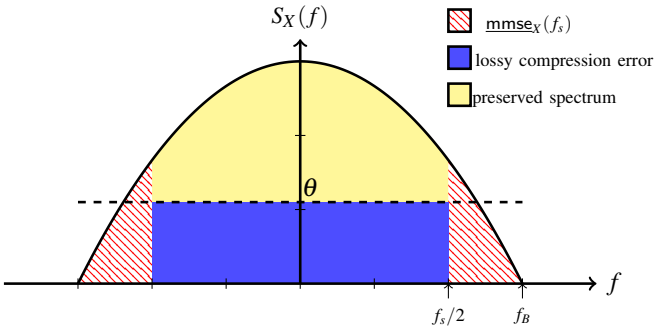


Fig. 2: Reverse waterfilling interpretation of (3): The function $D(f_s,R)$ of a unimodal $S_X(f)$ and zero noise is given by the sum of the sampling error and the lossy compression error.

by Pinsker [14] with a reverse waterfilling interpretation. A key idea in proving the source coding theorem which ties Pinsker's expression to the A/D conversion problem is to map $X(\cdot)$ into a discrete-time process based on sampling above its Nyquist frequency f_{Nyq} [1, Sec. 4.5.3]. The situation in which sampling at the Nyquist rate f_{Nyq} cannot be achieved due to system constraints [2] gives rise to the combined sampling and source coding problem depicted in Fig. 1 and solved in [4]. In this setting, $X(\cdot)$ is described by a rate R version of its sub-Nyquist samples $Y[\cdot]$. The minimal distortion in reconstruction taken over all such descriptions is denoted as $D(f_s,R)$. Under the assumption that $S_X(f)$ is unimodal and $H_a(f)$ is lowpass with cutoff freqeuncy $f_s/2$, $D(f_s,R)$ takes the following form [4, Eq. 9]

$$R(\theta) = \frac{1}{2}\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+\left[S_X(f)/\theta\right] df, \tag{1a}$$

$$D(\theta) = \underline{\mathrm{mmse}}_X(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min\left\{S_X(f),\theta\right\} df, \tag{1b}$$

where $\log^+(x) = \max\{0,\log(x)\}$ and

$$\underline{\mathrm{mmse}}_X(f_s) \triangleq \int_{\mathbb{R}\setminus\left(-\frac{f_s}{2},\frac{f_s}{2}\right)} S_X(f) df. \tag{2}$$

A waterfilling interpretation of (1) is illustrated in Fig. 2.

Assume that $X(\cdot)$ is band-limited to f_B. If $f_s > 2f_B$, then there is no loss of information in the sampling process, in which case we have

$$D(f_s,R) = D(R), \quad f_s \geq 2f_B, \tag{3}$$

where $D(R)$ is the (standard) quadratic DRF of the analog Gaussian source $X(\cdot)$, which is obtained by the celebrated reverse waterfilling expression of Pinsker [14]. In fact, it follows from [3] that if the energy of $X(\cdot)$ is not uniformly distributed over its bandwidth, then there exists a source coding rate R and a minimal sampling rate $f_{RD} < 2f_B$ such that (3) holds for all $f_s \geq f_{RD}$. This critical sampling rate can be computed by the equation

$$R = \frac{1}{2}\int_{-\frac{f_{RD}}{2}}^{\frac{f_{RD}}{2}} \log^+ \left[S_X(f)/S_X(f_{RD})\right] df. \tag{4}$$

It is shown in [3] that f_{RD} is monotonically increasing in R and approaches the Nyquist rate $2f_B$ as R goes to infinity. This result can be seen as an extension of the Shannon-Whittaker-Kotelnikov sampling theorem to the scenario where a finite bitrate constraint is imposed [15].

It also follows from [4] that a distortion arbitrarily close to $D(f_s,R)$ can be achieved by the following scheme:

(i) Filter-bank sampling at average frequency f_s with optimized pre-sampling filters and a sufficient number of sampling branches.
(ii) Vector quantizer with resolution of $Q = nR/f_s$ bits, where n is the block length.

B. Problem Formulation: Pulse Code Modulation

In this paper we study the distortion-rate performance of an A/D scheme which is similar to the one that achieves $D(f_s,R)$, where the vector quantizer is replaced by a scalar quantizer of resolution $q = R/f_s$ bits. This setting corresponds to the system model described in Fig. 3. The input process is an analog wide-sense stationary (WSS) process $X(\cdot) = \{X(t), t \in \mathbb{R}\}$ with PSD

$$S_X(f) \triangleq \int_{-\infty}^{\infty} \mathbb{E}\left[X(t+\tau)X(\tau)\right] e^{-2\pi i \tau f} d\tau.$$

The discrete-time process $Y[\cdot] = \{Y[n], n \in \mathbb{Z}\}$ is obtained by uniformly sampling the filtered process at frequency f_s, namely

$$Y[n] \triangleq \left(X(\cdot) \star h_a(\cdot)\right)(n/f_s), \quad n \in \mathbb{Z},$$

where $h_a(t)$ is the impulse response of the analog filter $H_a(f)$.

Let $\hat{Y}[n]$ be the process at the output of the quantizer at time n, and denote by $\eta[n]$ the quantization error, i.e.,

$$\hat{Y}[n] = Y[n] + \eta[n], \quad n \in \mathbb{Z}. \tag{5}$$

The variance of $\eta[n]$ is proportional to the size of the quantization bins, and decreases exponentially with the bit resolution q, provided the size of the bins decreases uniformly [16]. The non-linear relation between the quantizer input and its output complicates the analysis and usually calls for a simplifying assumption that linearizes the problem. A common assumption which we will adopt here is:

(A1) The process $\eta[\cdot]$ is i.i.d, uncorrelated with $Y[\cdot]$ and with variance

$$\sigma_\eta^2 = \frac{c_0}{(2^q-1)^2}. \tag{6}$$

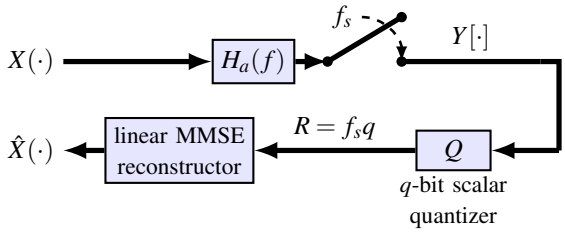


Fig. 3: System model of A/D with a scalar quantizer.

This assumption implies that the PSD of $\eta[\cdot]$ equals $S_\eta\left(e^{2\pi i\phi}\right) = \frac{c_0}{(2^q-1)^2}$ for any $\phi \in (-0.5, 0.5)$. The constant c_0 depends on statistical assumptions on the input signal. For example, if the amplitude of the input signal is bounded within the interval $(-A_m/2, A_m/2)$, then we can assume that the quantization bins are uniformly spaced and $c_0 = \frac{A_m}{12}$. If the input is Gaussian with variance σ^2 and the quantization rule is chosen according to the ideal point density allocation of the Lloyd algorithm [17], then [18, Eq. 10]

$$c_0 = \frac{\pi\sqrt{3}}{2}\sigma^2. \tag{7}$$

There exists a vast literature on the conditions under which assumption (A1) provides a good approximation to the system behavior. For example, in [16] it was shown that two consecutive samples $\eta[n]$ and $\eta[n+1]$ are approximately uncorrelated if the distribution of $Y[\cdot]$ is smooth enough, where this holds even if the sizes of the quantization bins are on the order of the variance of $Y[\cdot]$ [19]. This justifies the assumption that the process $\eta[\cdot]$ is white. Bennett [20] derived the following conditions under which $\eta[\cdot]$ and $Y[\cdot]$ are approximately uncorrelated: smooth PSD of $Y[\cdot]$, uniform quantization bins and a high quantizer resolution q. Since in our setting we are also interested in the low quantizer resolution regime, a better justification for this approximation is required. This will be the result of the following proposition, proof of which can be found in Appendix A.

Proposition 1. *The MMSE in estimating $X(\cdot)$ from $\hat{Y}[\cdot]$ in* (5) *is not smaller than the MMSE in estimating $X(\cdot)$ from the process*

$$\tilde{Y}[\cdot] \triangleq Y[n] + \tilde{\eta}[n], \quad n \in \mathbb{Z},$$

where $\tilde{\eta}[\cdot]$ is a stationary process possibly correlated with $Y[\cdot]$, with PSD $S_{\tilde{\eta}}\left(e^{2\pi i\phi}\right) = S_\eta\left(e^{2\pi i\phi}\right)$.

Proposition 1 implies that the assumption of an uncorrelated quantization noise and input signal can only increase the error, compared to an estimation scheme under the same marginal noise statistics that also takes into account the correlation between the samples and the quantization noise. We conclude that the analysis under assumption (A1) yields a good approximation to the true error if the quantizer resolution q is high, and provides an upper bound when q is low. The tightness of this upper bound can be learned from [21], where it was shown that PCM with a single bit

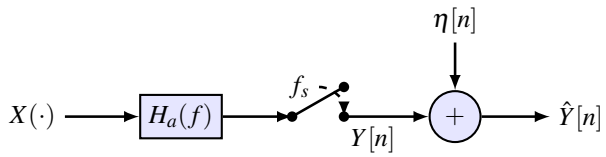


Fig. 4: Sampling and quantization system model.

quantizer leads to a reduction in the MSE of no more than 3db per octave more than an analysis that assumes (A1).

Under (A1), the relation between the input and the output of the system can be represented in the z domain by

$$\hat{Y}(z) = Y(z) + \eta(z). \tag{8}$$

This leads to the following relation between the corresponding PSDs:

$$\begin{aligned} S_{\hat{Y}}\left(e^{2\pi i\phi}\right) &= S_Y\left(e^{2\pi i\phi}\right) + S_\eta\left(e^{2\pi i\phi}\right) \\ &= f_s \sum_{k\in\mathbb{Z}} S_X(f - f_s k)\left|H_a(f - f_s k)\right|^2 + \sigma_\eta^2. \end{aligned} \tag{9}$$

A system that realizes the input-output realtion (8) is given in Fig. 4, where, in accordance with (A1), $\eta[\cdot]$ is white noise independent of $X(\cdot)$.

C. MMSE Estimation

We are interested in the MMSE in estimating $X(\cdot)$ from its noisy samples $\hat{Y}[\cdot]$. Namely, the minimal value of

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathbb{E}\left(X(t) - \hat{X}(t)\right)^2 \tag{10}$$

over all possible reconstruction methods of $X(\cdot)$ from $\hat{Y}[\cdot]$ of the form

$$\hat{X}(t) = \sum_{n\in\mathbb{Z}} w(t,n)\hat{Y}[n],$$

where $w(t,n)$ is square summable in n for every $t \in \mathbb{R}$. Standard linear estimation theory leads to the following proposition:

Proposition 2. *Consider the system in Fig. 4. The minimal time-averaged MSE* (10) *in linear estimation of $X(\cdot)$ from $\hat{Y}[\cdot]$ is given by*

$$\begin{aligned} &\mathsf{mmse} \triangleq \mathsf{mmse}_{X|\hat{Y}}(f_s, H_a) \\ &= \sigma_X^2 - \frac{1}{f_s}\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \frac{\sum_{k\in\mathbb{Z}} S_X^2(f - f_s k)\left|H_a(f - f_s k)\right|^2}{\sum_{k\in\mathbb{Z}} S_X(f - f_s k)\left|H_a(f - f_s k)\right|^2 + \sigma_\eta^2/f_s} df \end{aligned} \tag{11}$$

Proof: see Appendix B.

Note that in Proposition 2 we have not limited ourselves to band-limited input processes or to sub-Nyquist sampling. An expression for the optimal estimator $w^\star(t,n)$ can be derived from the proof. It can be shown to be of the form

$$w^\star(t,n) = w(t - n/f_s),$$

where the Fourier transform of $w(t)$ is

$$W(f) = \frac{H_a^*(f) S_X(f)}{\sum_{k\in\mathbb{Z}} \left|H_a(f)\right|^2 S_X(f - f_s k) + \sigma_\eta^2/f_s}.$$

The details are given in [15].

Using Hölder's inequality and monotonicity of the function $x \to \frac{x}{x+1}$, the integrand in (11) can be bounded for each f in the integration interval $(-f_s/2, f_s/2)$ by

$$\frac{(S^\star(f))^2}{S^\star(f) + \sigma_\eta^2/f_s}, \tag{12}$$

where

$$S^\star(f) = \sup_{k\in\mathbb{Z}} S_X(f - f_s k)\left|H_a(f - f_s k)\right|^2. \tag{13}$$

This leads to a lower bound on $\mathsf{mmse}_{X|\hat{Y}}(f_s, H_a)$. Under the assumption that $S_X(f)$ is unimodal in the sense that it is symmetric and non-increasing for $f > 0$, for each $f \in (-f_s/2, f_s/2)$ the supremum in (13) is obtained for $k = 0$. This implies that (12) is achievable if the pre-sampling filter is a low-pass filter with cut-off frequency $f_s/2$, namely

$$H_a^\star(f) = \begin{cases} 1, & |f| \le f_s/2, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

This choice of $H_a(f)$ in (11) leads to the following:

$$\mathsf{mmse}^\star_{X|Y}(f_s) = \underline{\mathsf{mmse}}_X(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \frac{S_X(f)}{1 + SNR(f)} df, \tag{15}$$

where $\underline{\mathsf{mmse}}_X(f_s)$ is defined in (2) and

$$SNR(f) \triangleq f_s S_X(f)/\sigma_\eta^2, \quad -\frac{f_s}{2} \le f \le \frac{f_s}{2}. \tag{16}$$

Henceforth, we will consider only processes with unimodal PSD, so that the MMSE under optimal pre-sampling filtering is given by (15). See [4] for the optimization of the expressions of the form (11) in the case where $S_X(f)$ is not unimodal.

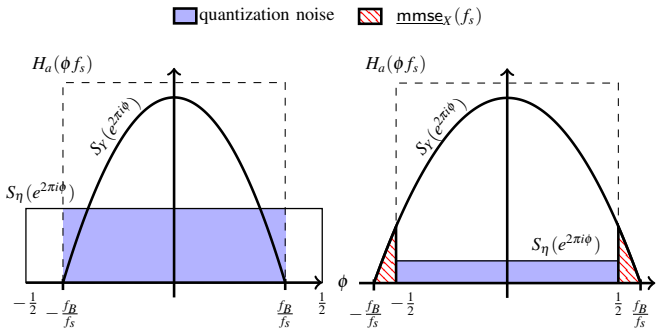
Since the SNR increases linearly in f_s, the MMSE of $X(\cdot)$ give $\hat{Y}[\cdot]$ decreases by a factor of $1/f_s$ for $f_s > 2f_B$ provided all other parameters are independent of f_s. In the next section we study (15) when, in addition, the quantizer resolution is inversely proportional to f_s, so as to keep a constant bitrate at the output as f_s varies.

III. Main Result: PCM under a Fixed Bitrate

In the PCM A/D conversion system of Fig. 3 with sampling frequency f_s and a quantizer resolution of q bits per sample, the amount of memory per time unit, or the bitrate at the output of the system, equals

$$R \triangleq q f_s$$

bits per time unit. Since in this model the A/D converter must use at least one bit per sample, we limit f_s to be smaller than the bitrate R. In this section we fix R and study the MMSE as a



(a) $f_s > 2f_B$ (b) $f_s < 2f_B$

Fig. 5: Spectral interpretation of Proposition 3: no sampling error when sampling above the Nyquist rate, but intensity of in-band quantization noise increases.

function of the sampling frequency f_s. Under this assumption, the variance of the quantization noise from (6) satisfies

$$\sigma_\eta^2 = \frac{c_0}{(2^q-1)^2} = \frac{c_0}{(2^{2R/f_s}-1)^2}. \tag{17}$$

The linear MMSE in estimating $X(\cdot)$ from $\hat{Y}$ under the optimal pre-sampling gives rise to an approximation to the operational distortion-function of the PCM, which we denote as $\tilde{D}(f_s,R)$. From (15) and (17) we obtain the following expression for $\tilde{D}(f_s,R)$:

Proposition 3. *The MMSE in estimating $X(\cdot)$ from $\hat{Y}[\cdot]$ assuming (A1) and $R = qf_s$ is as follows:*

$$\tilde{D}(f_s,R) = \underline{\mathsf{mmse}}_X(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \frac{S_X(f)}{1+SNR(f)}df \tag{18}$$

where

$$SNR(f) = SNR_{f_s,R}(f) = f_s\left(2^{R/f_s}-1\right)^2 \frac{S_X(f)}{c_0} \tag{19}$$

and $\underline{\mathsf{mmse}}_X(f_s)$ is given by (2)*.*

We will denote the two terms in the RHS of (18) as the *sampling error* and the *quantization error*, respectively. Fig. 6 shows the MMSE (18) as a function of f_s for a given R and various PSDs compared to their corresponding quadratic Gaussian iDRF under sub-Nyquist sampling (1). In Fig. 6 and in other figures throughout, we use the c_0 in (7) which corresponds to an optimal point density of the Gaussian distribution.

A. An Optimal Sampling Rate

The quantization error in (18) is an increasing function of f_s, whereas the sampling error $\underline{\mathsf{mmse}}_X(f_s)$ decreases in f_s. This situation is illustrated in Fig. 5. The sampling rate $f_s^\star$ that minimizes (18) is obtained at an equilibrium point where the derivatives of both terms are of equal magnitudes. Fig. 6 shows that $f_s^\star$ depends on the particular form of the input signal's PSD. If the signal is band-limited, we obtain the following result:

Proposition 4. *If $S_X(f) = 0$ for all $|f| > f_B$, then the sampling rate $f_s^\star$ that minimizes $\tilde{D}(f_s,R)$ is not bigger than $2f_B$.*

Proof: Note that $SNR_{f_s,R}(f)$ is an increasing function of f_s in the interval $0 \le f_s \le R$. Since we assume $X(\cdot)$ is band-limited, we have $\underline{\mathsf{mmse}}_X(f_s) = 0$ for $f_s \ge 2f_B$. This implies that $\tilde{D}(2f_B,R) \le \tilde{D}(f_s,R)$ for all $f_s > 2f_B$.

How much $f_s^\star$ is below $2f_B$ is determined by the derivative of $\underline{\mathsf{mmse}}_X(f_s)$, which equals $-2S_X(f_s/2)$. For example, in the case of the rectangular PSD:

$$\Pi(f) = \frac{\sigma^2}{2f_B}\begin{cases}1 & |f| \le f_B,\\ 0 & |f| > f_B,\end{cases} \tag{20}$$

the derivative of $-2S_X(f_s/2)$ for $f_s < 2f_B$ is $-\sigma^2$. The derivative of the second term in (18) is smaller than σ^2 for most choices of system parameters[1]. It follows that 0 is in the sub-gradient of $\tilde{D}(f_s,R)$ at $f_s = 2f_B$, and thus $f_s^\star = 2f_B$, i.e., Nyquist rate sampling is optimal when the energy of the signal is uniformly distributed over its bandwidth. Two more input signal examples are given below.

Example 1 (triangular PSD)**.** *Consider an input signal PSD*

$$\Lambda(f) = \frac{\sigma^2}{f_B}\max\left\{1-\frac{f}{f_B},0\right\}. \tag{21}$$

For any $f_s \le 2f_B$, we have

$$\underline{\mathsf{mmse}}_X(f_s) = \sigma^2 - \frac{\sigma^2}{f_B}\left(f_s - \frac{f_s^2}{4fB}\right).$$

Since the derivative of $\underline{\mathsf{mmse}}_X(f_s)$, which is $-2\Lambda(f_s/2)$, changes continuously from 0 to $-2\sigma^2/f_B$ as f_s varies from $2f_B$ to 0, we have $0 < f_s^\star < 2f_B$. The exact value of $f_s^\star$ depends on R and the ratio σ^2/c_0. It converges to $2f_B$ as the value of any of these two increases.

Example 2 (PSD of unbounded support)**.** *Consider an input signal PSD of the form*

$$S_G(f) = \frac{\sigma^2}{\sqrt{2\pi}f_0}e^{\frac{(f/f_0)^2}{2}}, \quad f \in \mathbb{R}, \tag{22}$$

where $f_0 > 0$. For a process with the PSD (22) *there exists a non-zero sampling error $\underline{\mathsf{mmse}}_X(f_s)$ for any finite sampling rate f_s, and therefore the argument in Proposition 4 does not hold.*

We can compare $f_s^\star$ in each of the examples above to the minimal sampling rate f_{DR} that achieves the quadratic distortion-rate function of a Gaussian process with the same PSD, given by (4). In the case of the PSD (21), the relation between f_{DR} and R was derived in [3]:

$$R = \frac{\sigma^2}{\ln 2}\left(\log\frac{1}{1-\frac{f_{DR}}{2f_B}} - \frac{f_{DR}}{2f_B}\right). \tag{23}$$

[1]This holds if $1 > \frac{c_0}{\sigma^2}\left(2^{0.5R/f_B}-1\right)^{-2}$.

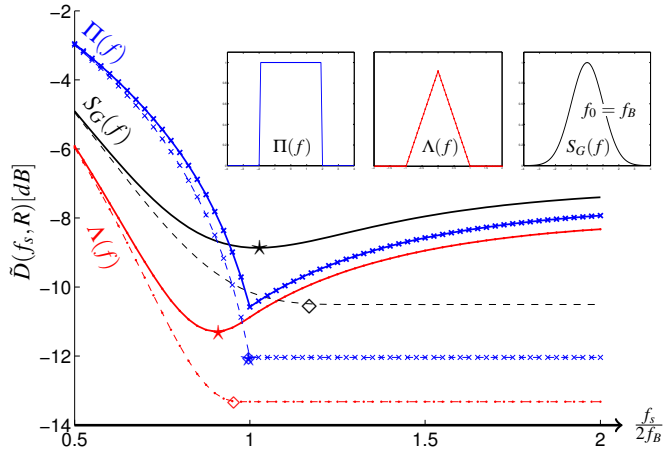


Fig. 6: MMSE as a function of f_s for a fixed R and various PSDs, which are given in the small frames. The dashed curves are the corresponding iDRF $D(f_s,R)$ given by (1). The rates $f_s^\star$ and f_{DR} corresponds to the $\star$ and $\diamond$, respectively.

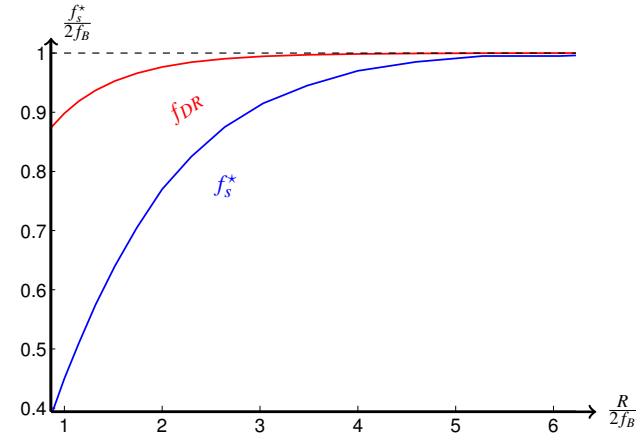


Fig. 7: Optimal sampling rates $f_s^\star$ and f_{DR} versus [3] R for the process with PSD $\Lambda(f)$ of (21).

We plot $f_s^\star$ and the corresponding f_{DR} in Fig. 7, as a function of R. It can be seen that $f_s^\star$ is smaller than f_{DR}, where both approach $2f_B$ as R increases. In the case of the PSD (22), the relation between f_{DR} and R can be computed from (4). This is plotted together with $f_s^\star$ versus R in Fig. 8. Note that since $S_G(f)$ is not band-limited, f_{DR} is not bounded in R since there is no sampling rate that guarantees perfect reconstruction for this signal.

B. Discussion

Under a fixed bitrate constraint, oversampling no longer reduces the MMSE since increasing the sampling rate reduces the quantizer resolution and increases the magnitude of the quantization noise. As illustrated in Fig. 5, for any f_s below the Nyquist rate the bandwidth of both the signal and the noise occupies the entire digital frequency domain, whereas the magnitude of the noise decreases as more bits are used in quantizing each sample.

It follows that $f_s^\star$ cannot be larger than the Nyquist rate as stated in Proposition 4, and is strictly smaller than Nyquist when the energy of $X(\cdot)$ is not uniformly distributed over its bandwidth, as in Example 1. In this case, some distortion due to sampling is preferred in order to increase the quantizer resolution. In other words, restricted to scalar quantization, the optimal rate R code is achieved by sub-sampling. This behavior of $\tilde{D}(f_s,R)$ is similar to the behavior of the information theoretic bound $D(f_s,R)$, as both provide an optimal sampling rate which balances sampling error and lossy compression error. On the other hand, oversampling introduces redundancy into the PCM representation, and yields a worse distortion-rate code than with $f_s = f_s^\star$. In this aspect the behavior of $\tilde{D}(f_s,R)$ is different than $D(f_s,R)$, since the latter does not penalize oversampling[2].

The trade-off between sampling rate and quantization precision is particularly interesting in the case where the signal is not band-limited: Although there is no sampling rate that guarantees perfect reconstruction, there is still a sampling rate that optimizes the aforementioned trade-off and minimizes the MMSE under a bitrate constraint.

The similarity between $f_s^\star$ and f_{DR} as a function of R suggests that in order to implement a sub-Nyquist A/D converter that operates close to the minimal information theoretic sampling rate f_{DR}, the principle of trading quantization bits with sampling rate must be taken into account. The observation that

$$f_s^\star \leq f_{DR} \tag{24}$$

in Examples 1 and 2 raises the conjecture as to whether (24) holds in general. This may be explained by the diminishing effect of reducing the sampling rate on the overall error. In other words, the fact that $\tilde{D}(f_s^\star,R) \geq D(R)$ implies that a distortion-rate achievable scheme is more sensitive to changes in the sampling rate than the sub-optimal implementation of A/D conversion via PCM. The dependency of $f_s^\star$ in the spectral energy distribution $S_X(f)$ has a time-domain explanation: for a fixed variance σ_X^2, two consecutive time samples taken at the Nyquist rate are more correlated (in their absolute value) when the PSD is not flat. Consequently, more redundancy is present after sampling than in the case where the PSD is flat. The main discovery of this paper is that part of this redundancy can be removed simply by sub-sampling, where this is in fact the optimal way to remove it when we are restricted to the PCM setting of Fig. 3.

IV. Conclusions

A/D conversion via pulse-code modulation under a fixed bitrate at the output introduces a trade-off between the

[2]This is because in the system model of Fig. 1, the encoder has the freedom to discard redundant information.

[3]The curves do not go further left since in our model we restrict the sampling rate to be smaller than the output bitrate R.

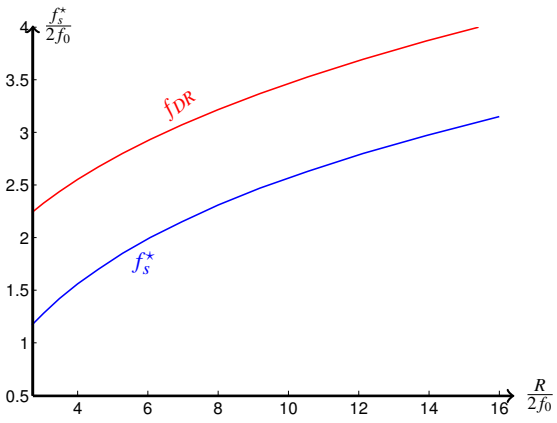


Fig. 8: Optimal sampling rate $f_s^\star$ and f_{DR} versus[3] R for the process with PSD (22).

sampling rate and the number of bits we use to quantize each sample. The optimal sampling rate that minimizes the MMSE obtained as a result of this trade-off is lower than the Nyquist rate. That is, our analysis shows that to minimize MMSE between the A/D input and output, some sampling distortion is preferred in order to increase the quantizer resolution.

ACKNOWLEDGMENT

This work was supported in part by the NSF Center for Science of Information (CSoI) under grant CCF-0939370, the BSF Transformative Science Grant 2010505.

REFERENCES

[1] T. Berger, *Rate-Distortion Theory*. Wiley Online Library, 1971.
[2] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, 2009.
[3] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "Sub-Nyquist sampling achieves optimal rate-distortion," in *Information Theory Workshop (ITW), 2015 IEEE*, Apr 2015.
[4] A. Kipnis, A. J. Goldsmith, Y. C. Eldar, and T. Weissman, "Distortion-rate function of sub-Nyquist sampled Gaussian sources," 2014, to appear in IEEE Transactions on Information Theory. [Online]. Available: http://arxiv.org/abs/1405.5329
[5] D. L. Neuhoff and S. S. Pradhan, "Information rates of densely sampled data: Distributed vector quantization and scalar quantization with transforms for Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5641–5664, 2013.
[6] B. Oliver, J. Pierce, and C. Shannon, "The philosophy of PCM," *Proc. IRE*, vol. 36, no. 11, pp. 1324–1331, Nov 1948.
[7] R. Gray, "Quantization noise spectra," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1220–1244, Nov 1990.
[8] N. Thao and M. Vetterli, "Reduction of the mse in r-times oversampled A/D conversion O(1/R) to O(1/R$\hat{2}$)," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 200–203, Jan 1994.
[9] J. de la Rosa, "Sigma-delta modulators: Tutorial overview, design guide, and state-of-the-art survey," *IEEE Trans. Circuits Syst.*, vol. 58, no. 1, pp. 1–21, Jan 2011.
[10] N. Thao and M. Vetterli, "Lower bound on the mean squared error in multi-loop Sigma Delta modulation with periodic bandlimited signals," in *The Twenty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol. 2, Oct 1994, pp. 1536–1540 vol.2.
[11] Z. Cvetkovic and M. Vetterli, "On simple oversampled A/D conversion in l2(r)," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 146–154, Jan 2001.
[12] C. S. Güntürk, "One-bit sigma-delta quantization with exponential accuracy," *Communications on Pure and Applied Mathematics*, vol. 56, no. 11, pp. 1608–1630, 2003.
[13] G. Hattab and M. Ibnkahla, "Multiband spectrum access: Great promises for future cognitive radio networks," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 282–306, March 2014.
[14] A. Kolmogorov, "On the shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inform. Theory*, vol. 2, no. 4, pp. 102–108, December 1956.
[15] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "Sampling stationary processes under bitrate constraints," in preparation.
[16] H. Viswanathan and R. Zamir, "On the whiteness of high-resolution quantization errors," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2029–2038, Jul 2001.
[17] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar 1982.
[18] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
[19] B. Widrow, "A study of rough amplitude quantization by means of nyquist sampling theory," *Circuit Theory, IRE Transactions on*, vol. 3, no. 4, pp. 266–276, Dec 1956.
[20] W. R. Bennett, "Spectra of quantized signals," *Bell System Technical Journal*, vol. 27, no. 3, pp. 446–472, 1948.
[21] N. Thao and M. Vetterli, "Lower bound on the mean-squared error in oversampled quantization of periodic signals using vector quantization analysis," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 469–479, Mar 1996.

APPENDIX A

Proof of Proposition 1:

Let $X[\cdot]$ and $Z[\cdot]$ be two jointly stationary processes and let

$$Y[n] = X[n] + Z[n], \quad n \in \mathbb{Z}.$$

The MMSE under linear estimation of $X[\cdot]$ from $Y[\cdot]$ is given by

$$E_{corr} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{S_X\left(e^{2\pi i\phi}\right) S_Z\left(e^{2\pi i\phi}\right) - \left|S_{XZ}\left(e^{2\pi i\phi}\right)\right|^2}{S_X\left(e^{2\pi i\phi}\right) + S_Z\left(e^{2\pi i\phi}\right) + 2\Re S_{XZ}\left(e^{2\pi i\phi}\right)} d\phi. \quad (25)$$

We will show that E_{corr} cannot exceeds the MMSE when the correlation between $X(\cdot)$ and $Z(\cdot)$ is zero. Let

$$S_{XZ}\left(e^{2\pi i\phi}\right) = \Re S_{XZ}\left(e^{2\pi i\phi}\right) + i\Im S_{XZ}\left(e^{2\pi i\phi}\right) =: u + iv,$$

where $u, v \in \mathbb{R}$. The integrand in (25) can be written as

$$\frac{S_X\left(e^{2\pi i\phi}\right) S_\eta\left(e^{2\pi i\phi}\right) - u^2 - v^2}{S_X\left(e^{2\pi i\phi}\right) + S_Z\left(e^{2\pi i\phi}\right) + 2u}. \quad (26)$$

Since $S_X\left(e^{2\pi i\phi}\right) S_\eta\left(e^{2\pi i\phi}\right) \geq \left|S_{XZ}\left(e^{2\pi i\phi}\right)\right|^2$ we have

$$u^2 + v^2 \leq S_X\left(e^{2\pi i\phi}\right) S_\eta\left(e^{2\pi i\phi}\right). \quad (27)$$

Note that (26) is positive and maximizing it is equivalent to maximizing (25). Since (26) is convex in u and v, it obtains its maximum over the boundary defined by (27). Specifically, the maximum of (26) in the domain (27) is obtain at $u = v = 0$. This implies that

$$E_{corr} \leq \frac{S_X\left(e^{2\pi i\phi}\right) S_Z\left(e^{2\pi i\phi}\right)}{S_X\left(e^{2\pi i\phi}\right) + S_Z\left(e^{2\pi i\phi}\right)}, \quad (28)$$

and the RHS of (28) is the expression for the MMSE under linear estimation when $S_{XZ}\left(e^{2\pi i\phi}\right) \equiv 0$, i.e., when $Z(\cdot)$ is uncorrelated with $X(\cdot)$.

APPENDIX B

In this Appendix we provide the proof of Proposition 2.

For $0 \leq \Delta \leq 1$ define

$$X_\Delta[n] \triangleq X\left((n+\Delta)T_s\right), \quad n \in \mathbb{Z},$$

where $T_s \triangleq f_s^{-1}$. Also define $\hat{X}_\Delta[n]$ to be the optimal MSE estimator of $X_\Delta[n]$ from $\hat{Y}[\cdot]$, that is

$$\hat{X}_\Delta[n] = \mathbb{E}\left[X_\Delta[n]|\hat{Y}[\cdot]\right], \quad n \in \mathbb{Z}.$$

The MSE in (10) can be written as

$$\begin{aligned}
&\mathsf{mmse}_{X|\hat{Y}} = \lim_{N\to\infty} \frac{1}{2N+1} \int_{-N}^{N+1} \mathbb{E}\left(X(t) - \hat{X}(t)\right)^2 dt \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \int_0^1 \mathbb{E}\left(X\left((n+\Delta)T_s\right) - \hat{X}\left((n+\Delta)T_s\right)\right)^2 d\Delta \\
&= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \int_0^1 \mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 d\Delta \\
&= \int_0^1 \mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 \Delta. \qquad (29)
\end{aligned}$$

Note that $S_{X_\Delta}\left(e^{2\pi i\phi}\right) = S_Y\left(e^{2\pi i\phi}\right)$ and $X_\Delta[\cdot]$ and $\hat{Y}[\cdot]$ are jointly stationary with cross-PSD

$$S_{X_\Delta\hat{Y}}\left(e^{2\pi i\phi}\right) = S_{X_\Delta}\left(e^{2\pi i\phi}\right) = f_s \sum_{k\in\mathbb{Z}} S_X\left(f_s(k-\phi)\right) e^{2\pi i\Delta(k-\phi)}.$$

Denote by $S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right)$ the PSD of the estimator obtained by the discrete Wiener filter for estimating $X_\Delta[\cdot]$ from $\hat{Y}[\cdot]$. We have

$$\begin{aligned}
S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) &= \frac{S_{X_\Delta\hat{Y}}\left(e^{2\pi i\phi}\right) S^*_{X_\Delta\hat{Y}}\left(e^{2\pi i\phi}\right)}{S_{\hat{Y}}\left(e^{2\pi i\phi}\right)} \\
&= \sum_{n,k} \frac{f_s^2 S_{X_a}\left(f_s(k-\phi)\right) S_{X_a}\left(f_s(n-\phi)\right) e^{2\pi i\Delta(k-n)}}{S_Y\left(e^{2\pi i\phi}\right) + S_\eta\left(e^{2\pi i\phi}\right)} \qquad (30)
\end{aligned}$$

Where $S_{X_a}(f) = S_X(f)\left|H_a(f)\right|^2$ is the PSD of the process at the output of the analog filter. The estimation error in Wiener filtering is given by

$$\begin{aligned}
&\mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 \\
&\quad = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_\Delta}\left(e^{2\pi i\phi}\right) d\phi - \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) dphi \\
&\quad = \sigma_X^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) d\phi. \qquad (31)
\end{aligned}$$

Equations (29), (30) and (31) leads to

$$\begin{aligned}
\mathsf{mmse}_{X|\hat{Y}} &= \int_0^1 \mathbb{E}\left(X_\Delta[n] - \hat{X}_\Delta[n]\right)^2 \Delta \\
&= \sigma_X^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^1 S_{X_\Delta|\hat{Y}}\left(e^{2\pi i\phi}\right) d\phi \\
&\stackrel{a}{=} \sigma_X^2 - \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{f_s \sum_{k\in\mathbb{Z}} S^2_{X_a}\left(f_s(k-\phi)\right)}{S_Y\left(e^{2\pi i\phi}\right) + S_\eta\left(e^{2\pi i\phi}\right)} d\phi, \qquad (32)
\end{aligned}$$

where (a) follows from (30) and the orthogonality of the functions $\left\{e^{2\pi xk}, k \in \mathbb{Z}\right\}$ over $0 \leq x \leq 1$. Equation (11) is obtained from (32) by changing the integration variable from ϕ to $f = \phi f_s$.