On coarse quantization of tight Gabor frame expansions

Özgür Yılmaz
Department of Mathematics
University of Maryland
College Park, MD 20742
E-mail: oyilmaz@math.umd.edu

09-17-2003

Abstract

This paper presents a coarse quantization algorithm (TF$\Sigma\Delta$-II) for tight Gabor frame expansions of certain functions in $L^2(\mathbb{R})$, an alternative to the TF$\Sigma\Delta$ of [10]. Using some a priori information about the function to be quantized, TF$\Sigma\Delta$-II produces an approximation in $L^2(\mathbb{R})$. For a tight Gabor frame with frame bound A, we prove that the approximation error corresponding to a kth-order TF$\Sigma\Delta$-II quantizer is of order $O(A^{-k})$. Furthermore, motivated by TF$\Sigma\Delta$-II, we construct an algorithm to coarsely quantize the Fourier coefficients of certain compactly supported functions.

1 Introduction

In the processing of audio signals, images, and video, one of the most important objectives is to obtain a digital representation of the signal of interest that is suitable for storage, transmission, and recovery. Since these signals are inherently analog, i.e., they have a continuous domain and a continuous range, this is usually done in two steps. The first step is finding a sequence that completely determines the original signal. In general this sequence is real or complex valued. Therefore a second step is needed to reduce the continuous range of this sequence to a discrete, possibly finite set. This second step is called *quantization*.

Given a signal f, obtaining the sequence $\{f_n\}_{n\in\mathbb{Z}}$ that completely determines f usually corresponds to expanding the original function f over some dictionary $\{\varphi_n\}_{n\in\mathbb{Z}}$, i.e.,

$$f = \sum_n f_n \varphi_n. \tag{1}$$

Such a dictionary is said to be *redundant* if the choice of $\{f_n\}_{n\in\mathbb{Z}}$ in (1) is not unique.

In various applications it is convenient to assume that the signals of interest are elements of some Hilbert space, e.g., bandlimited functions, $L^2(\mathbb{R})$, etc.. In this case, we can consider more structured dictionaries such as *frames*. A collection $\{\varphi_n\}$ in a Hilbert space H is said to be a *frame* if for each $f \in H$ we have

$$A\|f\|^2 \le \sum |\langle f, \varphi_n\rangle|^2 \le B\|f\|^2,$$

where the frame bounds $A > 0$ and $B < \infty$ are independent from f. The frame is called *tight* if $A = B$, and we can show that in this case, (1) holds for any f with $f_n := A^{-1}\langle f, \varphi_n\rangle$. An important remark is that the frame bound A of a tight frame with $\|\varphi_n\| = 1$ for all n "measures" the redundancy of the system. A=1 implies that $\{\varphi_n\}$ constitutes an orthonormal basis; the larger the frame bound $A > 1$ is, the more redundant is the frame.

Once an expansion of the form (1) is obtained, the next goal is to quantize the coefficients f_n. There are two different approaches to accomplish this: "fine quantization" and "coarse quantization". Given an expansion as in (1), one way to quantize the coefficients is replacing the real numbers f_n with $q_n = \delta$ round (f_n/δ); if the coefficients are complex, then we replace the real and imaginary parts separately. Here δ is the *step size* of the quantizer. Note that by decreasing the step size, and thus increasing the resolution of the quantizer, we can make $|f_n - q_n|$ arbitrarily small. In other words, the approximation error $f - \tilde{f}_\delta$ diminishes as δ approaches 0, where $\tilde{f}_\delta$ is defined by replacing f_n in (1) with q_n produced by a quantizer with step size δ. Such algorithms are usually called *fine quantization algorithms.*

A second approach exists if the expansion of the original function f is highly redundant. In this case, one can replace the coefficients f_n with "coarsely quantized" values q_n, and still have a good approximation. Instead of controlling the individual differences $|f_n - q_n|$, such an algorithm aims to produce q_n so that the reconstruction error $\|f - \tilde{f}\|$ is small. In particular, if we consider tight frame expansions, the algorithm is constructed such that the approximation error $f - \tilde{f}_A$ gets arbitrarily small as we increase the frame bound A. Such algorithms are called *coarse quantization algorithms.* Note that a coarse quantization algorithm exploits the redundancy of the expansion to compensate for the coarseness of the quantization.

In this paper we consider tight Gabor frame expansions of certain classes of functions in $L^2(\mathbb{R})$. Gabor frames are frames of $L^2(\mathbb{R})$ that are generated by shifting a fixed function $\varphi \in L^2(\mathbb{R})$ along a lattice $\Gamma = \tau_0\mathbb{Z} \times \xi_0\mathbb{Z}$ in the time-frequency plane: For $\varphi_{n,m}(t) := \varphi(t - n\tau_0)e^{im\xi_0 t}$, $\{\varphi_{n,m} : n, m \in \mathbb{Z}\}$ constitutes a frame in $L^2(\mathbb{R})$. Detailed discussions of Gabor frames can be found in [2, 5, 6].

For simplicity, we use the ordered triple (φ, τ_0, ξ_0), i.e., (φ, τ_0, ξ_0) to refer to the Gabor frame $\{\varphi_{n,m} : n, m \in \mathbb{Z}\}$ with $\varphi_{n,m}(t) = \varphi(t - n\tau_0)e^{im\xi_0 t}$. Suppose (φ, τ_0, ξ_0) is a tight Gabor frame with the frame bound A. Then for $f \in L^2(\mathbb{R})$, we have

$$f = \frac{1}{A} \sum \langle f, \varphi_{n,m} \rangle \varphi_{n,m}, \tag{2}$$

where equality is in the sense of L^2. Recall that the frame bound "measures" the redundancy of the frame expansion when the frame is tight and normalized such that each element has norm 1. For a tight Gabor frame (φ, τ_0, ξ_0), it is a standard result that the frame bound is given by $A = \frac{2\pi}{\tau_0\xi_0}$ as long as $\|\varphi\| = 1$ [4]. Therefore, if τ_0 and ξ_0 are small (so that $A >> 1$), the expansion is heavily redundant, thus it should be possible to coarsely quantize the frame coefficients of a given function and still have a good approximation. Indeed, such an algorithm, called TFΣΔ, was introduced in [10]. TFΣΔ produces weak type approximations of functions that are in the unit ball of the modulation space M^∞. In this paper, we will present an alternative coarse quantization algorithm, which we call TFΣΔ-II, for the tight Gabor frame expansions of those functions in the unit ball of the modulation space $M^\infty_{1/G}$, where $G : \mathbb{R}^2 \mapsto \mathbb{R}$ is a fixed "envelope", for which (2) converges almost everywhere (as well as in L^2). In this case, TFΣΔ-II produces approximations in $L^2(\mathbb{R})$.

The TFΣΔ-II algorithm (as well as the TFΣΔ of [10]) is inspired by sigma-delta quantization algorithms that are commonly used to coarsely quantize oversampled bandlimited functions [11]. Given a function f which is Ω-bandlimited, i.e., the Fourier transform $\hat{f}$ of f vanishes outside the interval $[-\Omega, \Omega]$, it is well known that

$$f(t) = \frac{1}{\lambda} \sum_n f(\frac{n\pi}{\lambda\Omega})\varphi(\frac{\Omega}{\pi}t - \frac{n}{\lambda}) \tag{3}$$

for an appropriately chosen function φ and for $\lambda > 1$, e.g., [3]. For such expansions the family of sigma-delta (ΣΔ) modulators provide efficient quantization algorithms. Given a function f in the unit ball of L^∞, a kth-order sigma-delta modulator yields a sequence $\{q_n^\lambda\}_{n\in\mathbb{Z}}$ with $q_n \in \{-1, 1\}$; this sequence has the additional property that its kth-order running sums track the kth-order running sums of $f_n^\lambda := f(\frac{n\pi}{\lambda\Omega})$, uniformly. More precisely,

$$|\sum_{m_{k-1}=N_k}^{M_k} \cdots \sum_{m_1=N_2}^{m_2} \sum_{n=N_1}^{m_1} f_n^\lambda - \sum_{m_{k-1}=N_k}^{M_k} \cdots \sum_{m_1=N_2}^{m_2} \sum_{n=N_1}^{m_1} q_n^\lambda| < C \tag{4}$$

where the value of the constant C does not depend on $N_1, \ldots, N_k$, M_k or $\{f_n^\lambda\}_{n\in\mathbb{Z}}$. For example, a first-order sigma-delta quantizer generates the q_n^λ, which satisfy (4) for $k = 1$, via the following recursion:

$$\begin{aligned} v_n - v_{n-1} &= f_n^\lambda - q_n^\lambda \\ q_n^\lambda &= sign(v_{n-1} + f_n^\lambda). \end{aligned} \tag{5}$$

In this case, one can show that [3]

- $|v_n| < 1 \quad \text{for all } n, \text{ if } v_0 \in (-1, 1)$ (6)
- $\|f - \tilde{f}_\lambda\|_{L^\infty} \leq \frac{1}{\lambda}\|g'\|_{L^1}, \text{ if } \text{supp}\hat{f} \subseteq [-\pi, \pi],$ (7)

where $\tilde{f}_\lambda := \lambda^{-1}\sum_n q_n^\lambda \varphi(\cdot - n/\lambda)$. In fact, this bound can be improved; [7] contains a proof that the error can be bounded pointwise by $C\lambda^{-4/3+\eta}$ where C depends on η and on the value of the derivative of the original function at the corresponding point. Moreover [8] shows that the mean-square error is of order $O(\lambda^{-3})$ in the case of a constant input. Finally, [1] shows that the mean-square error is of order $O(\lambda^{-3})$ for bandlimited functions satisfying certain mild conditions.

A k^{th}-order sigma-delta quantizer is defined by replacing the first-order backward difference operator in (5) by a k^{th}-order backward difference operator and adjusting the rule that determines q_n such that the $|v_n|$ stay uniformly bounded. In this case, one can prove that the L^∞ approximation error is of order $O(\lambda^{-k})$. Detailed discussions of higher-order schemes can be found in [3, 9].

In Section 2 we introduce the first-order TFΣΔ-II algorithm, a coarse quantization algorithm for certain functions in $L^2(\mathbb{R})$; this algorithm assumes a priori information about the function to be quantized; however, unlike the TFΣΔ of [10], the approximation is in $L^2(\mathbb{R})$. In Section 2.2 we define higher-order TFΣΔ-II algorithms and prove that for a k^{th}-order scheme the L^2-approximation error is $O(A^{-k})$ where A is the corresponding frame bound. We present numerical experiments for the first and second-order TFΣΔ-II in Section 2.3. Finally, in Section 3, we show that a scheme to coarsely quantize the Fourier coefficients of certain compactly supported functions can be constructed using the same strategy as in the construction of TFΣΔ-II.

2 The Time-Frequency Sigma-Delta Quantization Algorithm II (TFΣΔ-II)

In this section we will introduce an algorithm (TFΣΔ-II) to quantize the frame coefficients of certain functions in $L^2(\mathbb{R})$. This algorithm will produce an approximation to the original function in L^2 and we will prove that the corresponding L^2-approximation error is of order $O(A^{-k})$ for a k^{th}-order scheme. Note that these results are much stronger than the analogous results that were presented in [10] although the TFΣΔ-II algorithm is **not** translation invariant, unlike the TFΣΔ of [10]. Furthermore, the class of functions which can be quantized using TFΣΔ-II is smaller than the class of functions that can be quantized using TFΣΔ.

2.1 The First-Order TFΣΔ-II Algorithm

Let G be a fixed function in $L^2(\mathbb{R}^2)$ that is smooth with nice decay. Let (φ, τ_0, ξ_0) be a Gabor frame and denote by $\mathcal{B}_G^\varphi$ the collection of functions in $L^2(\mathbb{R})$ which satisfy

$$|\langle f, \varphi_{\tau,\xi}\rangle| < G(\tau, \xi), \tag{8}$$

where $\varphi_{\tau,\xi} = \varphi(t-\tau)e^{i\xi t}$. Note that when φ is a suitable function, $\mathcal{B}_G^{\varphi}$ coincides with the unit ball of the modulation space $M_{1/G}^{\infty}$ equipped with the norm $\|\cdot\|_{M_{1/G}^{\infty}} := \|V_{\varphi}(\cdot)G\|_{L^{\infty}}$. Here $V_{\varphi}f := \langle f, \varphi_{\tau,\xi}\rangle$.

Let f be in $\mathcal{B}_G^{\varphi}$. Denote the frame coefficients of f, $\langle f, \varphi_{n,m}\rangle$, by $c_{n,m}$; define $c_{n,m}^R$ and $c_{n,m}^I$ as the real and imaginary parts of $c_{n,m}$ respectively. Since $c_{n,m} = V_{\varphi}f(n\tau_0, m\xi_0)$ and since $f \in \mathcal{B}_G^{\varphi}$, we clearly have

$$\begin{aligned} |c_{n,m}^R| &< G(n\tau_0, m\xi_0), \quad \text{and} \\ |c_{n,m}^I| &< G(n\tau_0, m\xi_0). \end{aligned}$$

In other words, if we define $\tilde{c}_{n,m}^R := \frac{c_{n,m}^R}{G(n\tau_0, m\xi_0)}$ and $\tilde{c}_{n,m}^I := \frac{c_{n,m}^I}{G(n\tau_0, m\xi_0)}$, both $|\tilde{c}_{n,m}^R|$ and $|\tilde{c}_{n,m}^I|$ will be bounded by 1.

The algorithms that we consider in this paper quantize the real and imaginary parts of the frame coefficients in an identical manner. Therefore, to simplify notation, we will use the superscript S whenever an expression, an equation or a system of equations is valid for both $S = '' R''$ and $S = '' I''$. Now, define the first-order TF$\Sigma\Delta$-II scheme as follows:

$$\begin{aligned} u_{n,m}^S - u_{n-1,m}^S &= \tilde{c}_{n,m}^S - p_{n,m}^S \\ p_{n,m}^S &= \text{sign}(u_{n-1,m}^S + c_{n,m}^S) \\[1em] v_{n,m}^S - v_{n,m-1}^S &= u_{n,m}^S - r_{n,m}^S \\ r_{n,m}^S &= \text{sign}(v_{n,m-1}^S + u_{n,m}^S). \end{aligned} \tag{9}$$

Note that the scheme described in (9) is stable, i.e. the internal state variables, u^S, v^S, stay uniformly bounded: Since $|\tilde{c}_{n,m}^S|$ is bounded by 1, $|u_{n,m}^S|$ is bounded by 1 via (6); but this implies that $v_{n,m}^S$ is also bounded by 1, again because of (6). Clearly,

$$(\Delta_1\Delta_2 v^R)_{n,m} = \tilde{c}_{n,m}^R - (p_{n,m}^R + \Delta_1 r_{n,m}^R), \tag{10}$$

and similarly

$$(\Delta_1\Delta_2 v^I)_{n,m} = \tilde{c}_{n,m}^I - (p_{n,m}^I + \Delta_1 r_{n,m}^I); \tag{11}$$

thus we have

$$(\Delta_1\Delta_2 v)_{n,m} = \tilde{c}_{n,m} - (p_{n,m} + \Delta_1 r_{n,m}), \tag{12}$$

where $v_{n,m} = v_{n,m}^R + iv_{n,m}^I$, $p_{n,m} = p_{n,m}^R + ip_{n,m}^I$, $r_{n,m} = r_{n,m}^R + ir_{n,m}^I$, and $\tilde{c}_{n,m} := \frac{c_{n,m}}{G(n\tau_0, m\xi_0)}$. Multiplying both sides of (12) by $G(n\tau_0, m\xi_0)$ yields

$$G(n\tau_0, m\xi_0)(\Delta_1\Delta_2 v)_{n,m} = c_{n,m} - G(n\tau_0, m\xi_0)(p_{n,m} + \Delta_1 r_{n,m}). \tag{13}$$

This suggests us to define T_{TFII} as the mapping that maps the sequence $c = (c_{n,m})$ to the sequence $q = (q_{n,m})$ with

$$q_{n,m} = G(n\tau_0, m\xi_0)(p_{n,m} + \Delta_1 r_{n,m}). \tag{14}$$

Theorem 1. *Let (φ, τ_0, ξ_0) be a tight Gabor frame of $L^2(\mathbb{R})$ with frame bound A. Let $f \in \mathcal{B}_G^{\varphi}$, $c = (\langle f, \varphi_{n,m}\rangle)$, $q = T_{TFII}(c)$, and define $\tilde{f}_A := A^{-1}\sum_{n,m} q_{n,m}\varphi_{n,m}$. Suppose φ and G satisfy i-iv:*

i. $|\varphi| \star G_1 \in L^2$ where $\star$ stands for convolution and $G_1(s) = \int |\partial_1\partial_2 G(s,z)|dz$.

ii. $|\varphi'| \star G_2 \in L^2$ where $G_2(s) = \int |\partial_2 G(s,z)|dz$.

iii. $t(|\varphi| \star G_3)(t) \in L^2$ where $G_3(s) = \int |\partial_1 G(s,z)|dz$, and *iv.* $t(|\varphi'| \star G_4)(t) \in L^2$ *where* $G_4(s) = \int |G(s,z)|dz$.

where $\partial_i G$ *is the* i*th partial derivative of* G. *Then the* L^2*-approximation error satisfies*

$$\|f - \tilde{f}_A\|_{L^2} \leq \frac{C}{A}, \tag{15}$$

with $C := \| \, |\varphi| \star G_1\|_{L^2} + \| \, |\varphi'| \star G_2\|_{L^2} + \|t(|\varphi| \star G_3)(t)\|_{L^2} + \|t(|\varphi'| \star G_4)(t)\|_{L^2}$.

Proof. We start by writing the pointwise error, which is given by

$$f(t) - \tilde{f}_A(t) = \frac{1}{A}\sum_{n,m}(c_{n,m} - q_{n,m})\varphi_{n,m}(t) \tag{16}$$

for almost every t. Then,

$$\begin{aligned} f(t) - \tilde{f}_A(t) &= \frac{1}{A}\sum_{n,m} G(n\tau_0, m\xi_0)(\Delta_1\Delta_2 v^I)_{n,m}\varphi_{n,m}(t) && (17)\\ &= \frac{1}{A}\sum_{n,m} v_{n,m}\bar{\Delta}_2\bar{\Delta}_1 G(n\tau_0, m\xi_0)\varphi_{n,m}(t), && (18)\end{aligned}$$

where the first equality is due to (13); the second equality is the result of summation by parts. Note that the boundary terms disappear since G has nice decay in both its arguments. Now, denote $\bar{\Delta}_2\bar{\Delta}_1 G(n\tau_0, m\xi_0)\varphi_{n,m}$ by $I_{n,m}$. Then

$$I_{n,m}(t) = \bar{\Delta}_2\bar{\Delta}_1\Gamma(n\tau_0, m\xi_0, t), \tag{19}$$

where

$$\Gamma(s,z,t) = G(s,z)\varphi(t-s)e^{izt}. \tag{20}$$

Let us rewrite $I_{n,m}$ as

$$\begin{aligned} I_{n,m}(t) &= \bar{\Delta}_2(\Gamma(n\tau_0, m\xi_0, t) - \Gamma((n+1)\tau_0, m\xi_0, t)) && (21)\\ &= \bar{\Delta}_2\int_{(n+1)\tau_0}^{n\tau_0} \partial_1\Gamma(s, m\xi_0, t)ds && (22)\\ &= \int_{(n+1)\tau_0}^{n\tau_0} (\partial_1\Gamma(s, m\xi_0, t) - \partial_1\Gamma(s, (m+1)\xi_0, t))ds && (23)\\ &= \int_{(n+1)\tau_0}^{n\tau_0}\int_{(m+1)\xi_0}^{m\xi_0} \partial_2\partial_1\Gamma(s,z,t)dzds. && (24)\end{aligned}$$

Note that since both φ and G are smooth with nice decay, so is Γ; thus all the steps above are justified. Next, we substitute (24) in (18) and take the absolute value of both sides to get

$$\begin{aligned} |f(t) - \tilde{f}_A(t)| &\leq \frac{1}{A}\sum_{n,m}|v_{n,m}|\left|\int_{(n+1)\tau_0}^{n\tau_0}\int_{(m+1)\xi_0}^{m\xi_0} \partial_2\partial_1\Gamma(s,z,t)dzds\right| && (25)\\ &\leq \frac{1}{A}\|v\|_{l^\infty}\|\partial_2\partial_1\Gamma(\cdot,\cdot,t)\|_{L^1(\mathbb{R}^2)}, && (26)\end{aligned}$$

for almost every t. To complete the proof, we will write that $D(t) := \|\partial_2\partial_1\Gamma(\cdot,\cdot,t)\|_{L^1(\mathbb{R}^2)}$ is uniformly bounded in L^2, i.e. $\|D\|_{L^2}$ does not depend on the particular choice of f, which implies that the error $\|f - \tilde{f}_A\|_{L^2}$ is $O(A^{-1})$. A simple calculation yields

$$\begin{aligned} \partial_2\partial_1\Gamma(s,z,t) &= \varphi(s-t)e^{izt}(\partial_1\partial_2 G(s,z) + it\partial_1 G(s,z))\\ &\quad -\varphi'(s-t)e^{izt}(\partial_2 G(s,z) + itG(s,z)). && (27)\end{aligned}$$

Thus,

$$D(t) \le (G_1 \star |\varphi|)(t) + (G_2 \star |\varphi'|)(t) + |t|((G_3 \star |\varphi|)(t) + (G_4 \star |\varphi'|)(t)), \tag{28}$$

where G_i are defined as in the statement of the theorem. Finally since φ and G are chosen such that the conditions stated in Theorem 1 are satisfied, $\|D\|_{L^2}$ is finite and

$$\|D\|_{L^2} \le \| \, |\varphi| \star G_1\|_{L^2} + \| \, |\varphi'| \star G_2\|_{L^2} + \|t(|\varphi| \star G_3)(t)\|_{L^2} + \|t(|\varphi'| \star G_4)(t)\|_{L^2}, \tag{29}$$

thus we conclude

$$\|f - \tilde{f}_A\|_{L^2} \le \frac{C}{A} \tag{30}$$

with C as in the statement of the theorem. □

Remark 1. Note that (15) still holds up to some small correction term if the frame (φ, τ_0, ξ_0) is "almost tight". A frame is said to be *almost tight* if the ratio of the frame bounds is close to 1. Suppose (φ, τ_0, ξ_0) is a Gabor frame with frame bounds A and B. If we denote the quantity $B/A - 1$ by r, for any function $f \in L^2(\mathbb{R})$ we have

$$f = \frac{2}{(2+r)A} \sum \langle f, \varphi_{n,m} \rangle \varphi_{n,m} + Rf, \tag{31}$$

where $\|R\| \le r/(2+r)$. In this case, after defining

$$\tilde{f}_A := \frac{2}{(2+r)A} \sum q_{n,m} \varphi_{n,m} \tag{32}$$

we can apply the proof of Theorem 1 to show that

$$|f - \tilde{f}_A| \le \frac{2}{(2+r)} \frac{C}{A} + \frac{r}{2+r}. \tag{33}$$

Thus, the approximation error $\|f - \tilde{f}_A\|_{L^2}$ still has the same behavior when $r \approx 0$.

Remark 2. Theorem 1 lists several conditions on φ and G; these are all fulfilled if, for example, φ and G are in the Schwartz spaces $\mathcal{S}(\mathbb{R})$ and $\mathcal{S}(\mathbb{R}^2)$, respectively.

Remark 3. Clearly, the algorithm is not shift-invariant; knowing the exact index of a quantizer output is essential, because to construct $\tilde{f}_A$ we need to multiply the quantizer output $p_{n,m} + (\Delta_1 r)_{n,m}$, by $G(n\tau_0, m\xi_0)$.

Remark 4. The TF$\Sigma\Delta$-II algorithm suggests an algorithm to "coarsely quantize" the Fourier coefficients of a function with compact support; this will be discussed in Section 3.

2.2 Higher-order schemes

We will define the k^{th}-order TF$\Sigma\Delta$-II scheme as follows: Let (φ, τ_0, ξ_0) be a tight Gabor frame with frame bound A. Suppose f is in $\mathcal{B}_G^{\varphi}$. Let $c_{n,m} = \langle f, \varphi_{n,m} \rangle$, $c^R_{n,m}$, $c^I_{n,m}$, $\tilde{c}^R_{n,m}$ and $\tilde{c}^I_{n,m}$ be as in Section 2.1. Let the superscript S be as before and consider the recursion relations

$$\begin{aligned} (\Delta_1^k u^S)_{n,m} &= \tilde{c}^S_{n,m} - p^S_{n,m} \\ p^S_{n,m} &= \text{sign}(M(\Delta_1^0 u^S_{n-1,m}, \ldots, \Delta_1^{k-1} u^S_{n-1,m}, \tilde{c^S}_{n,m})) \end{aligned} \tag{34}$$

$$\begin{aligned} (\Delta_2^k v^S)_{n,m} &= \bar{u}^S_{n,m} - r^S_{n,m} \\ r^S_{n,m} &= \text{sign}(\Delta_2^0 v^S_{n,m-1}, \ldots, \Delta_2^{k-1} v^S_{n,m-1}, \bar{u}^R_{n,m}) \end{aligned} \tag{35}$$

where k is a positive integer, $\bar{u}^S := u^S/C_{k,M}$ and M is a function which guarantees that u^R, v^R, u^I and v^I are uniformly bounded in l^∞ by $C_{k,M}$. Note that the recursion relations (34) and (35) correspond to k^{th}-order standard sigma-delta quantizers with $c^S_{n,m}$ and $\bar{u}^S_{n,m}$ respectively as their input. Thus, since all these sequences are bounded in l^∞ by 1, such an M exists for any positive integer k due to [3]; for a wider class of admissible M when $k=2$, consult [9].

Now set $v = v^R + iv^I$, $p = p^R + ip^I$, and $r = r^R + ir^I$. Note that

$$C_{k,M} G(n\tau_0, m\xi_0)(\Delta_1^k \Delta_2^k v)_{n,m} = c_{n,m} - G(n\tau_0.m\xi_0)(p_{n,m} + C_{k,M}\Delta_1^k r_{n,m}), \tag{36}$$

which suggests us the following definition of T_{TFII-k}: T_{TFII-k} will denote the mapping that maps c to $\tilde{q}$ with $\tilde{q}_{n,m} := G(n\tau_0.m\xi_0)(p_{n,m} + C_{k,M}\Delta_1^k r_{n,m})$.

Theorem 2. *Let $(\varphi, \tau_o.\xi_0)$ be a tight Gabor frame with frame bound A. Suppose f is in $\mathcal{B}_G^\varphi$. Let $c_{n,m} = \langle f, \varphi_{n,m}\rangle$ and put $q = T_{TFII-k}(c)$ where $k \geq 1$ is an integer. Define $\tilde{f}_{A,k}$ by $\tilde{f}_{A,k} := \frac{1}{A}\sum_{n,m} q_{n,m}\varphi_{n,m}$. Suppose G and φ are chosen such that for l, l' in $\{0, \cdots, k\}$,*

$$t^{l'}(|\varphi^{(l)}| \star G_{l,l'})(t) \in L^2(\mathbb{R}) \tag{37}$$

where $G_{l,l'}$ is defined by

$$G_{l,l'}(s) := \int \binom{k}{l'} |\partial_2^{(k-l')}\partial_1^{(k-l)} G(s,z)| dz. \tag{38}$$

In this case, the approximation error satisfies

$$\|f - \tilde{f}_{A,k}\|_{L^2} \leq \frac{C}{A^k} \tag{39}$$

where C is given by

$$C := \sum_{l=0}^{k} \binom{k}{l} \sum_{l'=0}^{k} \|t^{l'}(|\varphi^{(l)}| \star G_{l,l'})\|_{L^2}. \tag{40}$$

Proof. Let us start by writing the error term:

$$\begin{aligned} f(t) - \tilde{f}_{A,k}(t) &= \frac{1}{A}\sum_{n,m}(c_{n,m} - q_{nm})\varphi_{n,m} \\ &= \frac{1}{A}\sum_{n,m} C_{k,M} G(n\tau_0, m\xi_0)(\Delta_1^k\Delta_2^k v)_{n,m}\varphi_{n,m} \\ &= \frac{C_{k,M}}{A}\sum_{n,m} v_{n,m}\bar{\Delta}_1^k\bar{\Delta}_2^k G(n\tau_0, m\xi_0)\varphi_{n,m}, \end{aligned} \tag{41}$$

where the second equality is due to (36); the third equality is obtained by partial summation. Now set $\Gamma(s,z,t) = G(s,z)\varphi(t-s)e^{izt}$, and rewrite (41) as

$$f(t) - \tilde{f}_{A,k}(t) = \frac{C_{k,M}}{A}\sum_{n,m} v_{n,m}\bar{\Delta}_1^k\bar{\Delta}_2^k\Gamma(n\tau_0, m\xi_0, t). \tag{42}$$

By techniques identical to those used in the proof of Theorem 9 of [10] we have

$$|f(t) - \tilde{f}_{A,k}(t)| \leq \frac{C_{k,M}\|v\|_{l^\infty}}{A^k}\|\partial_2^k\partial_1^k\Gamma(\cdot,\cdot,t)\|_{L^1}. \tag{43}$$

We complete the proof by estimating $\|\partial_2^k\partial_1^k\Gamma(\cdot,\cdot,t)\|_{L^1}$. Note that

$$\partial_2^k\partial_1^k\Gamma(s,z,t) = \sum_{l=0}^{k}(-1)^l\binom{k}{l}\varphi^{(l)}(t-s)\sum_{l'=0}^{k}\binom{k}{l'}(it)^{l'}\partial_2^{(k-l')}\partial_1^{(k-l)}G(s,z), \tag{44}$$

which yields

$$|\partial_2^k \partial_1^k \Gamma(s,z,t)| \le \sum_{l=0}^{k} \binom{k}{l} |\varphi^{(l)}(t-s)| \sum_{l'=0}^{k} \binom{k}{l'} |t|^{l'} |\partial_2^{(k-l')} \partial_1^{(k-l)} G(s,z)|. \tag{45}$$

By integrating both sides, we get

$$\|\partial_2^k \partial_1^k \Gamma(\cdot,\cdot,t)\|_{L^1(\mathbb{R}^2)} \le \sum_{l=0}^{k} \sum_{l'=0}^{k} |t|^{l'} \binom{k}{l} \int \left(|\varphi^{(l)}(t-s)| \int \binom{k}{l'} |\partial_2^{(k-l')} \partial_1^{(k-l)} G(s,z)| dz \right) ds. \tag{46}$$

Finally by setting $G_{l,l'}(s) := \int \binom{k}{l'} |\partial_2^{(k-l')} \partial_1^{(k-l)} G(s,z)| dz$, we obtain

$$|f(t) - \tilde{f}_{A,k}(t)| \le \frac{C_{k,M} \|v\|_{l^\infty}}{A^k} \sum_{l=0}^{k} \sum_{l'=0}^{k} \binom{k}{l} |t|^{l'} (|\varphi| \star G_{l,l'})(t). \tag{47}$$

which, by taking the L^2-norm of both sides, yields

$$\|f - \tilde{f}_{A,k}\|_{L^2} \le \frac{C}{A^k} \tag{48}$$

where $C = \frac{C_{k,M}\|v\|_{l^\infty}}{A^k} \sum_{l=0}^{k} \sum_{l'=0}^{k} \binom{k}{l} \| |t|^{l'} (|\varphi| \star G_{l,l'})(t) \|_{L^2}$. □

2.3 Numerical Experiment

In this section we will fix a Gabor frame and quantize the frame coefficients of a function in $\mathcal{B}_G^\varphi$ via the first- and second-order TFΣΔ-II algorithms. We shall consider the Gabor frame generated by the function $\varphi(t) = \pi^{1/4} e^{-\frac{t^2}{2}}$. One can show that (φ, τ_0, ξ_0) constitutes an almost tight Gabor frame (i.e., both frame bounds A and B are approximately equal to $2\pi/(\tau_0\xi_0)$ if $\tau_0 \approx \xi_0$ is sufficiently small. For all the frames we use in this section, $|r|$ is smaller than the arithmetical precision of the computer, where, as in Remark 1, we define r by $r := B/A - 1$.

Consider the function

$$f(t) = 0.5 e^{-i0.9t^3} e^{-0.05t^2}. \tag{49}$$

which clearly is in $\mathcal{B}_G^\varphi$. We use an FFT-based algorithm to compute the frame coefficients $\langle f, \varphi_{n,m} \rangle$. For simplicity, we use $G(\tau,\xi) = 2|\langle f, \varphi_{\tau,\xi} \rangle|$ and quantize the frame coefficients according to the algorithm given in (9) for the first-order scheme, and according to the algorithm given in (34)-(35) (with $k = 2$) for the second-order scheme. We apply the algorithms to different frame expansions of f with different frame bounds. The frames we use are (φ, τ_0, ξ_0) with $\tau_0 = \xi_0$ ranging from 0.2 to 1.5. Figure 1 shows the L^2-approximation error estimates as a function of the frame coefficient of the expansion. The graphs of $\tilde{f}_A$ which is obtained from the original function f via the first-order TFΣΔ-II is given in Figure 2. In Figure 3 we show the L^2-approximation error estimates that are obtained using the second-order TFΣΔ-II. Finally, in Figure 4 we present the graphs of $\tilde{f}_{A,2}$ that are obtained from f via the second-order TFΣΔ-II.

3 Coarse quantization of the Fourier coefficients of certain compactly supported functions

Let f be a function in $L^p(\mathbb{R})$ with $p > 1$ such that its support is on $[-\pi, \pi]$. Clearly we can extend f to a periodic function f_λ by setting $f_\lambda(t) = f(t)$ for $t \in [-\lambda\pi, \lambda\pi)$, and $f_\lambda(t - 2\lambda\pi) = f_\lambda(t)$. Then the Fourier series of f_λ is given by $f_\lambda(t) = \frac{1}{\sqrt{2\pi\lambda}} \sum_n \hat{f}(\frac{n}{\lambda}) e^{i\frac{n}{\lambda}t}$, where equality holds pointwise by

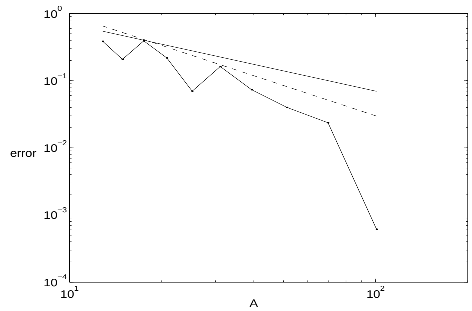
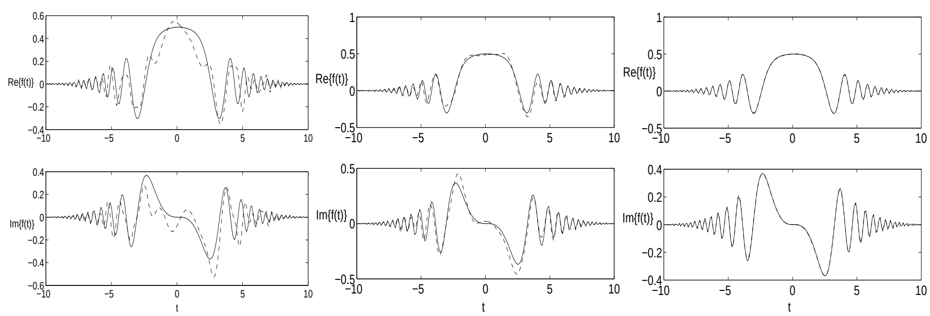


Figure 1: The L^2-approximation error $\|f - \tilde{f}_A\|_{L^2}$ vs. the frame bound A for the first-order case; $\tilde{f}_A$ is the approximation generated by the first-order TFΣΔ-II algorithm. The solid line is the graph of $\{(A, 7A^{-1})\}$, and the dashed one is the graph of $\{(A, 30A^{-1.5})\}$



(a) (b) (c)

Figure 2: The graph of $\tilde{f}_A$, the L^2-approximation of f obtained via the first-order TFΣΔ-II algorithm, for three different values of the frame bound A, along with the graph of f. In each figure, the top graph is the real part of f and $\tilde{f}_A$: the solid graph belongs to f and the dashed to $\tilde{f}_A$; the bottom graph is the imaginary part of f and $\tilde{f}_A$: the solid belongs to f and the dashed to $\tilde{f}_A$. Figure 2a is the graph of $\tilde{f}_{12.83}$ along with f; Figure 2b is the graph of $\tilde{f}_{25.16}$ and Figure 2c is the graph of $\tilde{f}_{100.64}$.

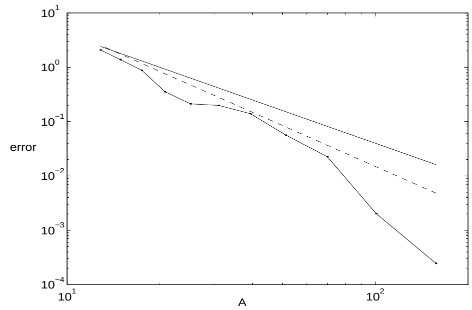
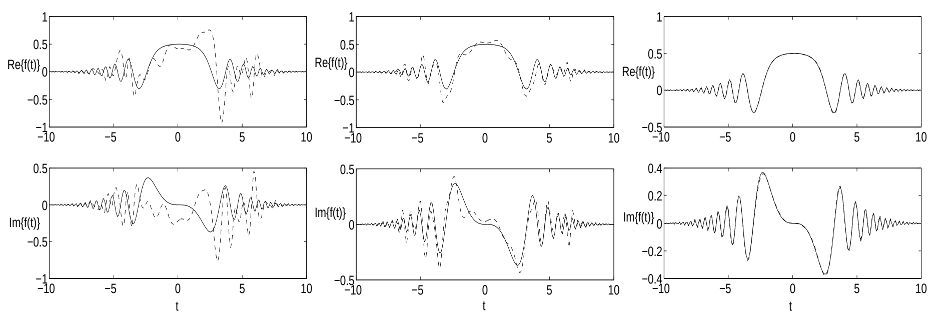


Figure 3: The L^2-approximation error $\|f - \tilde{f}_{A,2}\|_{L^2}$ vs. the frame bound A for the second order case; $\tilde{f}_{A,2}$ is the approximation generated by the second-order TFΣΔ-II algorithm. The solid line is the graph of $\{(A, 400A^{-2})\}$, and the dashed one is the graph of $\{(A, 1500A^{-2.5})\}$.



(a) (b) (c)

Figure 4: The graph of $\tilde{f}_{A,2}$, the L^2-approximation of f obtained via the second-order TFΣΔ-II algorithm, for three different values of the frame bound A, along with the graph of f. In each figure, the top graph is the real part of f and $\tilde{f}_{A,2}$: the solid graph belongs to f and the dashed to $\tilde{f}_{A,2}$; the bottom graph is the imaginary part of f and $\tilde{f}_{A,2}$: the solid belongs to f and the dashed to $\tilde{f}_{A,2}$. Figure 4a is the graph of $\tilde{f}_{12.83,2}$ along with f; Figure 4b is the graph of $\tilde{f}_{25.16,2}$ and Figure 4c is the graph of $\tilde{f}_{100.64,2}$.

our choice of f. Here $\hat{f}$ denotes the Fourier transform of f, i.e., $\hat{f}(\xi) := \frac{1}{\sqrt{2\pi}} \int f(t) e^{-i\xi t} dt$. Note that we also have

$$f(t) = f_\lambda(t) \chi_{[-\lambda\pi, \lambda\pi]}(t), \tag{50}$$

which for $t \in [-\lambda\pi, \lambda\pi]$ yields

$$f(t) = \frac{1}{\sqrt{2\pi}\lambda} \sum_n \hat{f}(\frac{n}{\lambda}) e^{i\frac{n}{\lambda}t}. \tag{51}$$

Now let us fix a continuously differentiable function G, i.e., $G \in C^1(\mathbb{R})$, such that G and G' are in $L^1(\mathbb{R})$. Let f be a function in $L^p(\mathbb{R})$ with $p > 1$ which is compactly supported with $\text{supp}(f) \subseteq [-\pi, \pi]$ and whose Fourier transform $\hat{f}$ satisfies

$$|\hat{f}(\xi)| \leq G(\xi). \tag{52}$$

Denote the collection of all such functions by $\mathcal{B}_G$. For $f \in \mathcal{B}_G$, we clearly have

$$|\frac{\hat{f}(\frac{n}{\lambda})}{G(\frac{n}{\lambda})}| < 1. \tag{53}$$

Therefore, we can use the standard first-order sigma-delta scheme to quantize the sequence $(\frac{\hat{f}(\frac{n}{\lambda})}{G(\frac{n}{\lambda})})$: Consider the recursion relations

$$\begin{aligned} (\Delta v)_n &= \frac{\hat{f}(\frac{n}{\lambda})}{G(\frac{n}{\lambda})} - \tilde{q}_n^\lambda \\ \tilde{q}_n^\lambda &= sign(v_{n-1} + \frac{\hat{f}(\frac{n}{\lambda})}{G(\frac{n}{\lambda})}) \end{aligned} \tag{54}$$

Since we have (53), we have $|v_n| \leq 1$ for all n by [3] assuming $|v_0| < 1$. Let us define T_G as the mapping that maps the sequence $(f(\frac{n}{\lambda}))$ to the sequence q^λ, where $q_n^\lambda := G(\frac{n}{\lambda}) q_n^\lambda$. Note that

$$G(\frac{n}{\lambda})(v_n - v_{n-1}) = \hat{f}(\frac{n}{\lambda}) - q_n^\lambda. \tag{55}$$

Now define

$$\tilde{f}_\lambda(t) := \frac{1}{\sqrt{2\pi}\lambda} \sum_{n \in \mathbb{Z}} q_n^\lambda e^{i\frac{n}{\lambda}t}. \tag{56}$$

Clearly, for $t \in [-\pi, \pi]$,

$$\begin{aligned} |f(t) - \tilde{f}_\lambda(t)| &= \frac{1}{\sqrt{2\pi}\lambda} |\sum_n (\hat{f}(\frac{n}{\lambda}) - q_n^\lambda) e^{i\frac{n}{\lambda}t}| \\ &= \frac{1}{\sqrt{2\pi}\lambda} |\sum_n G(\frac{n}{\lambda})(v_n - v_{n-1}) e^{i\frac{n}{\lambda}t}| \\ &= \frac{1}{\sqrt{2\pi}\lambda} |\sum_n v_n (G(\frac{n}{\lambda}) e^{i\frac{n}{\lambda}t} - G(\frac{n+1}{\lambda}) e^{i\frac{n+1}{\lambda}t})| \\ &\leq \frac{1}{\sqrt{2\pi}\lambda} \sum_n |G(\frac{n}{\lambda}) e^{i\frac{n}{\lambda}t} - G(\frac{n+1}{\lambda}) e^{i\frac{n+1}{\lambda}t})|; \end{aligned} \tag{57}$$

the second equality is the result of summation by parts (note that the boundary values vanish again since $G(\frac{n}{\lambda})$ tends to zero as $|n|$ approaches infinity); the final inequality follows because $|v_n|$

is bounded by 1. Now set $\Gamma(z,t) := G(z)e^{izt}$ and rewrite (57) as

$$\begin{aligned} |f(t)-\tilde{f}_\lambda(t)| &\leq \frac{1}{\sqrt{2\pi}\lambda}\sum_n |\Gamma(\frac{n}{\lambda},t)-\Gamma(\frac{n+1}{\lambda},t)| \\ &\leq \frac{1}{\sqrt{2\pi}\lambda}\sum_n \int_{\frac{n+1}{\lambda}}^{\frac{n}{\lambda}} |\partial_1\Gamma(z,t)|dz \\ &\leq \frac{1}{\sqrt{2\pi}\lambda}\|\partial_1\Gamma(\cdot,t)\|_{L^1}. \end{aligned} \tag{58}$$

Note that the second inequality is due to the fact that Γ is smooth.

Finally we will calculate $\|\partial_1\Gamma(\cdot,t)\|_{L^1}$. Clearly,

$$\partial_1\Gamma(z,t) = G'(z)e^{izt} + itG(z)e^{izt}, \tag{59}$$

which yields

$$\|\partial_1\Gamma(\cdot,t)\|_{L^1} \leq \frac{1}{\sqrt{2\pi}\lambda}(\|G'\|_{L^1} + |t|\|G\|_{L^1}). \tag{60}$$

But since f is supported on $[-\pi,\pi]$, we can conclude that

$$\sup_{t\in[-\pi,\pi]} |f(t)-\tilde{f}_\lambda(t)| \leq \frac{C}{\lambda} \tag{61}$$

with $C = \sqrt{2\pi}(\frac{1}{2\pi}\|G'\|_{L^1} + \frac{1}{2}\|G\|_{L^1})$. Thus we proved

Theorem 3. *Let $G \in C^1(\mathbb{R})$ be a fixed function such that G and G' are in $L^1(\mathbb{R})$. Suppose f is in $\mathcal{B}_G$. Then*

$$\sup_t |f(t)-\tilde{f}_\lambda(t)| \leq \frac{\sqrt{2\pi}(\frac{1}{2\pi}\|G'\|_{L^1} + \frac{1}{2}\|G\|_{L^1})}{\lambda}, \tag{62}$$

where $\tilde{f}_\lambda$ is defined as in (56).

Remark 5. It is straight-forward to define the higher-order versions of the above described scheme. The k^{th}-order scheme can be defined by replacing the first-order backward difference operator in (54) by a k^{th}-order backward difference operator, i.e. a k^{th}-order quantizer is defined by the following recursion relations:

$$\begin{aligned} (\Delta^k v)_n &= \frac{\hat{f}(\frac{n}{\lambda})}{G(\frac{n}{\lambda})} - q_n^\lambda \\ q_n^\lambda &= sign(M((\Delta^0 v)_{n-1},\ldots,(\Delta^{k-1}v)_{n-1}\frac{\hat{f}(\frac{n}{\lambda})}{G(\frac{n}{\lambda})})) \end{aligned} \tag{63}$$

where M is chosen such that v is uniformly bounded in l^∞. The existence of such M due to [3] for arbitrary k.

In this case, the approximation error is $O(\lambda^{-k})$ for functions in $\mathcal{B}_G$ with $G \in C^k(\mathbb{R})$ such that $G^{(j)} \in L^1(\mathbb{R})$ for $j = 0,1,\ldots,k$. The proof is similar to the proof of Theorem 2.

References

[1] W. Chen and B. Han. Improving the accuracy estimate for the first order sigma-delta modulator. *J. Amer. Math. Soc.* sunmitted in 2003.

[2] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992.

[3] I. Daubechies and R. DeVore. Reconstructing a bandlimited function from very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order. *Annals of Mathematics.* to appear.

[4] I. Daubechies and A.J.E.M. Janssen. Two theorems on lattice expansions. *IEEE Transactions on Information Theory*, 39(1):3–6, January 1993.

[5] H.G. Feichtinger and Thomas Strohmer, editors. *Gabor analysis and algorithms: Theory and applications.* Birkhäuser, Boston, 1998.

[6] Karlheinz Gröchenig. *Foundations of Time-Frequency Analysis.* Birkhäuser, Boston, 2001.

[7] C.S. Güntürk. Approximating a bandlimited function using very coarsely quantized data: Improved error estimates in sigma-delta modulation. *J. Amer. Math. Soc.* to appear.

[8] C.S. Güntürk. *Harmonic Analysis of Two Problems in Signal Quantization and Compression.* PhD thesis, Princeton University, 2000.

[9] Özgür Yılmaz. Stability analysis for several sigma-delta methods of coarse quantization of bandlimited functions. *Constructive Approximation*, 18:599–623, 2002.

[10] Özgür Yılmaz. Coarse quantization of highly redundant time-frequency representations of square-integrable functions. *Appl. Comput. Harmon. Anal.*, 14:107–132, 2003.

[11] S.R. Norsworthy, R.Schreier, and G.C. Themes, editors. *Delta-Sigma Data Converters.* IEEE Press, 1997.