On the Impact of Thresholding and Quantization on the Behavior of Bandlimited Systems and Signals

Holger Boche and Ullrich J. Mönich
Technische Universität Berlin, Lehrstuhl für Mobilkommunikation
Berlin, Germany, Email: {holger.boche, ullrich.moenich}@mk.tu-berlin.de

***Abstract*—In this paper we analyze the impact of thresholding and quantization on the approximation of bandlimited signals and systems by sampling series that only use samples of the signal taken at the Nyquist rate. We show that there are stable systems that become unstable if the samples are quantized. For these systems the approximation error is unbounded irrespective of how small the quantization step size is chosen. Furthermore, we completely characterize the systems for which the approximation is stable under thresholding and quantization. Surprisingly, this class of systems is the well-known class of bounded-input bounded-output (BIBO) stable systems. Moreover, we discuss the special case of finite impulse response (FIR) filters and give an upper bound for the approximation error that shows the dependence of the error on the filter length.**

***Index Terms*—Quantization, thresholding, approximation, linear time-invariant system, Shannon sampling series**

I. Introduction and Notation

The reconstruction of bandlimited signals from their samples is important not only from a theoretical point of view [1] but also for many practical applications. The principle of digital signal processing relies on the fact that certain bandlimited signals can be perfectly reconstructed from their samples. However, this is only true if the sample values are known exactly. In real applications this can never be realized because the quantization process in analog to digital conversion only has limited resolution [2], [3].

In this paper, we consider two non-linear distortions of the samples and analyze their effect on the approximation behavior of the sampling series. The first non-linear distortion is the quantization operator and the second is the threshold operator.

One application where the threshold operator is important is sensor networks. The sensors sample some bandlimited signal in space and time, and in order to save energy, the sensors transmit only if the absolute value of the sample exceeds some threshold. Thus, the receiver has to reconstruct the signal by using only the samples whose absolute value is larger or equal to the threshold.

So far we have discussed the approximation of signals from their disturbed samples. In [4] the more general case, where some transformation Tf of the signal f has to be approximated, was analyzed, and the convergence behavior of the sampling series was studied. However, in [4] and other publications that study the approximation of such transforms [5], the convergence analysis was done only for perfectly known sample values of f. In this paper we consider the more realistic case where the samples are disturbed by the threshold operator and quantization operator.

In order to continue, we need some notation and definitions. Let $\hat{f}$ denote the Fourier transform of a function f, where $\hat{f}$ is to be understood in the distributional sense. $L^p(\mathbb{R})$, $1 \leq p < \infty$, is the space of all pth-power Lebesgue integrable functions on $\mathbb{R}$, with the usual norm $\|\cdot\|_p$, and $L^\infty(\mathbb{R})$ is the space of all functions for which the essential supremum norm $\|\cdot\|_\infty$ is finite. For $\sigma > 0$ let $\mathcal{B}_\sigma$ be the set of all entire functions f with the property that for all $\epsilon > 0$ there exists a constant $C(\epsilon)$ with $|f(z)| \leq C(\epsilon)\exp((\sigma+\epsilon)|z|)$ for all $z \in \mathbb{C}$. The Bernstein space $\mathcal{B}_\sigma^p$ consists of all functions in $\mathcal{B}_\sigma$, whose restriction to the real line is in $L^p(\mathbb{R})$, $1 \leq p \leq \infty$. A function in $\mathcal{B}_\sigma^p$ is called bandlimited to σ. By the Paley-Wiener-Schwartz theorem, the Fourier transform of a function bandlimited to σ is supported in $[-\sigma, \sigma]$. For $\sigma > 0$ and $1 \leq p \leq \infty$ we denote by $\mathcal{PW}_\sigma^p$ the Paley-Wiener space of signals f with a representation $f(z) = 1/(2\pi)\int_{-\sigma}^{\sigma} g(\omega)\,\mathrm{e}^{iz\omega}\ \mathrm{d}\omega$, $z \in \mathbb{C}$, for some $g \in L^p(-\sigma, \sigma)$. If $f \in \mathcal{PW}_\sigma^p$ then $g(\omega) = \hat{f}(\omega)$. The norm for $\mathcal{PW}_\sigma^p$, $1 \leq p < \infty$, is given by $\|f\|_{\mathcal{PW}_\sigma^p} = (1/(2\pi)\int_{-\sigma}^{\sigma}|\hat{f}(\omega)|^p\,\mathrm{d}\omega)^{1/p}$.

II. Threshold and Quantization Operator

In this paper we analyze the effect of two non-linear distortions on the signal approximation behavior of sampling series. The first distortion is modeled by the threshold operator. For complex numbers $z \in \mathbb{C}$, the threshold operator κ_δ, $\delta > 0$, is defined by

$$\kappa_\delta z = \begin{cases} z & |z| \geq \delta \\ 0 & |z| < \delta. \end{cases}$$

Furthermore, for continuous signals $f : \mathbb{R} \to \mathbb{C}$, we define the threshold operator Θ_δ pointwise, i.e., $(\Theta_\delta f)(t) = \kappa_\delta f(t)$, $t \in \mathbb{R}$. In this paper, the threshold operator κ_δ is applied on the samples $\{f(k)\}_{k\in\mathbb{Z}}$ of signals $f \in \mathcal{PW}_\pi^1$, which gives the disturbed samples $\{\kappa_\delta f(k)\}_{k\in\mathbb{Z}}$. This is, of course, equivalent to applying the threshold operator Θ_δ on the signal f itself and then taking the samples, i.e., $\{(\Theta_\delta f)(k)\}_{k\in\mathbb{Z}}$. Then, the resulting samples $\{(\Theta_\delta f)(k)\}_{k\in\mathbb{Z}}$ are used to build

ISBN 978-3-8007-3211-1 © VDE VERLAG -GMBH · Berlin · Offenbach

an approximation

$$(A_\delta f)(t) := \sum_{k=-\infty}^{\infty} (\Theta_\delta f)(k) \frac{\sin(\pi(t-k))}{\pi(t-k)}$$
$$= \sum_{\substack{k=-\infty \\ |f(k)|\geq\delta}}^{\infty} f(k) \frac{\sin(\pi(t-k))}{\pi(t-k)} \qquad (1)$$

of the original signal f. By A_δ we denote the operator that maps $f \in \mathcal{PW}_\pi^1$ to $A_\delta f$ according to (1).

The second non-linear operator that we consider in this paper is the simple but frequently used uniform mid-tread quantization, where each complex number $z \in \mathbb{C}$ is quantized to $q_\delta z$, depending on the quantization step size $2\delta > 0$, according to the rule

$$q_\delta z = \left\lfloor \frac{\operatorname{Re} z}{2\delta} + \frac{1}{2} \right\rfloor 2\delta + \left\lfloor \frac{\operatorname{Im} z}{2\delta} + \frac{1}{2} \right\rfloor 2\delta i, \qquad (2)$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal than x. As can be seen in (2), the quantization is done separately for the real and the imaginary part of z. Furthermore, for continuous signals $f : \mathbb{R} \to \mathbb{C}$, we define the quantization operator Υ_δ, $\delta > 0$, pointwise, i.e., $(\Upsilon_\delta f)(t) = q_\delta f(t)$, $t \in \mathbb{R}$. For example, if the sample $f(k)$ is a real number and $f(k) \in [(2l-1)\delta, (2l+1)\delta)$ for some $l \in \mathbb{Z}$ then $(\Upsilon_\delta f)(k) = 2l\delta$. As in the case of the threshold operator, the resulting samples $\{(\Upsilon_\delta f)(k)\}_{k\in\mathbb{Z}}$ are used to build an approximation

$$(B_\delta f)(t) := \sum_{k=-\infty}^{\infty} (\Upsilon_\delta f)(k) \frac{\sin(\pi(t-k))}{\pi(t-k)} \qquad (3)$$

of the original signal f.

The convergence of the series in (1) and (3) is unproblematic. Since $f \in \mathcal{PW}_\pi^1$ we have $\lim_{t\to\infty} f(t) = 0$ by the Riemann-Lebesgue lemma, and it follows that the series in (1) and (3) have only finitely many summands. This implies that $A_\delta f \in \mathcal{PW}_\pi^2 \subset \mathcal{PW}_\pi^1$ and $B_\delta f \in \mathcal{PW}_\pi^2 \subset \mathcal{PW}_\pi^1$. In general, $A_\delta f$ and $B_\delta f$ are only approximations of f, and we want the approximation to be close to f if δ is sufficiently small.

Both approximation processes (1) and (3) are difficult to analyze, because the threshold operator Θ_δ and the quantization operator Υ_δ are non-linear. As a consequence, for $\delta > 0$, $A_\delta : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$ and $B_\delta : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$ are non-linear operators. We do the analysis for the space $\mathcal{PW}_\pi^1$ because this space is larger than the commonly used $\mathcal{PW}_\pi^2$-space of signals with finite energy and because the convergence behavior of sampling series for signals in $\mathcal{PW}_\pi^1$ is closely related to the convergence behavior of sampling series for bandlimited wide-sense stationary stochastic processes [6].

III. Stable LTI Systems

Up to now we have only discussed the signal approximation problem for f. However, in many signal processing applications the task is to approximate some processed version Tf of $f \in \mathcal{PW}_\pi^1$ and not f itself. One frequently used type of processing is the filtering of a signal by a stable linear time-invariant (LTI) system T. So, even more interesting than the mere signal reconstruction problem is the approximation problem where Tf has to be approximated by a sampling series, which uses only the samples of f.

Before we continue the discussion, we briefly review some definitions and facts about stable LTI systems. A linear system $T : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$ is called stable if the operator T is bounded, i.e., if $\|T\| = \sup_{\|f\|_{\mathcal{PW}_\pi^1}\leq 1} \|Tf\|_{\mathcal{PW}_\pi^1} < \infty$, and time-invariant if $(Tf(\,\cdot - a))(t) = (Tf)(t-a)$ for all $f \in \mathcal{PW}_\pi^1$ and $t, a \in \mathbb{R}$. Note that this kind of stability is with respect to the $\mathcal{PW}_\pi^1$-norm and thus is different from the commonly used bounded-input bounded-output (BIBO) stability. All BIBO stable systems are stable in our sense. Furthermore, our $\mathcal{PW}_\pi^1$-stability is equivalent to energy stability, i.e., L^2-stability, and important LTI systems like the Hilbert transform and the ideal low-pass filter belong to this class.

For every stable LTI system $T : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$ there exists exactly one function $\hat{h}_T \in L^\infty[-\pi, \pi]$ such that

$$(Tf)(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(\omega) \hat{h}_T(\omega)\, \mathrm{e}^{i\omega t}\ \mathrm{d}\omega \qquad (4)$$

for all $f \in \mathcal{PW}_\pi^1$. Note that $\hat{h}_T \in L^\infty[-\pi,\pi] \subset L^2[-\pi,\pi]$, and consequently $h_T \in \mathcal{PW}_\pi^2$. Conversely, every function $\hat{h}_T \in L^\infty[-\pi,\pi]$ defines a stable LTI system $T : \mathcal{PW}_\pi^1 \to \mathcal{PW}_\pi^1$. The operator norm of a stable LTI system T is given by $\|T\| = \|\hat{h}\|_\infty$.

IV. System Approximation and Thresholding

If the samples $\{f(k)\}_{k\in\mathbb{Z}}$ are known perfectly we can use

$$\sum_{k=-N}^{N} f(k)\, T(\operatorname{sinc}(\,\cdot - k))(t) = \sum_{k=-N}^{N} f(k) h_T(t-k) \qquad (5)$$

to obtain an approximation of Tf. The conditions under which (5) converges to Tf as N goes to infinity were analyzed in [4]. In this paper we analyze the signal approximation problem, where the samples are disturbed either by the non-linear threshold operator or by the non-linear quantization operator. More concretely, we want to approximate Tf either by

$$(TA_\delta f)(t) = \sum_{k=-\infty}^{\infty} (\Theta_\delta f)(k) h_T(t-k)$$

or by

$$(TB_\delta f)(t) = \sum_{k=-\infty}^{\infty} (\Upsilon_\delta f)(k) h_T(t-k).$$

As an example, we focus in the following on the threshold operator and will use the abbreviation $T_\delta := TA_\delta$. Nevertheless, the results are also true for the quantization operator. The approximation error can be upper bounded by

$$\begin{aligned} |(T_\delta f)(t) - (Tf)(t)| &\leq |(T_\delta f)(t)| + |(Tf)(t)| \\ &\leq |(T_\delta f)(t)| + \|Tf\|_{\mathcal{PW}_\pi^1} \\ &\leq |(T_\delta f)(t)| + \|T\|\, \|f\|_{\mathcal{PW}_\pi^1}. \qquad (6) \end{aligned}$$

Furthermore, if

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(t)| < \infty$$

we can conclude from (6) that

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(t) - (Tf)(t)| < \infty,$$

i.e., that the approximation error is bounded for all signals $f \in \mathcal{PW}_\pi^1$ with $\|f\|_{\mathcal{PW}_\pi^1} \le 1$.

The following theorem gives a necessary and sufficient condition for $\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1}|(T_\delta f)(t)|$ to be finite. In Corollary 2 we will see that the same condition is sufficient for a good approximation behavior of $(T_\delta f)(t)$.

Theorem 1. *Let T be a stable LTI system, $0 < \delta < 1/3$, and $t \in \mathbb{R}$. Then we have*

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(t)| < \infty$$

if and only if

$$\sum_{k=-\infty}^{\infty} |h_T(t-k)| < \infty. \tag{7}$$

Remark 1. The requirement $\delta < 1/3$ has the following reason. Since $\|f\|_\infty \le \|f\|_{\mathcal{PW}_\pi^1} \le 1$, we only consider signals whose peak value is bounded by 1. Thus, it only makes sense to consider thresholds δ that are smaller than or equal to 1. The specific value of $1/3$ is due to technical issues in the proof.

Corollary 1. *Let T be a stable LTI system, $0 < \delta < 1/3$, and $t \in \mathbb{R}$. If* (7) *is not fulfilled then*

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(t)| = \infty. \tag{8}$$

Corollary 1 shows that

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(Tf)(t) - (T_\delta f)(t)| = \infty$$

if the system T does not fulfill (7). Thus, the pointwise approximation error cannot be controlled, regardless of how small the threshold δ is chosen. Clearly, for every $f \in \mathcal{PW}_\pi^1$, $|(T_\delta f)(t)|$ is bounded. However, according to (8), for any level $L > 0$ we can find a signal $f_1 \in \mathcal{PW}_\pi^1$ with $\|f_1\|_{\mathcal{PW}_\pi^1} \le 1$ such that $|(T_\delta f)(t)| > L$.

Note that (7) is nothing else than the BIBO stability condition for discrete-time systems.

On the other hand, if (7) is fulfilled, then we have a good pointwise approximation behavior because the approximation error converges to zero as the threshold δ goes to zero.

Corollary 2. *Let T be a stable LTI system and $t \in \mathbb{R}$. If* (7) *is fulfilled then we have*

$$\lim_{\delta\to 0} \sup_{f\in\mathcal{PW}_\pi^1} |(Tf)(t) - (T_\delta f)(t)| = 0.$$

Proof: Taking the supremum on both sides of

$$\begin{aligned}
&|(Tf)(t) - (T_\delta f)(t)| \\
&= \left| \sum_{k=-\infty}^{\infty} f(k) h_T(t-k) - \sum_{k=-\infty}^{\infty} (\Theta_\delta f)(k) h_T(t-k) \right| \\
&= \left| \sum_{k=-\infty}^{\infty} (f(k) - (\Theta_\delta f)(k)) h_T(t-k) \right| \\
&\le \delta \sum_{k=-\infty}^{\infty} |h_T(t-k)|
\end{aligned} \tag{9}$$

proves the statement. ■

Remark 2. With (9) we have a universal bound for the approximation error, which is independent of f.

In order to prove Theorem 1 we need Lemma 1.

Lemma 1. *For all stable LTI systems T, $0 < \delta < 1/3$, and $t \in \mathbb{R}$ we have*

$$\frac{\delta}{2} \sum_{k=-\infty}^{\infty} |h_T(t-k)| \le \sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(t)| \le \sum_{k=-\infty}^{\infty} |h_T(t-k)| \tag{10}$$

Proof: The right inequality in (10) follows directly from

$$\begin{aligned}
|(T_\delta f)(t)| &= \left| \sum_{\substack{k=-\infty\\|f(k)|\ge\delta}}^{\infty} f(k) h_T(t-k) \right| \\
&\le \sum_{\substack{k=-\infty\\|f(k)|\ge\delta}}^{\infty} |f(k)||h_T(t-k)| \\
&\le \|f\|_{\mathcal{PW}_\pi^1} \sum_{k=-\infty}^{\infty} |h_T(t-k)|.
\end{aligned}$$

The left inequality in (10) needs some more reasoning. Let $0 < \delta < 1/3$ and $t \in \mathbb{R}$ be arbitrary but fixed. Furthermore, let $\mathcal{Z}^+ = \{k \in \mathbb{Z} : h_T(t-k) \ge 0\}$ and $\mathcal{Z}^- = \{k \in \mathbb{Z} : h_T(t-k) < 0\}$. For $0 < \eta < 1$ and $N \in \mathbb{N}$, consider the function

$$h^+(t,\eta,N) := \sum_{k=-2N+1}^{2N-1} h^+(k,\eta,N) \frac{\sin(\pi(t-k))}{\pi(t-k)},$$

where

$$h^+(k,\eta,N) = \begin{cases} 1+\eta, & k \in \mathcal{Z}^+ \cap [-N,N], \\ 1-\eta, & k \in \mathcal{Z}^- \cap [-N,N], \\ 2 - \frac{|k|}{N}, & N < |k| < 2N. \end{cases}$$

We have

$$\begin{aligned}
h^+(t,\eta,N) &= h^+(t,0,N) \\
&\quad + \eta \underbrace{\sum_{\substack{k=-N\\k\in\mathcal{Z}^+}}^{N} \frac{\sin(\pi(t-k))}{\pi(t-k)}}_{=:u_N^+(t)} - \eta \underbrace{\sum_{\substack{k=-N\\k\in\mathcal{Z}^-}}^{N} \frac{\sin(\pi(t-k))}{\pi(t-k)}}_{=:u_N^-(t)},
\end{aligned}$$

and it follows that

$$\|h^+(\,\cdot\,,\eta,N)\|_{\mathcal{PW}_\pi^1} \le \|h^+(\,\cdot\,,0,N)\|_{\mathcal{PW}_\pi^1} + \eta\|u_N^+\|_{\mathcal{PW}_\pi^1} + \eta\|u_N^-\|_{\mathcal{PW}_\pi^1}.$$

Since $\|h^+(\,\cdot\,,0,N)\|_{\mathcal{PW}_\pi^1} < 3$, which is proven in the appendix, and $\|u_N^+\|_{\mathcal{PW}_\pi^1} < \infty$ as well as $\|u_N^-\|_{\mathcal{PW}_\pi^1} < \infty$ for all $N \in \mathbb{N}$, there exists an $\eta_0 = \eta_0(N)$ with $0 < \eta_0 < 1$ such that $\|h^+(\,\cdot\,,\eta_0,N)\|_{\mathcal{PW}_\pi^1} < 3$. Now, let $g^+(t) := \delta h^+(t,\eta_0,N)$. Note that $\|g^+\|_{\mathcal{PW}_\pi^1} < 1$. We have

$$\begin{aligned}(T_\delta g^+)(t) &= \sum_{\substack{k=-\infty\\ |g^+(k)|\ge\delta}}^{\infty} g^+(k) h_T(t-k)\\ &= (1+\eta_0)\delta \sum_{\substack{k=-N\\ k\in\mathcal{Z}^+}}^{N} h_T(t-k)\\ &> \delta \sum_{\substack{k=-N\\ k\in\mathcal{Z}^+}}^{N} h_T(t-k)\end{aligned}$$

and consequently

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} (T_\delta f)(t) \ge \delta \sum_{\substack{k=-\infty\\ k\in\mathcal{Z}^+}}^{\infty} h_T(t-k). \tag{11}$$

Analogously to $h^+(t,\eta,N)$ we define

$$h^-(t,\eta,N) := \sum_{k=-2N+1}^{2N-1} h^-(k,\eta,N)\frac{\sin(\pi(t-k))}{\pi(t-k)},$$

where

$$h^-(k,\eta,N) = \begin{cases}-(1+\eta), & k\in\mathcal{Z}^-\cap[-N,N],\\ -(1-\eta), & k\in\mathcal{Z}^+\cap[-N,N],\\ -(2-\frac{|k|}{N}), & N<|k|<2N,\end{cases}$$

and the function $g^-(t) := \delta h^-(t,\eta_1,N)$, where $\eta_1 = \eta_1(N)$, $0 < \eta_1 < 1$, is chosen such that $\|h^-(\,\cdot\,,\eta_1,N)\|_{\mathcal{PW}_\pi^1} < 3$, which implies that $\|g^-\|_{\mathcal{PW}_\pi^1} < 1$. Moreover, we have

$$\begin{aligned}(T_\delta g^-)(t) &= \sum_{\substack{k=-\infty\\ |g^-(k)|\ge\delta}}^{\infty} g^-(k) h_T(t-k)\\ &= -(1+\eta_1)\delta \sum_{\substack{k=-N\\ k\in\mathcal{Z}^-}}^{N} h_T(t-k)\\ &= (1+\eta_1)\delta \sum_{\substack{k=-N\\ k\in\mathcal{Z}^-}}^{N} |h_T(t-k)|\\ &> \delta \sum_{\substack{k=-N\\ k\in\mathcal{Z}^-}}^{N} |h_T(t-k)|,\end{aligned}$$

and consequently

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} (T_\delta f)(t) \ge \delta \sum_{\substack{k=-\infty\\ k\in\mathcal{Z}^-}}^{\infty} |h_T(t-k)|. \tag{12}$$

Combining (11) and (12) finally gives

$$\begin{aligned}2\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} (T_\delta f)(t) &\ge \delta \sum_{\substack{k=-N\\ k\in\mathcal{Z}^+}}^{N} h_T(t-k) + \delta \sum_{\substack{k=-N\\ k\in\mathcal{Z}^-}}^{N} |h_T(t-k)|\\ &= \delta \sum_{k=-\infty}^{\infty} |h_T(t-k)|,\end{aligned}$$

which completes the second part of the proof. ■

Proof of Theorem 1: Theorem 1 follows directly from Lemma 1. ■

The following example illustrates Theorem 1 and shows that even for common stable LTI systems like the ideal low-pass filter there are problems because (7) is not fulfilled.

Example 1. If T_{L} is the ideal low-pass filter with $h_{T_{\mathrm{L}}}(t) = \sin(\pi t)/(\pi t)$ then we have

$$\sum_{k=-\infty}^{\infty} |h_{T_{\mathrm{L}}}(t-k)| = \infty$$

for all $t \in (0,1)$. It follows from Theorem 1 that, for $t \in (0,1)$ and $0 < \delta < 1/3$,

$$\begin{aligned}&\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_{\mathrm{L},\delta} f)(t)|\\ &\qquad= \sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} \left|\sum_{\substack{k=-\infty\\ |f(k)|\ge\delta}}^{\infty} f(k)\frac{\sin(\pi(t-k))}{\pi(t-k)}\right| = \infty.\end{aligned}$$

This shows that, for $t \in (0,1)$ and any δ with $0 < \delta < 1/3$, the approximation error $|(T_{\mathrm{L}} f)(t) - (T_{\mathrm{L},\delta} f)(t)|$ can be arbitrarily large depending on the signal $f \in \mathcal{PW}_\pi^1$.

Similar to Theorem 1, which characterizes the pointwise boundedness of $\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(t)|$, we can also give a necessary and sufficient condition for the uniform boundedness on the whole real axis.

Theorem 2. *Let T be a stable LTI system and $0 < \delta < 1/3$. We have*

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} \|T_\delta f\|_\infty < \infty \tag{13}$$

if and only if

$$\sup_{0\le t\le 1} \sum_{k=-\infty}^{\infty} |h_T(t-k)| < \infty. \tag{14}$$

Proof: Theorem 2 follows directly from Lemma 1 by taking the supremum $\sup_{t\in\mathbb{R}}$ of all parts of (10) and the fact that $\sum_{k=-\infty}^{\infty} |h_T(t-k)|$ is periodic with period 1. ■

Corollary 3. *Let T be a stable LTI system and $0 < \delta < 1/3$. We have* (13) *if and only if $h_T \in \mathcal{B}^1_\pi$, i.e., if and only if*

$$\int_{-\infty}^{\infty} |h_T(\tau)| \,\mathrm{d}\tau < \infty. \tag{15}$$

Proof: According to Nikol'skiĭ's inequality [7, p. 49], (14) is true if and only if $\int_{-\infty}^{\infty} |h_T(\tau)| \,\mathrm{d}\tau < \infty$. ■

Note that (15) is nothing else than the BIBO stability condition for continuous-time systems. Corollary 4 shows the good global approximation behavior of $T_\delta f$ if (15) is fulfilled.

Corollary 4. *Let T be a stable LTI system. If* (15) *is fulfilled then we have*

$$\lim_{\delta\to\infty} \sup_{f\in\mathcal{PW}^1_\pi} \|(Tf)(t) - (T_\delta f)(t)\|_\infty = 0.$$

Proof: Analogously to the proof of Corollary 2. ■

V. Quantization

The results in Section IV that were obtained for the threshold operator are also true if the non-linear distortion is the quantization operator.

Theorem 3. *Let T be a stable LTI system, $0 < \delta < 1/3$, and $t \in \mathbb{R}$. Then we have*

$$\sup_{\|f\|_{\mathcal{PW}^1_\pi} \le 1} |(TB_\delta f)(t)| < \infty$$

if and only if

$$\sum_{k=-\infty}^{\infty} |h_T(t-k)| < \infty.$$

Proof: Theorem 3 can be proven very similar to Theorem 1. ■

VI. FIR Filters

Since finite impulse response (FIR) systems are an important special case of stable LTI systems, we discuss some implications for those systems next.

Definition 1. We call a stable LTI system T finite impulse response system if $\hat{h}_T$ is a polynomial in $\mathrm{e}^{-i\omega}$, i.e., if $\hat{h}_T$ has the representation

$$\hat{h}_T(\omega) = \sum_{k=0}^{M} c_k \,\mathrm{e}^{-i\omega k}, \quad -\pi \le \omega \le \pi,$$

for some $M \in \mathbb{N}$ and $c_k \in \mathbb{C}$, $k = 0, \dots, M$.

A FIR system is called a finite impulse response system, because the discrete-time impulse response $\{h_T(k)\}_{k\in\mathbb{Z}}$ has only finitely many non-zero elements. This implies that $\sum_{k=-\infty}^{\infty} |h_T(k)| < \infty$ for every FIR system.

Corollary 3 shows that it is important to know whether $h_T \in \mathcal{B}^1_\pi$ or not because the boundedness of $\|T_\delta f\|_\infty$ is completely determined by this. For FIR systems T it is possible to classify when $h_T \in \mathcal{B}^1_\pi$, based on a property of $\hat{h}_T$, which is easy to check.

Lemma 2. *Let T be a FIR system. Then $h_T \in \mathcal{B}^1_\pi$ if and only if $\lim_{|\omega|\to\pi} \hat{h}_T(\omega) = 0$.*

Proof: "⇒": Since $h_T \in \mathcal{B}^1_\pi$, it follows that $\hat{h}_T$ is continuous [8, p. 153]. Therefore, $\lim_{|\omega|\to\pi} \hat{h}_T(\omega) = 0$.

"⇐": We have

$$\int_{-\infty}^{\infty} |h_T(t)| \,\mathrm{d}t = \int_{-\infty}^{\infty} |h_T(t)| \frac{1+|t|}{1+|t|} \,\mathrm{d}t$$
$$\le \left(\int_{-\infty}^{\infty} |h_T(t)|^2 (1+|t|)^2 \,\mathrm{d}t \right)^{1/2} \left(\int_{-\infty}^{\infty} (1+|t|)^{-2} \,\mathrm{d}t \right)^{1/2}.$$

Since the second integral is finite and h_T is bounded, all that remains to be shown is that

$$\left(\int_{-\infty}^{\infty} |h_T(t)|^2 t^2 \,\mathrm{d}t \right)^{1/2} < \infty.$$

Obviously, we have $\hat{h}_T \in L^1[-\pi,\pi]$ and $(\hat{h}_T)' \in L^1[-\pi,\pi]$ as well as $(\hat{h}_T)' \in L^2[-\pi,\pi]$, because $\hat{h}_T$ is a polynomial in $\mathrm{e}^{-i\omega}$ and $\lim_{|\omega|\to\pi} \hat{h}_T(\omega) = 0$. $(\hat{h}_T)'$ denotes the derivative of $\hat{h}_T$. Let $f^{\check{}}$ denote the inverse Fourier transform of a function f. Then we have

$$((\hat{h}_T)')^{\check{}}(t) = -it h_T(t),$$

and it follows that

$$\begin{aligned} \int_{-\infty}^{\infty} t^2 |h_T(t)|^2 \,\mathrm{d}t &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |(\hat{h}_T)'(\omega)|^2 \,\mathrm{d}\omega \\ &= \|(\hat{h}_T)'\|_{L^2[-\pi,\pi]} \\ &< \infty, \end{aligned}$$

where we used Parseval's theorem in the fist equality. ■

On the integer lattice, i.e., for $t = n \in \mathbb{Z}$, we have no approximation problems, because, according to (9),

$$\begin{aligned} |(Tf)(n) - (T_\delta f)(n)| &\le \delta \sum_{k=-\infty}^{\infty} |h_T(n-k)| \\ &< \infty \end{aligned}$$

for every fixed $\delta > 0$. The last inequality follows from the fact that T is a FIR system.

Since

$$|(Tf)(n) - (T_\delta f)(n)| \le |(T_\delta f)(n)| + \|T\| \, \|f\|_{\mathcal{PW}^1_\pi}$$

by the same steps as in (6), it is interesting to know how large $\sup_{\|f\|_{\mathcal{PW}^1_\pi}\le 1} |(T_\delta f)(n)|$ can get because we can then upper bound the approximation error on the integer lattice. Let M be the smallest natural number such that $h_T(k) = 0$ for all $k > M$. We have

$$\begin{aligned} \sum_{k=0}^{M} |h_T(k)| &\le \left(\sum_{k=0}^{M} 1 \right)^{1/2} \left(\sum_{k=0}^{M} |h_T(k)|^2 \right)^{1/2} \\ &= \sqrt{M+1} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |\hat{h}_T(\omega)|^2 \,\mathrm{d}\omega \right)^{1/2} \\ &\le \sqrt{M+1} \, \|\hat{h}_T\|_{L^\infty[-\pi,\pi]} \\ &= \sqrt{M+1} \, \|T\|, \end{aligned}$$

which gives an upper bound for

$$\sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(n)|.$$

That $\sqrt{M+1}$ is indeed the rate of growth can be easily seen by considering quadratic phase functions or "chirp sequences" with

$$h_{T_\mathrm{q}}(k) = \begin{cases} \frac{1}{\sqrt{M+1}} \exp\left(i\frac{k^2\pi}{M+1}\right) & 0 \le k \le M \\ 0 & \text{otherwise} \end{cases}$$

and

$$\begin{aligned} \hat{h}_{T_\mathrm{q}}(\omega) &= \sum_{k=0}^{M} h_{T_\mathrm{q}}(k)\,\mathrm{e}^{-ik\omega} \\ &= \frac{1}{\sqrt{M+1}} \sum_{k=0}^{M} \exp\left(i\frac{k^2\pi}{M+1}\right)\mathrm{e}^{-ik\omega}\,. \end{aligned}$$

It can be shown [9] that there exists a constant C_1, which is independent from M, such that $\|\hat{h}_{T_\mathrm{q}}\|_{L^\infty[-\pi,\pi]} \le C_1$. Moreover, it follows that

$$\sum_{k=0}^{M} |h_{T_\mathrm{q}}(k)| = \frac{1}{\sqrt{M+1}} \sum_{k=0}^{M} 1 = \sqrt{M+1}, \qquad (16)$$

which shows that T_q is a worst-case FIR system that achieves the $\sqrt{M+1}$ growth.

Thus, on the integer lattice $t = n \in \mathbb{Z}$ the worst-case approximation error increases as $\sqrt{M+1}$, because from (16) and Lemma 1 we obtain

$$\sup_{T\in\mathcal{T}_M} \sup_{\|f\|_{\mathcal{PW}_\pi^1}\le 1} |(T_\delta f)(n)| \ge C_2\sqrt{M+1},$$

where C_2 is a universal constant and $\mathcal{T}_M$ denotes the set of all FIR systems T with $\|T\| \le 1$ and $h_T(k) = 0$ for all $k > M$.

VII. Discussion

We have seen that, for $f \in \mathcal{PW}_\pi^1$, the class of stable LTI systems T that can be uniformly approximated by TA_δ and TB_δ is given by the set of LTI systems with $h_T \in \mathcal{B}_\pi^1$. This means that the class of stable LTI systems that are robust under thresholding and quantization is exactly the class of bounded-input bounded-output (BIBO) stable LTI systems. Further, we discussed the consequences for FIR filters and showed that the worst-case approximation error increases as $\sqrt{M+1}$, where M denotes the filter length.

Appendix
Proof of $\|h^+(\,\cdot\,,0,N)\|_{\mathcal{PW}_\pi^1} < 3$

The Fourier coefficients $F_N(k)$, $k \in \mathbb{Z}$, of the Fejér kernel

$$\hat{F}_N(\omega) = \frac{1}{N+1}\frac{\sin^2((N+1)\frac{\omega}{2})}{\sin^2(\frac{\omega}{2})}$$

are given by

$$F_N(k) = \begin{cases} 1 - \frac{|k|}{N} & |k| < N \\ 0 & |k| \ge N. \end{cases}$$

Thus, $h^+(k,0,N) = 2F_{2N}(k) - F_N(k)$, $k \in \mathbb{Z}$ and the Fourier transform of

$$h^+(t,0,N) = \sum_{k=-2N+1}^{2N-1} h^+(k,0,N)\frac{\sin(\pi(t-k))}{\pi(t-k)}$$

is

$$\hat{h}^+(\omega,0,N) = 2\hat{F}_{2N}(\omega) - \hat{F}_N(\omega), \quad |\omega| \le \pi.$$

As a consequence we obtain

$$\begin{aligned} \|h^+(\,\cdot\,,0,N)\|_{\mathcal{PW}_\pi^1} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |2\hat{F}_{2N}(\omega) - \hat{F}_N(\omega)|\,\mathrm{d}\omega \\ &< \frac{1}{2\pi}\int_{-\pi}^{\pi} 2\hat{F}_{2N}(\omega) + \hat{F}_N(\omega)\,\mathrm{d}\omega = 3, \end{aligned}$$

In the last line we can write "$<$" instead of "$\le$" because $\hat{F}_{2N}$ and $\hat{F}_N$ are both non-negative.

Acknowledgment

This work was partly supported by the German Research Foundation (DFG) under grant BO 1734/9-1.

An extended version of this paper was accepted for publication in the IEEE Transactions on Signal Processing [10].

References

[1] C. E. Shannon, "Communication in the presence of noise," in *Proceedings of the IRE*, vol. 37, no. 1, January 1949, pp. 10–21.

[2] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, October 1998.

[3] I. Daubechies, R. A. DeVore, C. S. Güntürk, and V. A. Vaishampayan, "A/D conversion with imperfect quantizers," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 874–885, March 2006.

[4] H. Boche and U. J. Mönich, "General behavior of sampling-based signal and system representation," in *Proceedings of the 2008 IEEE International Symposium on Information Theory*, July 2008, pp. 2439–2443.

[5] M. K. Habib, "Digital representations of operators on band-limited random signals," *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 173–177, January 2001.

[6] H. Boche and U. J. Mönich, "Non-uniform sampling — signal and system representation," in *Proceedings of the 2008 International Symposium on Information Theory and its Applications (ISITA2008)*, December 2008, pp. 1576–1581.

[7] J. R. Higgins, *Sampling Theory in Fourier and Signal Analysis – Foundations*. Oxford University Press, 1996.

[8] Y. Katznelson, *An Introduction to Harmonic Analysis*. Cambridge University Press, 2004.

[9] H. Boche and V. Pohl, "Boundedness behavior of the spectral factorization for polynomial data in the Wiener algebra," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3100–3107, July 2008.

[10] H. Boche and U. J. Mönich, "Complete characterization of stable bandlimited systems under quantization and thresholding," *IEEE Transactions on Signal Processing*, to be published.