Lower bounds for the error decay incurred by coarse quantization schemes

Felix Krahmer * Rachel Ward †

June 27, 2011

Abstract

Several analog-to-digital conversion methods for bandlimited signals used in applications, such as $\Sigma\Delta$ quantization schemes, employ coarse quantization coupled with oversampling. The standard mathematical model for the error accrued from such methods measures the performance of a given scheme by the rate at which the associated reconstruction error decays as a function of the oversampling ratio λ. It was recently shown that exponential accuracy of the form $O(2^{-\alpha\lambda})$ can be achieved by appropriate one-bit Sigma-Delta modulation schemes. However, the best known achievable rate constants α in this setting differ significantly from the general information theoretic lower bound. In this paper, we provide the first lower bound specific to coarse quantization, thus narrowing the gap between existing upper and lower bounds. In particular, our results imply a quantitative correspondence between the maximal signal amplitude and the best possible error decay rate. Our method draws from the theory of large deviations.

1 Introduction

Many signals of practical engineering interest are naturally produced in analog form; at the same time, it is becoming more efficient and robust to store and transmit signals in digital form. Therefore, the study of accurate and

*Hausdorff Center for Mathematics, Universität Bonn, Bonn, Germany
†Courant Institute of Mathematical Science, New York University, New York, NY, USA

tractable methods for analog-to-digital (A/D) conversion, or the approximation of real-valued signals using a finite alphabet, is of great importance in modern signal processing.

In the setting of A/D conversion, the signal of interest $x(t)$ is often modeled as a bounded bandlimited function. According to the well-known Shannon-Nyquist sampling theorem, such functions are completely determined by their values $x_n = x(\frac{n}{\lambda})$ sampled at frequency λ greater than the signal bandwidth. The original signal can be reconstructed from these samples using convolutional decoding of the form $x(t) = \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} x_n g(t - \frac{n}{\lambda})$. Exact equality can be obtained by choosing the function g so that its Fourier transform has compact support, and approximates the characteristic function of the frequency support of $x(t)$. In that case, the reconstruction formula represents an ideal low-pass filter. Conversion between analog and digital representations for $x(t)$ may be achieved by replacing the input sequence (x_n) by a sequence (q_n) of *quantized values* chosen from a finite set such that the signal

$$\widetilde{x}(t) = \sum_{n\in\mathbb{Z}} q_n g\big(t - \frac{n}{\lambda}\big) \tag{1}$$

formed by replacing the x_n's with the q_n's yields a good approximation of x. In applications, one is often forced to approximate the ideal low-pass filter g by a filter φ satisfying additional constraints, as for example compact (time) support.

In addition, one sometimes restricts attention to recovering the values x_j on the sampling grid only. Consequently, such a quantization scheme fixes a finite-length reconstruction filter φ_n, and approximate recovery is then obtained if

$$x_j \approx \tilde{x}_j = \sum \varphi_{j-n} q_n. \tag{2}$$

In this paper we will focus on the continuous scenario (1), but we will allow for (almost) arbitrary reconstruction kernels φ. Similar techniques extend to corresponding results for the discrete scenario (2).

Quantization schemes employed in practice

In *pulse code modulation*, the sampling frequency λ is close to the critical sampling frequency, and the quantized value q_n is taken to be a truncated binary representation of the sample x_n. To increase the accuracy of this

approximation, one takes longer binary expansions of each sample. In particular, if m bits are allotted to each truncated binary expansion, then the distortion $\|x - \widetilde{x}\|_{L^\infty}$ decreases like $O(2^{-m})$.

On the other hand, the set of admissible values for q_n in *oversampled coarse quantization* methods is restricted to a fixed alphabet $\mathcal{A}$ of reasonably small size, and more accurate approximations are obtained by increasing the sampling rate λ. In the extreme case of one-bit quantization, one chooses the alphabet $\mathcal{A}_1 = \{-1, +1\}$. For K-bit quantization, the q_n are taken from the set $\mathcal{A}_K$ consisting of 2^K evenly spaced values in the closed interval $[-1, 1]$. The number of bits spent per unit time interval in this setting is $m = \lambda \log_2 |\mathcal{A}_\mathcal{K}| = \lambda K$. From the viewpoint of circuit engineering, oversampled coarse quantization is associated to low-cost analog hardware, because increasing the sampling rate is cheaper than refining the quantization. Consequently, oversampling data converters are often used for low to medium-bandwidth signals, such as audio signals [10] and, more recently, for wireless communication [5]. Further advantages of oversampled coarse quantization methods include a built-in redundancy and robustness against errors resulting from imperfections in the analog circuit implementation. This robustness comes as a consequence of the more 'democratic' distribution of bit significance in the reconstruction formula, see [1]; in the extreme case of one-bit quantization, the individual bits $q_n \in \{-1, 1\}$ carry *equal* significance.

Our work in relation to prior advances

In this paper, we show that these advantages of coarse quantization come with the price of sub-optimal accuracy of the resulting convolutional approximation. It is well-known (see, for example, [8], [7]) that no quantization scheme spending m bits per Nyquist interval can beat the error decay of $O(2^{-m})$ achieved by pulse code modulation. This optimal rate of decay is not possible for coarse quantization in the discrete setting (2), shown in the work of Calderbank and Daubechies [1]. Until now, tighter lower bounds for coarse quantization are available only under the white noise hypothesis, where one assumes that the quantization error $x_n - q_n$ is distributed like Gaussian white noise, and in conjunction with additional technical assumptions [3]. In contrast, the lower bounds we shall provide hold for *any* K-bit quantization scheme, without any additional assumptions, and independent of the encoding algorithm used to generate the q_n.

As the main contribution of this paper, we provide an explicit lower bound

on the error decay achievable by K-bit quantization. Normalizing such that the q_n's are chosen from an evenly spaced alphabet with endpoints -1 and 1, and such that the bandlimited functions of interest are bounded in amplitude by $\mu < 1$, we will show that the rate of decay of $\|\widetilde{x}_\lambda - x\|_\infty$ is bounded below by $O\big(2^{-\alpha m}\big)$, where $\alpha = 1 - K^{-1}(1 - h(\frac{1+\mu}{2}))$, and h is the *unbiased binary entropy function* $h(u) = -\big((1-u)\log_2(1-u) + u\log_2 u\big)$. In fact, the best known upper bounds for K-bit quantization are also of the form $O(2^{-rm})$. Such a bound was first achieved via a construction by Güntürk [7] of a family of one-bit $\Sigma\Delta$ quantization schemes. These constructions were later refined by Deift, Güntürk, Krahmer [2], yielding the rate constant $r \approx 0.102$. As this rate constant is achieved only over input signals of maximal amplitude $\mu \leq .05$, this upper bound does not stand in contradiction to our lower bound which gives $r > \alpha \geq .9982$ when $\mu = .05$ and $K = 1$. On the other hand, our lower bound implies that the best possible rate constant tends to zero as $\mu \to 1$.

Organization of the paper

After precisely setting up the problem and clarifying our notation in Section 2, we summarize our results in Section 3. In Section 4, we recall important concepts and results from the theory of large deviations. In that section, we also recall results from the theory of Banach spaces which we use in the proof of our main theorem, which is presented in Section 5.

2 Notation and setup

Before continuing, let us introduce the notation used in this paper. We use the Landau O-notation $f(x) = O(h(x))$ (and $f(x) = o(h(x))$) to imply that for some $M > 0$ (or any $M > 0$, respectively), there exists a real number u_0 such that $|f(u)| \leq M|h(u)|$ for all $u \geq u_0$. Let $\mathcal{S}$ denote the Schwartz space of rapidly decreasing functions on $\mathbb{R}$. For the Fourier transform, we use the normalization

$$\widehat{x}(\omega) := \int_{-\infty}^{\infty} x(t)\exp\left(-2\pi i\omega t\right)dt. \tag{3}$$

We define the class $\mathcal{B}_\Omega(\mathbb{R})$ of Ω-bandlimited functions to be the space of real-valued continuous functions in $L^\infty(\mathbb{R})$ whose Fourier transforms (in the distributional sense) have support contained in $[-\Omega/2, \Omega/2]$. Henceforth,

we will normalize $\Omega = 1$. The classical sampling theorem for bandlimited functions states that if $\lambda > 1$, then any function x in the class $\mathcal{B}_1(\mathbb{R})$ and having bounded amplitude can be recovered from its samples $\{x(\frac{n}{\lambda})\}_{n\in\mathbb{Z}}$ as a weighted sum of translates of an averaging kernel $g \in L^1(\mathbb{R})$ via the formula

$$x(t) = \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} x\Big(\frac{n}{\lambda}\Big)g\Big(t - \frac{n}{\lambda}\Big), \tag{4}$$

where g is any kernel whose Fourier transform satisfies

$$\widehat{g}(\omega) \;\; = \;\; \begin{cases} 1, & \text{if } |\omega| \le \pi \\ 0, & \text{if } |\omega| \ge \lambda_0\pi \end{cases} \tag{5}$$

for some arbitrary λ_0 with $\lambda \ge \lambda_0 > 1$. Note that with such a g, the reconstruction formula (4) describes an ideal low-pass filter. Note also that any such kernel g with finite frequency support must have infinite (time) support, according to the uncertainty principle. Such ideal filters with infinite-support are cumbersome to construct, and in practice are often approximated by kernels having finite support. In this case, the reconstruction formula (4) holds at most approximately. A priori it is not clear that this approximation always has a negative effect on the accuracy of the associated quantization schemes. For this reason, in the subsequent analysis we will not restrict the choice of the filter by more than a simple smoothness condition. We will use, however, the normalization arising naturally in the ideal case. There one has by (5) that

$$\int g(t)dt = \widehat{g}(0) = 1; \tag{6}$$

we adapt this normalization for general kernels φ.

A *K-bit quantization scheme* assigns, to each input function x and to each sampling rate $\lambda \ge \lambda_0$, a sequence of evenly-spaced q_n^λ from an alphabet $\mathcal{A}_K$ of size $|\mathcal{A}_K| = 2^K$ in such a way that the approximation

$$\widetilde{x}_\lambda(t) = \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} q_n^\lambda \varphi\Big(t - \frac{n}{\lambda}\Big) \tag{7}$$

approaches $x(t)$ as $\lambda \to \infty$. Consequently, the approximation quality resulting from a particular sequence $\{q_n^\lambda\}_{n\in\mathbb{Z}}$ of quantized values together with a reconstruction kernel φ is commonly assessed by the reconstruction error,

$$e^K_{\lambda,q^\lambda}(t) := x(t) - \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} q_n^\lambda \varphi(t - \frac{n}{\lambda}), \qquad q_n^\lambda \in \mathcal{A}_K \tag{8}$$

and its supremum norm. We shall normalize the K-bit quantization alphabet $\mathcal{A}_K$ so as to have extreme values $+1$ and -1.

With this normalization on the alphabet in place and the kernel normalization (6), the approximate reconstruction in (7) is essentially a weighted local average of the q_n^λ. Hence we can not expect good approximation for $\|x\|_{L^\infty} > 1$. For this reason, we fix $\mu < 1$ and work with the space $\mathcal{B}_1^\mu(\mathbb{R}, \mu)$ defined to be the class of functions in $\mathcal{B}_1(\mathbb{R})$ with amplitude bounded by μ on the whole real line. Thus we will study

$$E_K^\mu(\lambda) := \sup_{x \in \mathcal{B}_1(\mathbb{R},\mu)} \inf_{q_n^\lambda \in \mathcal{A}_K} \|e_\lambda^K\|_{L^\infty}. \tag{9}$$

3 Summary of results

Our main result concerns a lower bound on the rate of decay for K-bit quantization of bandlimited functions in terms of the maximal amplitude μ:

Theorem 3.1. *Consider a K-bit quantization scheme associated to a reconstruction kernel $\varphi \in \mathcal{S}$, normalized so that $\int \varphi(t)dt = 1$. If the optimal rate of decay for such a scheme satisfies $E_K^\mu(\lambda) = O(2^{-\alpha K \lambda})$, then*

$$\alpha \leq 1 - K^{-1}\left(1 - h\left(\frac{1+\mu}{2}\right)\right),$$

where $h(p) = -\big(p \log_2 p + (1-p)\log_2(1-p)\big)$ is the binary entropy function.

Theorem 3.1 represents a quantitative improvement over the general lower bound, which for K-bit quantization reads

$$E_K^\mu \geq O(2^{-K\lambda}), \tag{10}$$

as well as over the corresponding strict inequality in the discrete case (as mentioned above).

The lower bound provided in Theorem 3.1 is most markedly improved over the previous lower bound (10) in the case of one-bit quantization, $K = 1$. In this case, the bound reduces to $\alpha \leq h(\frac{1+\mu}{2})$. In Figure 3, we compare our lower bound with the best-known upper bounds from [2] in this setting. Observe that in the limit as $\mu \to 1$, the upper and lower bounds both yield $\alpha = 0$. For small μ, however, there is a considerable gap between the lower

bounds provided in this paper and the best-known constructive upper bounds in [2]. A possible explanation for that fact is that our lower bounds hold for arbitrary bit sequences, while there need not be a constructive procedure to find the optimal bit sequence from a signal.

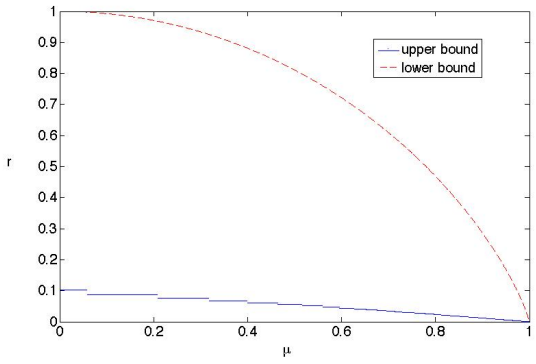

Figure 1: The rate constants α corresponding to upper and lower bounds for the error decay as a function of μ and for 1-bit quantization.

Intuition behind Theorem 3.1

That the performance of K-bit quantization schemes should depend on the maximal amplitude μ can be understood as follows. Among the 2^{KN} sequences of length N comprised of elements $q_n \in \mathcal{A}_K$, most of the sums $\sum_{n=0}^{N} q_n$ will have an average near zero. Now the values of the reconstructed function $\widetilde{x} = \sum_{|n|\leq N} q_n \varphi(t - \frac{n}{\lambda})$ are computed as a local average of the q_n's, hence most of the possible $\widetilde{x}$ are localized near zero as well. The larger μ, the larger the function values to be represented; the disproportion increases.

Positive time sampling

We note that in practice, the input signal $x(t)$ is accessible only for positive time $t \geq 0$, so one needs to reconstruct it from positive-time samples $x(\frac{n}{\lambda}), n \in \mathbb{N}$ only. That is, it is more realistic to consider approximations of the form

$$\widetilde{x}_{\lambda}^{+}(t) = \frac{1}{\lambda} \sum_{n \in \mathbb{N}} q_n^{\lambda} \varphi(t - \frac{n}{\lambda}). \tag{11}$$

Accordingly, one may set a calibration time T_0 over which the approximation need not hold, and measure the reconstruction error through

$$E_K^{\mu,+}(\lambda) := \sup_{x \in \mathcal{B}_\infty(\mathbb{R},\mu)} \sup_{t \geq T_0} \left| x(t) - \frac{1}{\lambda} \sum_{n \in N} q_n^\lambda \varphi(t - \frac{n}{\lambda}) \right|. \tag{12}$$

If the calibration time $T_0 = T_0(\lambda)$ is sufficiently long as a function of λ, then the effect of using only positive-time samples can be controlled. Following the lines of [7], one obtains the following corollary to Theorem 3.1:

Corollary 3.2. *If one sets $T_0 = T_0(\lambda)$ as in (20) below, and if the optimal rate of decay for $E_K^{\mu,+}(\lambda)$ satisfies $E_K^{\mu,+}(\lambda) = O(2^{-\alpha K \lambda})$, then*

$$\alpha \leq 1 - K^{-1}\Big(1 - h\Big(\frac{1+\mu}{2}\Big)\Big). \tag{13}$$

4 Background

4.1 Inequalities from the theory of large deviations

In order to make the intuition behind Theorem 3.1 rigorous, we need some results from the theory of large deviations for Bernoulli random variables. Recall that a Bernoulli random variable X with bias p takes values in the set $\{0,1\}$ with $\mathbb{P}(X=1) = p$. The *relative entropy* between two Bernoulli distributions with associated biases p and a is given by $H = H(a,p) := -\Big(a \log_2\left(\frac{a}{p}\right) + (1-a)\log_2\left(\frac{1-a}{1-p}\right)\Big)$. In the particular case $p = 1/2$, the relative entropy function $H(a,1/2)$ simplifies to $H(a,1/2) = h(a) - 1$ where $h(a) = -\left(a\log_2(a) + (1-a)\log_2(1-a)\right)$ is the binary entropy function. For a sequence of independent Bernoulli random variables B_j with bias p, denote by $S_n := \sum_{j=1}^n B_j$ the sequence of their partial sums. A basic result in the theory of large deviations for Bernoulli sums reads

Proposition 4.1. *For $p < a < 1$, and for $n \in \mathbb{N}$, one has*

$$P_n(a) := P\left(S_n \geq na\right) \leq 2^{-nH}. \tag{14}$$

Among any sum of independent and identically distributed (i.i.d) random variables X_j supported on $[0,1]$ and with expected value $\mathbb{E}X_j = p$, the Bernoulli sum presents the slowest exponential rate of convergence towards zero for the probabilities of large deviation:

Proposition 4.2. *Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables on $[0,1]$ with $\mu = \mathbb{E}[X_n] = p$. Then for $p < a < 1$, and for $n \in \mathbb{N}$, one has*

$$P\left(\sum_{j=1}^{n} X_j \geq na\right) \leq P_n(a) \leq 2^{-nH}. \tag{15}$$

For more details on large deviations for Bernoulli sums (including a detailed discussion of Proposition 4.1), we refer the reader to [6]. A more complete introduction to the theory of large deviations can be found in [4]. For a proof of Proposition 4.2, see [4] and [9].

4.2 Kolmogorov ε-entropy

We need a few concepts from the theory of Banach spaces (cf. [8]). Let Y be a Banach space and $X \subset Y$ a compact subset. A set $\{f_i\}_{i\in I}$, $f_i \in Y$, is called an *ε-net of X in Y* if each $x \in X$ satisfies $\|x - f_i\|_\infty \leq \varepsilon$ for some $i \in I$. Let $N = N_\varepsilon$ be the smallest number of functions $f_1, \ldots, f_N \in Y$ forming an ε-net of X in Y. The quantity

$$H_\varepsilon := \log_2 N_\varepsilon \tag{16}$$

is the Kolmogorov ε-entropy (or metric entropy) of X in Y.

Recall that we use the notation $\mathcal{B}_1(I,\mu)$ to refer to the class of functions $x : I \to [-\mu, \mu]$ that are restrictions (to the interval I) of functions in $\mathcal{B}_1(\mathbb{R},\mu)$. This is a compact subspace of $C(I)$ with respect to the norm $\|\cdot\|_\infty$. The Kolmogorov ε-entropy of $B_1(I,\mu)$ in $C(I)$ is shift invariant and can thus be denoted by $H_\varepsilon(|I|)$. It is known [8] that the average Kolmogorov ε-entropy (per unit interval) of this space, defined by

$$\bar{H}_\varepsilon := \lim_{|I|\to\infty} \frac{1}{|I|} H_\varepsilon(|I|) \tag{17}$$

exists and has the asymptotic behavior

$$\bar{H}_\varepsilon = (1 + o(1)) \log_2 \frac{\mu}{\varepsilon} \quad \text{as } \epsilon \to 0 \ . \tag{18}$$

Note that we may rewrite $\log\left(\frac{\mu}{\varepsilon}\right) = \log\left(\frac{1}{\varepsilon}\right)(1 + o(1))$ as $\varepsilon \to 0$, so that the asymptotic behavior of $\bar{H}_\varepsilon$ is independent of μ.

The average Kolmogorov ε-entropy of the space

$$\mathcal{B}_1^\delta(\mathbb{R},\mu) := \mathcal{B}_1(\mathbb{R},\mu)\cap\left\{f\in L^\infty \mid \forall x : f(t)\in[\mu-\delta,\mu]\right\} = \mu-\frac{\delta}{2}+\mathcal{B}_1(\mathbb{R},\delta/2). \tag{19}$$

has the same asymptotic behavior as that of $\mathcal{B}_1(\mathbb{R},\mu)$. To see this, we use that adding a constant does not change the ε-entropy.

5 Proof of Theorem 3.1

We are now equipped with the necessary tools to prove our main result, Theorem 3.1. We proceed by contradiction; more specifically we will show that under the assumption that $E_K^\mu \le C2^{-\alpha K\lambda}$ for fixed constants $\alpha > 1 - K^{-1}(1-h(\frac{1+\mu}{2}))$ and $C>0$ and all $\lambda>1$, one can construct ε-nets for spaces of the type $\mathcal{B}_1^\delta(I,\mu)$ that violate the asymptotic bounds for the average Kolmogorov ε-entropy given in Section 4.2.

5.1 An ε-net for the whole space $\mathcal{B}_1(I,\mu)$

Let us restrict our attention to compact intervals of the form $I=[-a,a]$. Then, closely following [7], we introduce $T_0(\lambda)$ for all $\lambda>1$ to be the smallest number that satisfies

$$\int_{T_0(\lambda)-\frac{1}{\lambda}}^{\infty} \rho(s)ds \le 2^{-\alpha K\lambda}, \tag{20}$$

where $\rho\in L^1(\mathbb{R})$ is even symmetric on $\mathbb{R}$, monotonically decreases on $\mathbb{R}^+$, and bounds $|\varphi|$ from above everywhere. This quantity can be interpreted as the margin that needs to be added to control the tail behavior of φ. For this reason, we consider the larger 'padded' interval $\widetilde{I}=[-a-T_0(\lambda),a+T_0(\lambda)]$, its dilation $\lambda\widetilde{I}=[-\lambda(a+T_0(\lambda)),\lambda(a+T_0(\lambda))]$, and the truncated approximation

$$\widetilde{f}_\lambda(t) = \frac{1}{\lambda}\sum_{\mathbb{Z}\cap\lambda\widetilde{I}} q_n^\lambda \varphi\left(t-\frac{n}{\lambda}\right). \tag{21}$$

Restricting to $t\in I$, this function is close to any possible extension of the form $\widetilde{x}_\lambda = \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} q_n^\lambda\varphi\left(t-\frac{n}{\lambda}\right)$. Indeed, for $n\in\mathbb{Z}\setminus\lambda\widetilde{I}$ one has $\left|t-\frac{n}{\lambda}\right| > T_0(\lambda)$,

so that

$$|\widetilde{x}_\lambda(t) - \widetilde{f}_\lambda(t)| \leq \frac{1}{\lambda} \sum_{\mathbb{Z}\setminus\lambda\widetilde{I}} \left|\varphi\left(t - \frac{n}{\lambda}\right)\right| \leq 2 \int_{T_0(\lambda)-\frac{1}{\lambda}}^{\infty} \rho(s)ds \leq 2^{-\alpha K\lambda+1}; \quad (22)$$

Recall that $E_K^\mu(\lambda) = \sup_{x\in\mathcal{B}_1(\mathbb{R},\mu)} \inf_{q_n^\lambda\in\mathcal{A}_K} \|x - \widetilde{x}_\lambda\|_{L^\infty(\mathbb{R})} =\leq C2^{-\alpha K\lambda}$ is assumed. Then

$$\|x - \widetilde{f}_\lambda\|_{L^\infty(I)} \leq \|x - \widetilde{x}_\lambda\|_{L^\infty(I)} + \|\widetilde{x}_\lambda - \widetilde{f}_\lambda\|_{L^\infty(I)} \quad (23)$$

$$\leq C'2^{-\alpha K\lambda} =: \varepsilon. \quad (24)$$

That is, for this choice of ε, the $\widetilde{f}_\lambda$'s form an ε-net for the space $\mathcal{B}_1(I,\mu)$. It is clear that as x varies in the set $\mathcal{B}_1(\mathbb{R},\mu)$, the resulting ε-net F_λ has cardinality at most $2^{K|\mathbb{Z}\cap\lambda\widetilde{I}|}$.

5.2 An ε-net for the reduced space $\mathcal{B}_1^\delta(I,\mu)$

By our main assumption there is a fixed constant α_0 such that $\alpha > \alpha_0 > 1 - K^{-1}(1 - h(\frac{1+\mu}{2}))$. By continuity of h, we may fix $\delta > 0$ sufficiently small that $\alpha_0 \geq 1-K^{-1}(1-h(\frac{1}{2}+\frac{\mu-5\delta}{2}))$. For this choice of δ, we will now estimate the size of the ε-net F_λ^δ arising in the same way as F_λ when x varies only over $\mathcal{B}_1^\delta(\mathbb{R},\mu)$. Note that δ may depend on μ but is independent of λ. Hence we can assume without loss of generality that λ is large enough to ensure $\varepsilon \leq \delta$. Note that for all t one has $x(t) \geq \mu - \delta$, thus $\widetilde{x}_\lambda(t) \geq \mu - \delta - \varepsilon \geq \mu - 2\delta$, and, by (22), $\widetilde{f}_\lambda(t) \geq \mu - 3\delta$ for $t \in I$. Consequently,

$$F_\lambda^\delta \subset F_\lambda \cap \left\{f \in \mathcal{B}_1(\mathbb{R},\mu) \mid \forall t \in I : f(t) \geq \mu - 3\delta\right\}. \quad (25)$$

Let $G_\lambda = \left[\mathcal{A}_K\right]^{\mathbb{Z}\cap\lambda\widetilde{I}}$, and consider the subset of this class given by

$$G_\lambda^\delta := \left\{ q_n^\lambda \in G_\lambda : \forall t \in I : \frac{1}{\lambda}\sum_{\mathbb{Z}\cap\lambda\widetilde{I}} q_n^\lambda \varphi\left(t - \frac{n}{\lambda}\right) \geq \mu - 3\delta \right\}. \quad (26)$$

Now consider a random variable $\mathbf{Q}$ distributed according to a uniform probability measure on G_λ. We observe that

$$|F_\lambda^\delta| \leq |G_\lambda^\delta| = \mathbb{P}(\mathbf{Q} \in G_\lambda^\delta)|G_\lambda| = \mathbb{P}(\mathbf{Q} \in G_\lambda^\delta)2^{K|\mathbb{Z}\cap\lambda\widetilde{I}|} \leq \mathbb{P}(\mathbf{Q} \in G_\lambda^\delta)2^{\lceil\lambda K(|I|+2T_0(\lambda))\rceil}. \quad (27)$$

We would now like to estimate $\mathbb{P}(\mathbf{Q} \in G_\lambda^\delta)$:

1. Note that $\mathbf{Q}$ agrees in distribution with a sequence of identically distributed independent variables Q_n, $n \in \mathbb{Z} \cap \lambda\widetilde{I}$, which have support on $[-1,1]$ and expectation $\mathbb{E}\big(Q_n\big) = 0$. Consequently, one obtains

$$\begin{aligned}\mathbb{P}(\mathbf{Q} \in G_\lambda^\delta) \leq & \mathbb{P}\left(\forall j \in \mathbb{Z} \cap \lambda I : \frac{1}{\lambda} \sum_{n \in \mathbb{Z} \cap \lambda\widetilde{I}} Q_n \varphi\left(\frac{j-n}{\lambda}\right) \geq \mu - 3\delta\right) \\ \leq & \mathbb{P}\left(\frac{1}{\lambda} \sum_{j \in \mathbb{Z} \cap \lambda I} \sum_{n \in \mathbb{Z} \cap \lambda\widetilde{I}} Q_n \varphi\left(\frac{j-n}{\lambda}\right) \geq |\mathbb{Z} \cap \lambda I|(\mu - 3\delta)\right) \\ = & \mathbb{P}\left(\sum_{n \in \mathbb{Z} \cap \lambda\widetilde{I}} c_n Q_n \geq |\mathbb{Z} \cap \lambda I|(\mu - 3\delta)\right), \end{aligned} \tag{28}$$

   where $c_n = \frac{1}{\lambda} \sum_{j \in \mathbb{Z} \cap \lambda I} \varphi\left(\frac{j-n}{\lambda}\right)$.

2. We would like to bound the coefficients c_n. By assumption, we have $\varphi \in \mathcal{S}$. Therefore, we may apply the Poisson Summation Formula,

$$\frac{1}{\lambda} \sum_{j \in \mathbb{Z}} \varphi\left(\frac{j-n}{\lambda}\right) = \sum_{k \in \mathbb{Z}} \widehat{\varphi}(k\lambda) = \widehat{\varphi}(0) + O(\lambda^{-1}) = 1 + O(\lambda^{-1}). \tag{29}$$

   From now on, we assume that $a > T_0(\lambda)$. This assumption makes sense as in the definition of the average Kolmogorov ε-entropy, one lets $|I| \to \infty$ for each fixed λ. Then the interval $\widehat{I} := [-(a - T_0(\lambda)), a - T_0(\lambda)]$ is non-empty; it satisfies $\widehat{I} \subset I \subset \widetilde{I}$.

   Now for $n \in \lambda\widehat{I}$, we have

$$\left|\frac{1}{\lambda} \sum_{j \in \mathbb{Z} \cap \lambda I} \varphi\left(\frac{j-n}{\lambda}\right) - 1\right| \leq 2\delta + O(\lambda^{-1}) \tag{30}$$

   as a consequence of (20). Furthermore, for $n \notin \lambda\widehat{I}$, we use the crude estimate

$$\left|\frac{1}{\lambda} \sum_{j \in \mathbb{Z} \cap \lambda I} \varphi\left(\frac{j-n}{\lambda}\right) - 1\right| \leq \|\rho\|_1 + 1 + \rho(0) =: D \tag{31}$$

   in conjunction with the bound

$$|\mathbb{Z} \cap (\lambda\widetilde{I} \setminus \lambda\widehat{I})| \leq 4\lambda T_0(\lambda) + 4 =: N(\lambda). \tag{32}$$

3. We now apply these bounds for the coefficients c_n in (28). As $\widehat{I} \subset I$, and $|\widetilde{I}| = O(\lambda)$, we obtain

$$\mathbb{P}(\mathbf{Q} \in G_\lambda^\delta) \leq \mathbb{P}\left(\sum_{\mathbb{Z}\cap\lambda\widetilde{I}} Q_n \geq |\mathbb{Z}\cap\lambda I|(\mu - 5\delta - O(\lambda^{-1})) - DN(\lambda)\right) \tag{33}$$

$$\leq \mathbb{P}\left(\sum_{\mathbb{Z}\cap\lambda I} Q_n \geq |\mathbb{Z}\cap\lambda I|(\mu - 5\delta) - D'N(\lambda)\right). \tag{34}$$

Rescaling the random variables Q_n to yield independent and identically distributed random variables supported on $[0,1]$ with expectation equal to $1/2$, we may apply Propositions 4.2 and 4.1 to bound the probability of such a large deviation from the mean:

$$\log_2 \mathbb{P}(\mathbf{Q} \in G_\lambda^\delta) \leq -|\mathbb{Z}\cap\lambda I| h\Big(.5\big(1 + \mu - 5\delta + C'N(\lambda)\big)\cdot|\mathbb{Z}\cap\lambda I|^{-1}\big)\Big) \tag{35}$$

$$\leq -(\lambda|I| - 1)\Big(h\Big(.5\big(1+\mu-5\delta\big)\cdot\big(1+O(|I|^{-1})\big)\Big) - 1\Big). \tag{36}$$

Combining our estimate for $\mathbb{P}(\mathbf{Q} \in G_\lambda^\delta)$ with (27), we obtain that as x varies in $\mathcal{B}_1^\delta(\mathbb{R},\mu)$, the f_λ vary on a set F_λ^δ of cardinality at most

$$N := 2^{\lambda|I|\left(K-1+h(\frac{1}{2}+\frac{\mu-5\delta}{2})\right)\cdot(1+O(|I|^{-1}))} \tag{37}$$

As a consequence, for each $\lambda > 1$, there are arbitrarily long intervals I such that, for each I, there is an ε-net of $\mathcal{B}_1^\delta(I,\mu)$ with at most N elements.

5.3 Towards a contradiction

We have found that the size of a ε-net of $\mathcal{B}_1^\delta(I,\mu)$ is bounded above by N, as given by equation (37). Thus, we may bound the Kolmogorov ε-entropy H_ε of the space $\mathcal{B}_1^\delta(I,\mu)$ (see section 4.2) by

$$H_\varepsilon(|I|) \leq \log_2(N) = \lambda|I|\Big(K - 1 + h\Big(\frac{1}{2} + \frac{\mu-5\delta}{2}\Big)\Big)\big(1 + O(|I|^{-1})\big). \tag{38}$$

As $|I| \to \infty$, we may bound the average Kolmogorov ε-entropy $\bar{H}_\varepsilon = \lim_{|I|\to\infty} \frac{H_\epsilon(|I|)}{|I|}$ by $\bar{H}_\varepsilon \leq \lambda\Big(K - 1 + h(\frac{1}{2} + \frac{\mu-5\delta}{2})\Big)$. Note that this also gives a bound on the average Kolmogorov ε-entropy for the larger space $\mathcal{B}_1$. Recalling that $\varepsilon = C'2^{-\alpha K\lambda}$, or $\log_2 \frac{1}{\varepsilon} = \alpha K\lambda - \log_2 C'$, and also recalling our hypothesis that $\alpha > \alpha_0 \geq 1 - K^{-1} + K^{-1}h(\frac{1+\mu-5\delta}{2})$, we arrive at the chain of inequalities

$$\alpha - \frac{\log_2 C'}{K\lambda} \leq \left(\log_2 \frac{1}{\varepsilon}\right) \frac{1 - K^{-1} + K^{-1}h\big(\frac{1}{2} + \frac{\mu-5\delta}{2}\big)}{\bar{H}_\varepsilon} \leq \left(\log_2 \frac{1}{\varepsilon}\right) \frac{\alpha_0}{\bar{H}_\varepsilon} \quad (39)$$

This establishes a contradiction to (18), as $\lambda \to \infty$ and consequently $\varepsilon \to 0$. □

Remark The assumption that the kernel $\varphi \in \mathcal{S}$ in our main theorem is stronger than necessary. We used this assumption only to apply the Poisson Summation Formula in the proof of Theorem 3.1; a weaker but more technical requirement for this formula to hold is that $|\varphi(t)| + |\widehat{\varphi(t)}| \leq C(1+t)^{-(1+\delta)}$. In particular, our proof also works for twice continuously differentiable kernels φ with compact support, a scenario resembling filters used in practice.

Acknowledgments

The authors would like to thank Sinan Güntürk for interesting discussions on this topic as well as the Hausdorff Center for Mathematics, Bonn, and the Summer School on "Theoretical Foundations and Numerical Methods for Sparse Recovery" at the RICAM, Linz, where large parts of the project were completed. They gratefully acknowledge the support of a National Science Foundation Postdoctoral Research Fellowship (Ward) and the Charles M. Newman Fellowship at the Courant Institute (Krahmer).

References

[1] A. Calderbank and I. Daubechies. The pros and cons of democracy. *IEEE Transactions on Information Theory*, 48:1721–1725, 2002.

[2] P. Deift, S. Güntürk, and F. Krahmer. An optimal family of exponentially accurate one-bit sigma-delta quantization schemes. submitted.

[3] M. Derpich, E. Silva, D. Quevado, and G. Goodwin. On optimal perfect reconstruction feedback quantizers. *IEEE Transactions on Signal Processing*, 56:3871–3890, 2008.

[4] R. Ellis. *Entropy, Large Deviations and Statistical Mechanics*. Springer Publication, 2005.

[5] I. Galton. Delta-sigma data conversion in wireless transceivers. *Microwave Theory and Techniques, IEEE Transactions on*, 50(1):302 –315, Jan 2002.

[6] L. Gordon and R. Arratia. Tutorial on large deviations for the binomial distribution. *Bulletin of Mathematical Biology*, 51(1), 1989.

[7] C. Güntürk. One-bit sigma-delta quantization with exponential accuracy. *Communications on Pure and Applied Mathematics*, (11):1608–1630, 2003.

[8] A. Kolmogorov and M. Tihomirov. ϵ-entropy and ϵ-capacity of sets in function spaces. *Uspehi Mat Nauk*, (2):3–86, 1959.

[9] C. Leon and F. Perron. Extremal properties of sums of Bernoulli random variables. *Statistics and Probability Letters*, (62):345–354, 2003.

[10] S. Norsworthy, R. Schreier, and G. Temes. *Delta-Sigma Data Converters: Theory, Design and Simulation*. New York: IEEE Press, 1997.