Higher order 1-bit Sigma-Delta modulation on a circle

Olga GRAF & Felix KRAHMER & Sara KRAUSE-SOLBERG
Technical University of Munich
Boltzmannstrasse 3, 85748 Garching, Germany
{olga.graf,felix.krahmer,sara.krause-solberg}@ma.tum.de

***Abstract*—Manifold models in data analysis and signal processing have become more prominent in recent years. In this paper, we will look at one of the main tasks of modern signal processing, namely, at analog-to-digital (A/D) conversion in connection with a simple manifold model (circle). We will focus on Sigma-Delta modulation which is a popular method for A/D conversion of bandlimited signals that employs coarse quantization coupled with oversampling. Classical Sigma-Delta schemes would provide mismatches and large errors at the initialization point if the signal to be converted is defined on a circle. In this paper, our goal is to get around these problems for higher order Sigma-Delta schemes. Our results show how to design an update for the second and third order schemes based on the reconstruction error analysis such that for the updated scheme the reconstruction error is improved.**

I. INTRODUCTION

A. Analog-to-digital conversion

Whenever data is processed using computers, analog-to-digital (A/D) conversion is used. It comprises two stages: sampling and quantization. As a result of the latter, each value of a discrete signal is mapped to some element from a finite alphabet. The extreme case when such alphabet consists of only two elements $\{-1, 1\}$ is called *1-bit quantization*.

Digital-to-analog (D/A) conversion, i.e., a process of reconstructing the original signal from quantized values, is usually carried out by applying an appropriate low-pass filter. The accuracy of such reconstruction serves as a main criteria for the quality assessment of the quantization scheme.

A popular method for the quantization of bandlimited functions is the so called *Sigma-Delta ($\Sigma\Delta$) modulation*. This method couples coarse quantization alphabet with substantial oversampling which, in its turn, makes the design of cheap analog circuits of low complexity possible. $\Sigma\Delta$ modulation has been known to circuit engineers since the 1963 pioneering work [1] of Inose and Yasuda; a rigorous mathematical study was initiated by Daubechies and DeVore in [2] in the early 2000's. Since then, the mathematical literature on this method of quantization has grown rapidly. Early papers on $\Sigma\Delta$ modulation focused on the reconstruction accuracy as a function of oversampling rate in the context of *bounded bandlimited functions on the real line* [3], [4], [5], [6]. Furthermore, $\Sigma\Delta$ modulation schemes were extended to *finite frame expansions* in [7], [8], [9]. Also, a number of papers provide results for $\Sigma\Delta$ modulation in *compressed sensing setting* [10], [11], [12].

This research was supported by the DFG Collaborative Research Center TRR109, Discretization in Geometry and Dynamics.

Despite the growing importance of *manifold models* in data and signal processing, there is still very little quantization literature available for such models, both for $\Sigma\Delta$ modulation and other quantization methods. For the special case of Grassmannian manifolds, which arise naturally in wireless communications, there appeared some papers studying their quantization properties [13]. Recently, the problem of recovering an unknown data point on a given manifold from 1-bit quantized random measurements was studied in [14]. What concerns $\Sigma\Delta$ modulation, the first attempt to generalize the results for bandlimited functions on the real line to those on the one-dimensional torus, was made by the authors in [15]. The main motivation behind this attempt was and continues to be the connection between $\Sigma\Delta$ and digital halftoning [16]. In order to pave the way to apply halftoning for priniting on closed surfaces, one first needs to understand the behavior of such approaches for the simplest closed one-dimensional manifold, the circle. This paper is a follow-up work where we extend our previous results to higher order $\Sigma\Delta$ schemes.

B. Our contribution

In [15], we focused on the first order $\Sigma\Delta$ scheme on a circle. In order to avoid mismatches in the scheme caused by this setting and to improve the reconstruction error, we proposed an update, namely, a small constant function shift by an appropriate amount.

In this paper, the contributions are two fold:

1) We extend the error analysis in [15] to the case of m-th order $\Sigma\Delta$ schemes (see Proposition 2).

2) Motivated by the error analysis, we show how to find an update (uniform or non-uniform shift) for the second and third order $\Sigma\Delta$ schemes such that the shifted function is recovered with the accuracy of $O(N^{-m})$, $m = 2, 3$ (see Theorems 1, 2).

II. PRELIMINARIES

A. Basics of $\Sigma\Delta$ modulation

Given a (finite or infinite) sequence of samples $(y_n)_{n\in\mathbb{Z}}$, the 1-bit $\Sigma\Delta$ quantizer runs the following iteration for $n \in \mathbb{Z}$:

$$\begin{aligned} v_n &= (h * v)_n + y_n - q_n, \\ q_n &= \operatorname{sign}((h * v)_n + y_n), \end{aligned} \tag{1}$$

where q_n are the quantized values, v_n are the state variables, h is the feedback filter described by the recurrence relation $(h*v)_n = \sum_{j=0}^{k} h_j v_{n-j}$, and $\operatorname{sign}(x)$ is the signum function.

Sometimes, it is useful to rewrite the first line in (1) in terms of another state variable u_n. If $y_n - q_n$ can be rewritten as an m-th order backward finite difference of some bounded sequence u_n, i.e.,

$$\Delta^m u_n = y_n - q_n, \tag{2}$$

then it is said that (1) is the *m-th order $\Sigma\Delta$ quantizer*.

The first order $\Sigma\Delta$ quantizer (i.e., with the feedback filter $h = (0,1)$) is known to be stable. In case of higher order schemes, the feedback filter should be chosen carefully, so that the stability is guaranteed. The following stability criterion provides a sufficient condition for the stability of a $\Sigma\Delta$ scheme (1).

Proposition 1 (Stability Criterion)**.** *[3], [17] Consider a $\Sigma\Delta$ quantizer given by the recurrence relation* (1)*. If*

$$\|h\|_{\ell^1} \le 2 - \|y\|_{\ell^\infty}, \tag{3}$$

then the modulator is stable for all inputs y_n.

Stability is a crucial property for $\Sigma\Delta$ modulation, and the reconstruction error analysis of $\Sigma\Delta$ modulation typically relies on it, namely, on the boundedness of the state variables v_n (or u_n) for given y_n and q_n.

A reconstruction error bound for 1-bit $\Sigma\Delta$ quantizer in [2] establishes a polynomial error decay of $O(\lambda^{-m})$, where λ is a factor, by which the function is oversampled with respect to its Nyquist frequency. Later, an exponential error decay of $O(2^{-r\lambda})$ with $r \approx 0.102$ was achieved in [3], [4] by combining $\Sigma\Delta$ schemes of different orders. This is the best known error decay rate, which is also known to be optimal [5], [6]. In this paper, we will follow the approach of [4], but only for fixed orders, hence expecting polynomial error bounds similar to those obtained in [2]. This will be sufficient for achieving our main goal: understanding how to update the $\Sigma\Delta$ modulation scheme to make it fit the circular framework better and assessing the improvement in accuracy of reconstruction after the update.

B. Analog-to-digital conversion on a circle

Let us briefly review sampling, quantization and reconstruction of K-bandlimited functions (i.e., $\operatorname{supp}(\widehat{f}) \subset [-K, K]$) $f \in L^2(\mathbb{T})$ whose domain is a unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. Uniformly sampling f on a unit circle at a rate $N/2\pi$ ($N \in \mathbb{N}$) gives the samples

$$y_n := f\left(2\pi n/N\right), \qquad n \in \{0, 1, ..., N-1\}. \tag{4}$$

Note that there is a one-to-one correspondence between the samples of 2π-periodic functions on the real line and the samples of functions on a unit circle. However, there will be no such correspondence between the quantized values.

Motivated by Shannon's sampling theorem, we define a reproducing kernel $\varphi^K \in L^2(\mathbb{T})$ such that

$$\widehat{\varphi}^K(\xi) = 1, \quad |\xi| \le K \quad \text{and} \quad \widehat{\varphi}^K(\xi) = 0, \quad |\xi| > K. \tag{5}$$

By the Fourier inversion theorem and some computations, we have

$$\varphi^K(x) = \sin\left((2K+1)\tfrac{x}{2}\right) / \sin\left(\tfrac{x}{2}\right). \tag{6}$$

In fact, (6) is the Dirichlet kernel whose convolution with any 2π-periodic function gives the n-th degree Fourier series approximation of that function. Therefore, (6) is indeed the reproducing kernel for K-bandlimited functions on $\mathbb{T}$.

Taking the Fourier series expansion of $\widehat{f}$ and using the inverse Fourier transform yields

$$f(t) = \frac{1}{N}\sum_{n=0}^{N-1} f\left(\frac{2\pi n}{N}\right)\varphi^K\left(t - \frac{2\pi n}{N}\right) \tag{7}$$

for any $N = 2\lambda K + 1$, where $\lambda \in \mathbb{R}_{\ge 1}$ is the oversampling parameter. Formula (7) is the analog of Shannon's interpolation formula for functions defined on a unit circle $\mathbb{T}$.

The standard approach to function recovery from its quantized values on $\mathbb{R}$ uses the Shannon's interpolation formula, where the samples are replaced by the quantized values. For the recovery from quantized values on $\mathbb{T}$ we have, consequently,

$$f_r(t) = \frac{1}{N}\sum_{n=0}^{N-1} q_n \varphi^K\left(t - \frac{2\pi n}{N}\right), \tag{8}$$

where $f_r(t)$ denotes the reconstructed function.

Let the instantaneous error $e(t)$ at time t be given by the pointwise difference

$$|f(t) - f_r(t)| = \frac{1}{N}\left|\sum_{n=0}^{N-1}(y_n - q_n)\,\varphi^K\left(t - \frac{2\pi n}{N}\right)\right|.$$

The quality of reconstruction can be measured by a variety of functional norms on $e(t)$. Here, we will use the norm $\|e\|_{L^\infty}$ which is one of the standard choices.

III. Error analysis for the second and higher order $\Sigma\Delta$ schemes on a circle

Error analysis for the first order $\Sigma\Delta$ scheme on a circle was given by the authors in [15]. Before extending it to the m-th order $\Sigma\Delta$ schemes, we will present the following two lemmas (the proofs will be omitted due to space limitations).

Lemma 1. *Denote $\varphi_n^K := \varphi^K\left(t - \frac{2\pi n}{N}\right)$. Then, for the sequences $(\varphi_n^K)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ we have*

$$\begin{aligned}\sum_{n=0}^{N-1}\Delta^m u_n \varphi_n^K &= (-1)^m \sum_{n=0}^{N-1} u_n \Delta^m \varphi_{n+m}^K \\ &+ \sum_{k=1}^{m}(-1)^k \Delta^{m-k} u_{-1}\Delta^{k-1}\varphi_{k-1}^K \\ &+ \sum_{k=1}^{m}(-1)^{k+1}\Delta^{m-k} u_{N-1}\Delta^{k-1}\varphi_{N+k-1}^K.\end{aligned}$$

Lemma 2. *Denote φ_n^K as in Lemma 1. Then, for the k-th order finite difference of φ_k^K we have for some $\tau \in \left(t - \frac{2\pi k}{N}, t\right)$*

$$\Delta^k \varphi_k^K = (-1)^k \left(\frac{2\pi}{N}\right)^k (\varphi^K)^{(k)}(\tau),$$

where $(\varphi^K)^{(k)}(\tau)$ is the k-th order derivative.

Now we can proceed with the error estimate.

Proposition 2 (Error analysis for m-th order $\Sigma\Delta$)**.** *Suppose $f \in L^2(\mathbb{T})$ and $\varphi^K \in L^2(\mathbb{T})$ satisfies* (5)*. Suppose $y_n :=$*

$f(2\pi n/N)$ *serves as an input to the* m*-th order* ($m \geq 2$) $\Sigma\Delta$ *quantizer given by* (1) *and* (2) *with the stability criterion* (3) *satisfied. Assume that the state variables are initialized as follows:* $u_{-1} = u_{-2} = ... = u_{-m} := 0$.

Then, for all $t \in \mathbb{T}$,

$$
\begin{aligned}
&|f(t) - f_r(t)| \\
&\leq \frac{(2\pi)^{m-1}\|u\|_{\ell^\infty}}{N^m}\left(\left\|(\varphi^K)^{(m)}\right\|_{L^1} + \left\|(\varphi^K)^{(m-1)}\right\|_{L^\infty}\right) \\
&+ \frac{1}{N}\sum_{k=1}^{m-1}\left(\frac{2\pi}{N}\right)^{k-1}\left\|(\varphi^K)^{(k-1)}\right\|_{L^\infty}|\Delta^{m-k}u_{N-1}|. \quad (9)
\end{aligned}
$$

Proof sketch. Due to the initialization of u_n and the fact that $\Delta^{k-1}\varphi_{N+k-1} = \Delta^{k-1}\varphi_{k-1}$ we can simplify the result in Lemma 1. Then, we have

$$
\begin{aligned}
&|f(t) - f_r(t)| \\
&= \frac{1}{N}\left|\sum_{n=0}^{N-1}(y_n - q_n)\varphi_n^K\right| = \frac{1}{N}\left|\sum_{n=0}^{N-1}\Delta^m u_n\varphi_n^K\right| \\
&\leq \frac{1}{N}\Big(\sum_{n=0}^{N-1}\left|(-1)^m u_n\Delta^m\varphi_{n+m}^K\right| \\
&+ \sum_{k=1}^{m}\left|(-1)^{k+1}\Delta^{m-k}u_{N-1}\Delta^{k-1}\varphi_{k-1}^K\right|\Big). \quad (10)
\end{aligned}
$$

Now rewrite the second sum in (10) as two summands by extracting the last term, apply Lemma 2 and estimate the result in terms of infinity norm. We get

$$
\begin{aligned}
&\sum_{k=1}^{m}\left|(-1)^{k+1}\Delta^{m-k}u_{N-1}\Delta^{k-1}\varphi_{k-1}^K\right| \\
&\leq \|u\|_{\ell^\infty}\left(\frac{2\pi}{N}\right)^m\left\|(\varphi^K)^{(m-1)}\right\|_{L^\infty} \\
&+ \sum_{k=1}^{m-1}\left(\frac{2\pi}{N}\right)^{k-1}\left\|(\varphi^K)^{(k-1)}\right\|_{L^\infty}|\Delta^{m-k}u_{N-1}|. \quad (11)
\end{aligned}
$$

Let us now evaluate the first sum in (10).

$$
\sum_{n=0}^{N-1}\left|(-1)^m u_n\Delta^m\varphi_{n+m}^K\right| \leq \|u\|_{\ell^\infty}\left(\frac{2\pi}{N}\right)^{m-1}\left\|(\varphi^K)^{(m)}\right\|_{L^1}.
$$

Inserting the above result and (11) into (10) completes the proof. □

Essentially, Proposition 2 tells us that whilst the first term of the error estimate (9) gives the error of $O(N^{-m})$ for the m-th order $\Sigma\Delta$ quantizer, then the second term always gives a sum of larger errors of order up to $O(N^{-1})$, hence increasing the order of the $\Sigma\Delta$ scheme without making any changes to the existing scheme becomes meaningless.

IV. Modification of the $\Sigma\Delta$ scheme

The question to address in this section is how to improve the reconstruction error (9) by modifying the m-th order $\Sigma\Delta$ modulation scheme described in Section II.A.

We propose to use the recurrence relation (1) once again on updated samples

$$
\tilde{y}_n := y_n + \delta_n, \quad (12)
$$

where some small δ_n is added at each iteration. We denote the resulting updated variables $\tilde{u}_n$ and $\tilde{q}_n$. We require that $\tilde{y}_n$ are the samples of a K-bandlimited function $\tilde{f}(t)$ and denote the error between this function and its reconstruction $\tilde{e}(t) := |\tilde{f}(t) - \tilde{f}_r(t)|$. In what follows, we will show that finding an appropriate sequence $(\delta_n)_{n\in\mathbb{N}}$ leads to the error $\tilde{e}(t)$ of $O(N^{-m})$ for the m-th order $\Sigma\Delta$ quantizer for $m = 2, 3$ and improves the distribution of the error $|f(t) - \tilde{f}_r(t)|$ around the circle.

Clearly, updates δ_n yield a certain new error $e_\delta(t) := |f(t) - \tilde{f}(t)|$. We will take some effort to analyze our setting and conclude that making the $\Sigma\Delta$ scheme consistent with the circular framework (which we haven't done yet, rather using the default scheme designed originally for the real line) will naturally require $e_\delta(t)$.

Let us write a lower bound estimate of $\left\|f - f_r\right\|_{L^\infty}$ using the fact that the error maximum is greater than its average.

$$
\begin{aligned}
&\left\|f - f_r\right\|_{L^\infty} \\
&\geq \frac{1}{N^2}\left|\sum_{k=0}^{N-1}\sum_{n=0}^{N-1}(y_n - q_n)\varphi^K\left(\frac{2\pi k}{N} - \frac{2\pi n}{N}\right)\right| \\
&= \frac{1}{N}\left|\sum_{n=0}^{N-1}(y_n - q_n)\frac{1}{N}\sum_{k=0}^{N-1}\varphi^K\left(\frac{2\pi k}{N} - \frac{2\pi n}{N}\right)\right| \\
&= \frac{1}{N}\left|\sum_{n=0}^{N-1}y_n - \sum_{n=0}^{N-1}q_n\right|. \quad (13)
\end{aligned}
$$

It is clear that $\sum_{n=0}^{N-1}y_n - \sum_{n=0}^{N-1}q_n$ is different from zero for an arbitrary function since $\sum_{n=0}^{N-1}y_n \in \mathbb{R}$ and $\sum_{n=0}^{N-1}q_n \in \mathbb{Z}$. Using the initialization $u_{-1} = u_{-2} = ... = u_{-m} := 0$ and computing the telescoping sum, we have $\sum_{n=0}^{N-1}\Delta^m u_n = \Delta^{m-1}u_{N-1}$ which together with (2) yields

$$
\sum_{n=0}^{N-1}y_n - \sum_{n=0}^{N-1}q_n = \Delta^{m-1}u_{N-1}. \quad (14)
$$

Therefore, for the lower bound to be greater than or equal to zero in (13), we need a vector of updates $(\delta_n)_{n=0}^{N-1}$ whose entries sum up to $\Delta^{m-1}u_{N-1}$.

Although the error $e_\delta(t)$ cannot be avoided in general, we can argue that this is a reasonable trade-off. Firstly, in some applications the update could be subtracted later from the reconstructed function. Secondly, in the applications where this is not the case (e.g., digital halftoning), an error caused by a well-designed update (which is constant or smoothly varying) can be less audible/visible than highly oscillating initial error.

A. Second order $\Sigma\Delta$ scheme

It was shown in [15] that for the first order scheme it is sufficient to choose a constant update $\delta = -N^{-1}v_{N-1}$ in order to eliminate the boundary term of summation by parts in the error estimate. From (9) we see that in the second order case there is only one summand larger than of $O(N^{-2})$ and it involves the remainder $|\Delta u_{N-1}|$. Therefore, we will choose a constant update similar to that in [15], namely,

$$
\delta = -N^{-1}\Delta u_{N-1}. \quad (15)
$$

The following theorem ensures that this approach is valid.

Theorem 1 (Uniform update for 2-nd order $\Sigma\Delta$)**.** *Let the assumptions in Proposition 2 hold and assume the order of* $\Sigma\Delta$ *quantizer* $m = 2$*. Then, using in* (1) *the updated samples* (12) *with* δ *given by* (15) *leads to* $\Delta\tilde{u}_{N-1} = 0$ *and the error* $\tilde{e}(t) := |\tilde{f}(t) - \tilde{f}_r(t)|$ *is of* $O(N^{-2})$.

Proof sketch. The proof follows from the main result in [15] by defining $\omega_n := \Delta u_n$ and working with the first order relation

$\Delta\omega_n = y_n - q_n$. The error estimate is obtained by rewriting (9) with updated variables. $\square$

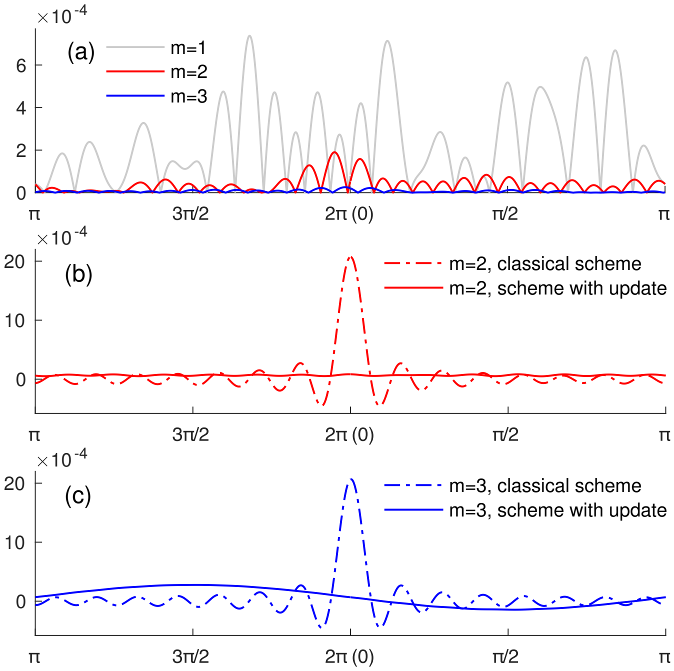

Fig. 1. Reconstruction errors for classical and updated $\Sigma\Delta$ schemes on a circle. In all experiments we use $f(t) = 0.1\sin(5t)\cos(10t) + 0.2$ with bandwidth $K = 15$, $N = 3002$ (in (a)) or $N = 6002$ (in (b), (c)) and feedback filters $(0,1)$, $(0,4/3,0,0,-1/3)$, $(0,91/72,0,0,0,0,0,-13/36,0,0,0,0,0,7/72)$ for the 1-st, 2-nd and 3-rd order schemes, respectively. (a) Comparison of errors $\tilde{e}(t) := |\tilde{f}(t) - \tilde{f}_r(t)|$ for $\Sigma\Delta$ schemes of order $m = 1,2,3$. (b) Differences $f(t) - f_r(t)$, $f(t) - \tilde{f}_r(t)$ for the 2-nd order $\Sigma\Delta$ scheme (cf. Theorem 1). (c) Differences $f(t) - f_r(t)$, $f(t) - \tilde{f}_r(t)$ for the 3-rd order $\Sigma\Delta$ scheme (cf. Theorem 2).

B. *Third order $\Sigma\Delta$ scheme*

The third order case is still an object of further research, but we would like to share our partial results. We have two unwanted large terms in (9) involving the remainders $|\Delta^2 u_{N-1}|$ and $|\Delta u_{N-1}|$. Here, we cannot achieve $\tilde{e}(t)$ of $O(N^{-3})$ by a constant update. In order to find a suitable sequence of updates $(\delta_n)_{n\in\mathbb{N}}$, we first examine how the updates propagate with each iteration. For the third order $\Sigma\Delta$ quantizer without the updates we can set up a system of equations

$$\begin{pmatrix} u_n \\ u_{n-1} \\ u_{n-2} \end{pmatrix} = \begin{pmatrix} 3 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} u_{n-1} \\ u_{n-2} \\ u_{n-3} \end{pmatrix} + \begin{pmatrix} y_n - q_n \\ 0 \\ 0 \end{pmatrix} = M\mathbf{u}_{n-1} + (y_n - q_n)\mathbf{e}_1. \tag{16}$$

For the updated quantization scheme, we have

$$\begin{pmatrix} \tilde{u}_{N-1} \\ \tilde{u}_{N-2} \\ \tilde{u}_{N-3} \end{pmatrix} = M^N \mathbf{u}_{-1} + \sum_{n=0}^{N-1} M^{N-1-n}(y_n + \delta_n - \tilde{q}_n)\mathbf{e}_1 = \sum_{n=0}^{N-1} M^{N-1-n}(q_n + \delta_n - \tilde{q}_n)\mathbf{e}_1 + \mathbf{u}_{N-1}. \tag{17}$$

It is empirically satisfied in many cases that the sum of the propagation of $(q_n - \tilde{q}_n)$ is equal to 0. In the first step we assume the validity of the latter. Next, we will compute explicitly the propagation of δ_n using the Jordan decomposition $M = VJV^{-1}$. Omitting some simplification steps, we get

$$M^{N-1-n}\delta_n\mathbf{e}_1 = VJ^{N-1-n}V^{-1}\delta_n\mathbf{e}_1 = \delta_n VJ^{N-1-n}\mathbf{e}_3 = \frac{\delta_n}{2}\begin{pmatrix} (N-n+1)(N-n) \\ (N-n)(N-n-1) \\ (N-n-2)(N-n-1) \end{pmatrix}. \tag{18}$$

Analyzing various sequences $(\delta_n)_{n\in\mathbb{N}}$ is out of the scope of this paper. Here, we will present a certain update which achieves our goal of having $\tilde{e}(t)$ of $O(N^{-3})$ for the third order $\Sigma\Delta$ quantizer.

Theorem 2 (Harmonic update for 3-rd order $\Sigma\Delta$)**.** *Let the assumptions in Proposition 2 hold and assume the order of $\Sigma\Delta$ quantizer $m = 3$. Set $\delta_n = \delta^1 w_n^1 + \delta^2 w_n^2$ where $\delta^1, \delta^2 \in \mathbb{R}$ are the unknown scalars and $w^1 = (1,\dots,1) \in \mathbb{R}^N$, $w^2 = \left(\sin(\frac{2\pi n}{N})\right)_{n=0}^{N-1}$ are vectors. Assume in (17) $\sum_{n=0}^{N-1} M^{N-1-n}(q_n - \tilde{q}_n)\mathbf{e}_1 = 0$. The unknowns can be obtained from*

$$\begin{pmatrix} \delta^1 \\ \delta^2 \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{N} \\ -\frac{2}{N}\tan(\frac{\pi}{N}) & (1+\frac{1}{N})\tan(\frac{\pi}{N}) \end{pmatrix} \begin{pmatrix} \Delta u_{N-1} \\ \Delta^2 u_{N-1} \end{pmatrix}. \tag{19}$$

Then, using in (1) *the updated samples* (12) *with δ_n obtained from* (19) *leads to $\Delta^2\tilde{u}_{N-1} = 0$ and $\Delta\tilde{u}_{N-1} = 0$. The error $\tilde{e}(t) := |\tilde{f}(t) - \tilde{f}_r(t)|$ is of $O(N^{-3})$.*

Proof sketch. For given δ_n, we carry out computations in (18) further and get

$$\sum_{n=0}^{N-1} M^{N-1-n}\delta_n\mathbf{e}_1 = \frac{1}{2}\begin{pmatrix} \frac{N(N^2+3N+2)}{3} & \frac{N(N+1)}{2}\cot(\frac{\pi}{N}) \\ \frac{N(N^2-1)}{3} & \frac{N(N-1)}{2}\cot(\frac{\pi}{N}) \\ \frac{N(N^2-3N+2)}{3} & \frac{N(N-3)}{2}\cot(\frac{\pi}{N}) \end{pmatrix} \begin{pmatrix} \delta^1 \\ \delta^2 \end{pmatrix}.$$

Now insert the above result in (17) and multiply from the left by $\left(\begin{smallmatrix} 1 & -1 & 0 \\ 1 & -2 & 1 \end{smallmatrix}\right)$. Now we require that the resulting left-hand side $(\Delta\tilde{u}_{N-1}, \Delta^2\tilde{u}_{N-1})^T = 0$. This gives the condition

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{N(N+1)}{2} & \frac{N}{2}\cot(\frac{\pi}{N}) \\ N & 0 \end{pmatrix} \begin{pmatrix} \delta^1 \\ \delta^2 \end{pmatrix} + \begin{pmatrix} \Delta u_{N-1} \\ \Delta^2 u_{N-1} \end{pmatrix}. \tag{20}$$

Solving the above equation for $(\delta^1, \delta^2)^T$ gives the result in (19). The error estimate is obtained by rewriting (9) with updated variables. $\square$

Remark. It is easy to see from (19) that $\delta^1 = O(N^{-1})$ and $\delta^2 = O(\tan(\frac{\pi}{N}))$. It is an interesting open question whether it is possible to design $\delta_n = \delta^1 w_n^1 + \delta^2 w_n^2$ in such a way that δ^2 is of smaller order. This would not reduce the overall contribution of $O(N^{-1})$ to the error made by δ_n, however, it might be preferable in applications.

We complement theoretical results with numerical demonstration in Fig. 1.

V. Conclusions and Future Work

In this paper, we justified the necessity of updating the classical $\Sigma\Delta$ modulation scheme if the function to be quantized is defined on a circle. We proposed the sequence of updates for the second and third order $\Sigma\Delta$ schemes and complemented our results with the reconstruction error analysis. For the schemes of order $m \geq 3$, designing the sequence of updates becomes less trivial. The important open question is how to find an optimal update for a $\Sigma\Delta$ scheme of arbitrary order. Furthermore, it would be of interest to extend the results to more complicated manifold models.

References

[1] H. Inose and Y. Yasuda, "A unity bit coding method by negative feedback," in *Proc. IEEE*, 1963, vol. 51(11), pp. 1524–1535.

[2] I. Daubechies and R. DeVore, "Approximating a bandlimited function using very coarsely quantized data: a family of stable Sigma-Delta modulators of arbitrary order," *Ann. Math.*, vol. 158(2), pp. 679–710, 2003.

[3] C. Güntürk, "One-bit Sigma-Delta quantization with exponential accuracy," *Comm. Pure Appl. Math.*, vol. 56(11), pp. 1608–1630, 2003.

[4] P. Deift, C. S. Güntürk, and F. Krahmer, "An optimal family of exponentially accurate one-bit Sigma-Delta quantization schemes," *Comm. Pure Appl. Math.*, vol. 64(7), pp. 883–919, 2011.

[5] F. Krahmer and R. Ward, "Lower bounds for the error decay incurred by coarse quantization schemes," *Appl. Comput. Harmonic Anal.*, vol. 32(1), pp. 131–138, 2012.

[6] A. R. Calderbank and I. Daubechies, "The pros and cons of democracy," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1721–1725, 2002.

[7] F. Krahmer, R. Saab, and R. Ward, "Root-exponential accuracy for coarse quantization of finite frame expansions," *IEEE Trans. Inform. Theory*, vol. 58(2), pp. 1069–1079, 2012.

[8] J. Benedetto, A. Powell, and Ö. Yilmaz, "Sigma-Delta ($\sigma\delta$) quantization and finite frames," *IEEE Trans. Inform. Theory*, vol. 52(5), pp. 1990–2005, 2006.

[9] J. J. Benedetto, A. M. Powell, and Ö. Yilmaz, "Second-order Sigma-Delta ($\sigma\delta$) quantization of finite frame expansions," *Appl. Comput. Harmonic Anal.*, vol. 20(1), pp. 126–148, 2006.

[10] F. Krahmer, R. Saab, and Ö. Yilmaz, "Sigma-Delta quantization of sub-gaussian frame expansions and its application to compressed sensing," *Inform. Inference*, vol. 3(1), pp. 50–58, 2014.

[11] C. Güntürk, M. Lammers, A. Powell, R. Saab, and Ö. Yilmaz, "Sobolev duals for random frames and $\sigma\delta$ quantization of compressed sensing measurements," *Found. Comp. Math.*, vol. 13(1), pp. 1–36, 2013.

[12] R. Saab, R. Wang, and Ö. Yilmaz, "Quantization of compressive samples with stable and robust recovery," *Appl. Comput. Harmonic Anal.*, vol. 44(1), pp. 123–143, 2018.

[13] B. Mondal, S. Dutta, and R. W. Heath Jr., "Quantization on the Grassmann manifold," *IEEE Trans. Signal Proc.*, vol. 55(8), pp. 4208–4216, 2007.

[14] M. Iwen, F. Krahmer, S. Krause-Solberg, and J. Maly, "On recovery guarantees for one-bit compressed sensing on manifolds," *Preprint*, 2018.

[15] S. Krause-Solberg, O. Graf, and F. Krahmer, "One-bit Sigma-Delta modulation on a closed loop," in *Proc. IEEE Stat. Signal Proc. Workshop (SSP '18)*, 2018.

[16] T. Kite, B. Evans, A. Bovik, and T. Sculley, "Digital halftoning as 2-D Delta-Sigma modulation," in *Proc. Intl. Conf. Image Proc.*, 1997, vol. 1, pp. 799–802.

[17] R. Schreier and M. Snelgrove, "Stability in a general Sigma-Delta modulator," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Proc.*, 1991, pp. 1769–1772.