WHITE NOISE HYPOTHESIS FOR UNIFORM QUANTIZATION ERRORS

DAVID JIMENEZ, LONG WANG, AND YANG WANG

Abstract. The White Noise Hypothesis (WNH) assumes that in the uniform pulse code modulation (PCM) quantization scheme the errors in individual channels behave like white noise, i.e. they are independent and identically distributed random variables. The WNH is key to estimating the mean square quantization error (MSE). But is the WNH valid? In this paper we take a close look at the WNH. We show that in a redundant system the errors from individual channels can never be independent. Thus to an extent the WNH is invalid. Our numerical experients also indicate that with coarse quantization the WNH is far from being valid. However, as the main result of this paper we show that with fine quantizations the WNH is essentially valid, in which the errors from individual channels become asymptotically *pairwise* independent, each uniformly distributed in $[-\Delta/2, \Delta/2)$, where Δ denotes the stepsize of the quantization.

1. Introduction

In processing, analysing and storing of analog signals it is often necessary to make atomic decompositions of the signal using a given set of *atoms*, or *dictionary* $\{\mathbf{v}_j\}$. In this approach, a signal $\mathbf{x}$ is represented as a linear combination of $\{\mathbf{v}_j\}$,

$$\mathbf{x} = \sum_j c_j \mathbf{v}_j.$$

In practice $\{\mathbf{v}_j\}$ is a finite set. Furthermore, for the purpose of error correction, recovery from data erasures or robustness, redundancy is built into $\{\mathbf{v}_j\}$, i.e. more elements than needed are in $\{\mathbf{v}_j\}$. Instead of a true basis, $\{\mathbf{v}_j\}$ is chosen to be a frame. Since $\{\mathbf{v}_j\}$ is a finite set, we may without loss of generality assume $\{\mathbf{v}_j\}_{j=1}^N$ are vectors in $\mathbb{R}^d$ with $N \geq d$.

Let $F = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N]$ be the $d \times N$ matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_N$. We say $\{\mathbf{v}_j\}_{j=1}^N$ is a *frame* if F has rank d. Let $\lambda_{\max} \geq \lambda_{\min} > 0$ be the maximal and minimal

1991 *Mathematics Subject Classification.* Primary 42C15.

Key words and phrases. Frames, tight frame, pulse code modulation, quantization, white noise hypothesis.

The third author is supported in part by the National Science Foundation, grant DMS-0456538.

eigenvalues of FF^T, respectively. It is easily checked that

$$\lambda_{\min}\|\mathbf{x}\|^2 \leq \sum_{j=1}^N |\mathbf{x}\cdot\mathbf{v}_j|^2 \leq \lambda_{\max}\|\mathbf{x}\|^2. \tag{1.1}$$

$\lambda_{\max}$ and $\lambda_{\min}$ are called the *upper and lower frame bounds* for the frame, respectively. If $\lambda_{\max} = \lambda_{\min} = \lambda$, in which case $FF^T = \lambda I_d$, we call $\{\mathbf{v}_j\}_{j=1}^N$ a *tight frame* with frame constant λ. Note that any signal $\mathbf{x} \in \mathbb{R}^d$ can be easily reconstructed using the data $\{\mathbf{x}\cdot\mathbf{v}_j\}_{j=1}^N$. Set $\mathbf{y} = [\mathbf{x}\cdot\mathbf{v}_1, \mathbf{x}\cdot\mathbf{v}_2, \cdots, \mathbf{x}\cdot\mathbf{v}_N]^T$. Then $\mathbf{y} = F^T\mathbf{x}$ and

$$(FF^T)^{-1}F\mathbf{y} = (FF^T)^{-1}FF^T\mathbf{x} = \mathbf{x}.$$

Let $G = (FF^T)^{-1}F = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N]$. The set of columns $\{\mathbf{u}_j\}_{j=1}^N$ of G is called the *canonical dual frame* of the frame $\{\mathbf{v}_j\}_{j=1}^N$. We have the reconstruction

$$\mathbf{x} = \sum_{j=1}^N (\mathbf{x}\cdot\mathbf{v}_j)\,\mathbf{u}_j. \tag{1.2}$$

If $\{\mathbf{v}_j\}_{j=1}^N$ is a tight frame with frame constant λ, then $G = \lambda^{-1}F$, and we have the reconstruction

$$\mathbf{x} = \frac{1}{\lambda}\sum_{j=1}^N (\mathbf{x}\cdot\mathbf{v}_j)\,\mathbf{v}_j. \tag{1.3}$$

In digital applications, quantizations will have to be performed. The simplest scheme is the Pulse Code Modulation (PCM) quantization scheme, in which the coefficients $\{\mathbf{x}\cdot\mathbf{v}_j\}_{j=1}^N$ are quantized. In this paper we consider exclusively *uniform quantizations.* Let $\mathcal{A} = \Delta\mathbb{Z}$ where $\Delta > 0$ is the quantization step. With uniform quantization a real value t is replaced with the value in $\mathcal{A}$ that is the closest to t. So, in our setting, t is replaced with $Q_\Delta(t)$ given by

$$Q_\Delta(t) := \left\lfloor \frac{t}{\Delta} + \frac{1}{2} \right\rfloor \Delta.$$

Thus, given a frame $\{\mathbf{v}_j\}_{j=1}^N$ and its canonical dual frame $\{\mathbf{u}_j\}_{j=1}^N$, instead of using the data $\{\mathbf{x}\cdot\mathbf{v}_j\}_{j=1}^N$ and (1.2) to obtain a perfect reconstruction, we use the data $\{Q_\Delta(\mathbf{x}\cdot\mathbf{v}_j)\}_{j=1}^N$ and obtain an imperfect reconstruction

$$\tilde{\mathbf{x}} = \sum_{j=1}^N Q_\Delta\left(\mathbf{x}\cdot\mathbf{v}_j\right)\mathbf{u}_j. \tag{1.4}$$

This raises the following question: How good is the reconstruction? This question has been studied in terms of both the worst case error and the mean square error (**MSE**), see e.g. [13].

Note that the error from the reconstruction is

$$\mathbf{x}-\tilde{\mathbf{x}}=\sum_{j=1}^{N}\tau_{\Delta}\left(\mathbf{x}\cdot\mathbf{v}_j\right)\mathbf{u}_j, \tag{1.5}$$

where $\tau_{\Delta}(t):=t-Q_{\Delta}(t)=\left(\left\{\frac{t}{\Delta}+\frac{1}{2}\right\}-\frac{1}{2}\right)\Delta$, with $\{\cdot\}$ denoting the fractional part. While an *a priori* error bound is relatively straightforward to obtain, the *mean square error* $\mathbf{MSE}:=\mathcal{E}\left(\|\mathbf{x}-\tilde{\mathbf{x}}\|^2\right)$, assuming certain probability distribution for $\mathbf{x}$, is much harder. To simplify the problem, the so-called *White Noise Hypothesis* (**WNH**), is employed by engineers and mathematicians in this area (see e.g. [2, 3, 13]). The **WNH** asserts the following:

- Each $\tau_{\Delta}\left(\mathbf{x}\cdot\mathbf{v}_j\right)$ is uniformly distributed in $[-\Delta/2,\Delta/2)$; hence it has mean 0 and variance $\Delta^2/12$.
- $\{\tau_{\Delta}\left(\mathbf{x}\cdot\mathbf{v}_j\right)\}_{j=1}^{N}$ are independent random variables.

With the **WNH** it is an easy derivation, which we furnish in the next section, that the **MSE** is given by

$$\mathcal{E}\left(\|\mathbf{x}-\tilde{\mathbf{x}}\|^2\right)=\frac{\Delta^2}{12}\sum_{j=1}^{d}\lambda_j^{-1}=\frac{\Delta^2}{12}\sum_{j=1}^{N}\|\mathbf{u}_j\|^2. \tag{1.6}$$

where $\{\lambda_j\}$ are the eigenvalues of FF^T.

Note that using (1.6) the **MSE** for quantization decreases by a factor of 4 if we decrease Δ by a factor of 2. It amounts to an increase in signal to noise ratio of approximately 6dB ($10\log_{10}4\approx 6$). This is often referred to as the *6dB-per-bit-rule*.

The **WNH** is often called *Bennett's White Noise Assumption* [2, 3]. Bennett studied quantization error (distortion) in his fundamental paper [4] in the scalar setting. He demonstrated that under the assumption that the scalar random variable has a smooth density, the quantization error behaves like uniformly distributed "random noise" when Δ is small, resulting in the **MSE** to be approximately $\Delta^2/12$. Bennett also studied quantization errors in the nonuniform quantization setting, which can often be reduced to the uniform setting by the use of companders. The current interest in the **WNH** stems from the study of vector quantization, in which several correlated signals are quantized simultaneously such as in our setting. A vast literature on vector quantization and on vector quantization errors exist, and for an excellent and comprehensive survey on vector quantization see Gray and

Neuhoff [14]. A weaker form of the **WNH**, which states that the error components are approximately uncorrelated in the high resolution setting, i.e. when Δ is small, is often found in engineering literatures without rigorous proofs (see [11] and the discussion in [22]). A rigorous proof of this weaker form of the **WNH** was first given in Viswanathan and Zamir [22]. More precisely, they proved that if two random variables X, Y have a joint density function then $\frac{1}{\Delta^2}\mathcal{E}\left(\tau_\Delta(X)\tau_\Delta(Y)\right) \longrightarrow 0$ as $\Delta \to 0$. Viswanathan and Zamir also proved similar results in the nonuniform quantization setting, under much stronger assumptions.

It should be pointed out that much of the advantage of vector quantization comes from the fact that the quantizations are *not* necessarily performed independently on each channel. As a result of it there are many interesting and challenging mathematical problems in nonuniform vector quantization. While the focus of this paper is on uniform quantization, we hope it will be a very first step in resolving the problem in the more general setting.

The objective of this paper is a rather modest one. Given the vast literature on quantization errors and some of the general confusions regarding the **WNH**, this paper aims to provide complete analysis and rigorous mathematical theorems on the behavior of quantization errors. These results are by no means difficult, and they are also rather intuitive. Nevertheless we feel there is a need to have them written down. If nothing else we hope this paper will serve to clarify things on the **WNH** in the uniform quantization setting. As a very simple result we show under the assumption that the distribution of $\mathbf{x}$ has a density (absolutely continuous), the components of the quantization errors $\{\tau_\Delta\left(\mathbf{x}\cdot\mathbf{v}_j\right)\}_{j=1}^N$ can *never* be independent if $N > d$. However, we show that asymptotically the **WNH** is almost valid by proving stronger and more general results than that in [22]. More precisely, we prove that if a set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ are linearly independent then the normalized quantization errors $\left\{\frac{1}{\Delta}\tau_\Delta\left(\mathbf{x}\cdot\mathbf{u}_j\right)\right\}_{j=1}^k$ converge in distribution to independent and uniformly distributed random variables as $\Delta \to 0^+$. Applying it to the frame setting, we show that if the vectors $\{\mathbf{v}_j\}_{j=1}^N$ are pairwise linearly independent then $\{\tau_\Delta\left(\mathbf{x}\cdot\mathbf{v}_j\right)\}_{j=1}^N$ becomes asymptotically *pairwise independent* and thus *pairwise uncorrelated*, and each $\tau_\Delta\left(\mathbf{x}\cdot\mathbf{v}_j\right)$ becomes asymptotically uniformly distributed on $[-\Delta/2, \Delta/2]$. These slightly weaker assumptions are sufficient to lead to the **MSE** given by (1.6) asymptotically. Furthermore, we also characterize completely the asymptotic behavior of the **MSE** if some $\mathbf{v}_j$'s are parallel. These and other results are stated and proved in subsequent sections.

Several people had given us helpful suggestions on this paper. But in particular we wish to express our gratitude to the referee, who not only read the manuscript very carefully but provided us with a number of valuable suggestions, particularly on the vast engineering literature regarding vector quantization.

2. *A Priori* Error Bound and MSE under the WNH

In this section we derive *a priori* error bound and a formula for the **MSE** under the **WNH**. These results are not new. We include them for self-containment. We use the following settings throughout this section: Let $\{\mathbf{v}_j\}_{j=1}^N$ be a frame in $\mathbb{R}^d$ with corresponding frame matrix $F = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N]$. The eigenvalues of FF^T are $\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d = \lambda_{\min} > 0$. Let $\{\mathbf{u}_j\}_{j=1}^N$ be the canonical dual frame with corresponding matrix $G = (FF^T)^{-1}F$. For any $\mathbf{x} = \sum_{j=1}^N (\mathbf{x}\cdot\mathbf{v}_j)\,\mathbf{u}_j$, using the quantization alphabet $\mathcal{A} = \Delta\mathbb{Z}$ we have the PCM quantized reconstruction

$$\tilde{\mathbf{x}} = \sum_{j=1}^N Q_\Delta\left(\mathbf{x}\cdot\mathbf{v}_j\right)\mathbf{u}_j.$$

Proposition 2.1. *For any $\mathbf{x} \in \mathbb{R}^d$ we have*

$$\|\mathbf{x} - \tilde{\mathbf{x}}\| \leq \frac{1}{2}\sqrt{\frac{N}{\lambda_{\min}}}\Delta. \tag{2.1}$$

If in addition $\{\mathbf{v}_j\}_{j=1}^N$ is a tight frame with frame constant λ, then

$$\|\mathbf{x} - \tilde{\mathbf{x}}\| \leq \frac{1}{2}\sqrt{\frac{N}{\lambda}}\Delta. \tag{2.2}$$

Proof. We have $\mathbf{x} - \tilde{\mathbf{x}} = \sum_{j=1}^N \tau_\Delta\left(\mathbf{x}\cdot\mathbf{v}_j\right)\mathbf{u}_j = G\mathbf{y}$, where $\mathbf{y} = \left[\tau_\Delta\left(\mathbf{x}\cdot\mathbf{v}_1\right), \ldots, \tau_\Delta\left(\mathbf{x}\cdot\mathbf{v}_N\right)\right]^T$. Thus $\|\mathbf{x} - \tilde{\mathbf{x}}\|^2 = \mathbf{y}^T G^T G\mathbf{y} \leq \rho\left(G^T G\right)\|\mathbf{y}\|^2$ where $\rho(\cdot)$ denotes the spectral radius. Now $\rho(G^T G) = \rho(GG^T) = \rho((FF^T)^{-1}) = \lambda_{\min}^{-1}$. Observe that $|\tau_\Delta\left(\mathbf{x}\cdot\mathbf{v}_j\right)| \leq \Delta/2$. Thus $\|\mathbf{y}\|^2 \leq N(\Delta/2)^2$. This yields an *a priori* error bound (2.1). The bound (2.2) is an immediate corollary. ■

Proposition 2.2. *Under the* **WNH**, *the* **MSE** *is*

$$\mathcal{E}\left(\|\mathbf{x} - \tilde{\mathbf{x}}\|^2\right) = \frac{\Delta^2}{12}\sum_{j=1}^d \lambda_j^{-1} = \frac{\Delta^2}{12}\sum_{j=1}^N \|\mathbf{u}_j\|^2. \tag{2.3}$$

In particular, if $\{\mathbf{v}_j\}_{j=1}^N$ is a tight frame with frame constant λ, then

$$\mathcal{E}\left(\|\mathbf{x}-\tilde{\mathbf{x}}\|^2\right) = \frac{d}{12\lambda}\Delta^2. \tag{2.4}$$

Proof. Denote $G^TG = [b_{ij}]_{i,j=1}^N$ and again let $\mathbf{y} = [\tau_\Delta(\mathbf{x}\cdot\mathbf{v}_1), \dots, \tau_\Delta(\mathbf{x}\cdot\mathbf{v}_N)]^T$. Note that with the **WNH**, $\mathcal{E}(y_iy_j) = \mathcal{E}(\tau_\Delta(\mathbf{x}\cdot\mathbf{v}_i)\tau_\Delta(\mathbf{x}\cdot\mathbf{v}_j)) = (\Delta^2/12)\delta_{ij}$. Now $\mathbf{x}-\tilde{\mathbf{x}} = G\mathbf{y}$ and hence

$$\mathcal{E}\left(\|\mathbf{x}-\tilde{\mathbf{x}}\|^2\right) = \mathcal{E}\left(\mathbf{y}^TG^TG\mathbf{y}\right) = \sum_{i,j=1}^N b_{ij}\mathcal{E}\left(y_iy_j\right) = \sum_{i=1}^N b_{ii}\frac{\Delta^2}{12} = \frac{\Delta^2}{12}\mathrm{tr}(G^TG).$$

Finally, $\mathrm{tr}(G^TG) = \sum_{j=1}^N \|\mathbf{u}_j\|^2$, and $\mathrm{tr}(G^TG) = \mathrm{tr}(GG^T) = \mathrm{tr}((FF^T)^{-1}) = \sum_{j=1}^d \lambda_j^{-1}$. ■

Remark: The **MSE** formulae (2.3-2.4) still hold if the independence of $\{\tau_\Delta(\mathbf{x}\cdot\mathbf{v}_j)\}_{j=1}^N$ in the **WNH** is replaced with the weaker condition that $\{\tau_\Delta(\mathbf{x}\cdot\mathbf{v}_j)\}_{j=1}^N$ are uncorrelated.

3. A Closer Look at the WNH

The **WNH** asserts that the error components $\{\tau_\Delta(\mathbf{x}\cdot\mathbf{v}_j)\}_{j=1}^N$ are independent and identically distributed random variables. Intuitively this cannot be true if $N > d$. This is indeed the case in general. The following is a simple result.

Theorem 3.1. *Let $\mathbf{X} \in \mathbb{R}^d$ be an absolutely continuous random vector. Let $\{\mathbf{v}_j\}_{j=1}^N$ be nonzero vectors in $\mathbb{R}^d$ with $N > d$. Then the random variables $\{\tau_\Delta(\mathbf{X}\cdot\mathbf{v}_j)\}_{j=1}^N$ are not independent.*

Proof. Let F be the frame matrix for the frame $\{\mathbf{v}_j\}$. Then $\dim(\mathrm{range}(F^T)) \le d$, and therefore $\mathcal{L}(\mathrm{range}(F^T)) = 0$ where $\mathcal{L}$ is the Lebesgue measure on $\mathbb{R}^N$. Let $\mathbf{Y} = [Y_1, \dots, Y_N]^T := F^T\mathbf{X}$, and let $\tilde{\mathbf{Y}} = [Q_\Delta(Y_1), \dots, Q_\Delta(Y_N)]^T$ be the quantized $\mathbf{Y}$. Denote $\mathbf{Z} = \mathbf{Y} - \tilde{\mathbf{Y}} = [Z_1, \dots, Z_N]^T$. Note that $Y_j = \mathbf{v}_j \cdot \mathbf{X}$, so each Y_j is absolutely continuous, and therefore so is each Z_j. If $\{Z_j\}$ are independent, then $\mathbf{Z}$ must be absolutely continuous.

Now, Set $\Omega := \mathrm{range}(F^T) + \Delta\mathbb{Z}^N$. Then $\mathcal{L}(\Omega) = 0$ because $\Delta\mathbb{Z}^N$ is a countable set. However, $\mathbf{Z}$ takes values in Ω so $P(\mathbf{Z} \in \Omega) = 1$. This contradicts the absolute continuity of $\mathbf{Z}$. ■

Remark: Actually for Theorem 3.1 to hold we only need to assume that $\mathbf{X}$ has an absolutely continuous component, i.e. $\mathbf{X} = \mathbf{X}_c + \mathbf{X}_s$ where $\mathbf{X}_c \ne 0$ is absolutely continuous and $\mathbf{X}_s$

is singular. However, the theorem can fail without the absolute continuity condition, even if each component of $\mathbf{X}$ may be absolutely continuous. The simplest example is to take $\mathbf{X} = [X, -X]^T$ where X is any random variable and $\mathbf{v}_1 = [1,1]^T$ and $\mathbf{v}_2 = [1,-1]^T$.

Even when $N = d$ the **WNH** holds only under rather strict conditions. The following is another simple result.

Proposition 3.2. *Let $\mathbf{X} = [X_1, \dots, X_m]^T$ be a random vector in $\mathbb{R}^m$ whose distribution has density function $g(x_1, \dots, x_m)$.*

(1) *The error components $\{\tau_\Delta(X_j)\}_{j=1}^m$ are independent if and only if there exist complex numbers $\{\beta_j(n) : 1 \le j \le m, n \in \mathbb{Z}\}$ such that*
$$\widehat{g}\left(\frac{a_1}{\Delta}, \dots, \frac{a_m}{\Delta}\right) = \beta_1(a_1) \cdots \beta_m(a_m) \tag{3.1}$$
for all $[a_1, \dots, a_m]^T \in \mathbb{Z}^m$.

(2) *Let $h_j(t)$ be the marginal density of X_j. Then $\{\tau_\Delta(X_j)\}_{j=1}^m$ are identically distributed if and only if*
$$\sum_{n\in\mathbb{Z}} h_j(t - n\Delta) = H(t) \quad a.e.$$
for some $H(t)$ independent of j. They are uniformly distributed on $[-\Delta/2, \Delta/2]$ if and only if $H(t) = 1/\Delta$ a.e..

Proof. To prove (1) denote $\mathcal{I}_\Delta = [-\Delta/2, \Delta/2]$ and $\mathbf{Y} = [\tau_\Delta(X_1), \dots, \tau_\Delta(X_m)]^T$. Observe that $\mathbf{Y}$ has a density
$$G(\mathbf{y}) := \sum_{\mathbf{a}\in\mathbb{Z}^m} g(\mathbf{y} - \Delta\mathbf{a}) \tag{3.2}$$
for $\mathbf{y} \in \mathcal{I}_\Delta^m$. The density $G(\mathbf{y})$ is periodic with period Δ, and it is well known that its Fourier series is given by $G(\mathbf{y}) = \sum_{\mathbf{a}\in\mathbb{Z}^m} c_{\mathbf{a}} e^{2i\pi\frac{\mathbf{a}}{\Delta}\cdot\mathbf{y}}$, where $c_{\mathbf{a}} = \widehat{g}\left(\frac{\mathbf{a}}{\Delta}\right)$. But $\{Y_j\}_{j=1}^m$ are independent if and only if on $\mathcal{I}_\Delta^m$ we have $g(y_1, \dots, y_m) = g_1(y_1) \cdots g_m(y_m)$. This happens if and only if
$$\widehat{g}\left(\frac{a_1}{\Delta}, \frac{a_2}{\Delta}, \dots, \frac{a_m}{\Delta}\right) = h_1\left(\frac{a_1}{\Delta}\right) h_2\left(\frac{a_2}{\Delta}\right) \cdots h_m\left(\frac{a_m}{\Delta}\right)$$
for all $\mathbf{a} = [a_1, \dots, a_m]^T \in \mathbb{Z}^m$, with $h_j(\xi) = \widehat{g}_i(\xi)$. This part of the theorem is proved by setting $\beta_j(n) = h_j(n)$.

The proof of (2) follows directly from the fact that the density of $\tau_\Delta(X_j)$ is $\sum_{n\in\mathbb{Z}} h_j(t - \Delta n)$ for $t \in \mathcal{I}_\Delta$. ■

Proposition 3.2 puts strong constraints on the distribution of $\mathbf{x}$ for the **WNH** to hold. Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with joint density $f(\mathbf{x})$. Let $\{\mathbf{v}_j\}_{j=1}^d$ be linearly independent, and let $\mathbf{Y} = [\mathbf{X} \cdot \mathbf{v}_1, \mathbf{X} \cdot \mathbf{v}_2, \dots, \mathbf{X} \cdot \mathbf{v}_d]^T$. Then the joint density of $\mathbf{Y}$ is $g(\mathbf{y}) = |\det(F)|^{-1} f\left((F^T)^{-1}\mathbf{y}\right)$ where $F = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$. Thus, both the independence and the identical distribution assumptions in the **WNH**, even for $N = d$, will be false unless very exact conditions are met. For instance, if we take $\mathbf{X}$ to be Gaussian and F to be unitary, then the independence property is satisfied only when F diagonalizes the covariance matrix of $\mathbf{X}$.

Corollary 3.3. *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with joint density $f(\mathbf{x})$ and $\{\mathbf{v}_j\}_{j=1}^d$ be linearly independent vectors in $\mathbb{R}^d$. Let $\mathbf{Y} = F^T\mathbf{X} = [\mathbf{X} \cdot \mathbf{v}_1, \dots, \mathbf{X} \cdot \mathbf{v}_N]^T$ and $g(\mathbf{y}) = |\det(F)|^{-1} f\left((F^T)^{-1}\mathbf{y}\right)$ where $F = [\mathbf{v}_1, \dots, \mathbf{v}_d]$.*

(1) *$\{\tau_\Delta(Y_j)\}_{j=1}^d$ are independent random variables if and only if there exist complex numbers $\{\beta_j(n) : 1 \le j \le d, n \in \mathbb{Z}\}$ such that*
$$\widehat{g}\left(\frac{a_1}{\Delta}, \dots, \frac{a_d}{\Delta}\right) = \beta_1(a_1) \cdots \beta_d(a_d) \tag{3.3}$$
for all $[a_1, \dots, a_d]^T \in \mathbb{Z}^d$.

(2) *Let $h_j(t) = \int_{\mathbb{R}^{d-1}} g(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_d)\, dx_1 \cdots\, dx_{j-1}\, dx_{j+1} \dots\, dx_d$. Then $\{\tau_\Delta(X_j)\}_{j=1}^d$ are identically distributed if and only if $\sum_{n\in\mathbb{Z}} h_j(t - n\Delta) = H(t)$ a.e. for some $H(t)$ independent of j. They are uniformly distributed on $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ if and only if $H(t) = 1/\Delta$ a.e..*

Proof. We only have to observe that $g(\mathbf{y})$ is the density of $\mathbf{Y}$ and that h_j is the marginal density of Y_j. The corollary now follows directly from the theorem. ■

From a practical point of view, with coarse quantization the **MSE** cannot be estimated simply by (1.6). Thus the "6-dB-per-bit" rule may not apply. We shall demonstrate this with numerical results. However, with high resolution quantization the formula (1.6) becomes increasingly accurate. We show this in the next section.

4. Asymptotic Behavior of Errors: Linear Independence Case

In many practical applications such as music CD, fine quantizations with 16 bits or more have been adopted. Although the **WNH** is not valid in general, with fine quantizations we prove here that a weaker version of the **WNH** is close to being valid, which yields an asymptotic formula for the PCM quantized **MSE**. Our result here strengthens an asympototic result in [22].

We again consider the same setup as before. Let $\{\mathbf{v}_j\}_{j=1}^N$ be a frame in $\mathbb{R}^d$ with corresponding frame matrix $F = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N]$. The eigenvalues of FF^T are $\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d = \lambda_{\min} > 0$. Let $\{\mathbf{u}_j\}_{j=1}^N$ be the canonical dual frame with corresponding matrix $G = (FF^T)^{-1}F$. For any $\mathbf{x} \in \mathbb{R}^d$ we have $\mathbf{x} = \sum_{j=1}^N (\mathbf{x} \cdot \mathbf{v}_j)\, \mathbf{u}_j$. Using the quantization alphabet $\mathcal{A} = \Delta\mathbb{Z}$ we have the PCM reconstruction (1.4). Note that $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}(\Delta)$ as it depends on Δ. With the **WNH** we obtain the **MSE**

$$\mathbf{MSE} = \mathcal{E}\left(\|\mathbf{x} - \tilde{\mathbf{x}}\|^2\right) = \frac{\Delta^2}{12}\sum_{j=1}^N \lambda_j^{-1}.$$

To study the asymptotic behavior of the error components, we study as $\Delta \to 0^+$ the normalized quantization error

$$\frac{1}{\Delta}(\mathbf{x} - \tilde{\mathbf{x}}) = \sum_{j=1}^N \frac{1}{\Delta}\tau_\Delta\left(\mathbf{x} \cdot \mathbf{v}_j\right)\mathbf{u}_j. \tag{4.1}$$

Theorem 4.1. *Let $\mathbf{X} \in \mathbb{R}^d$ be an absolutely continuous random vector. Let $\mathbf{w}_1, \ldots, \mathbf{w}_m$ be linearly independent vectors in $\mathbb{R}^d$. Then*

$$\left[\frac{1}{\Delta}\tau_\Delta\left(\mathbf{X} \cdot \mathbf{w}_1\right), \ldots, \frac{1}{\Delta}\tau_\Delta\left(\mathbf{X} \cdot \mathbf{w}_m\right)\right]^T$$

converges in distribution as $\Delta \to 0^+$ to a random vector uniformly distributed in $[-1/2, 1/2]^m$.

Proof. Denote $Y_j = \mathbf{X} \cdot \mathbf{w}_j$. Since $\{\mathbf{w}_j\}$ are linearly independent, $\mathbf{Y} = [Y_1, \ldots, Y_m]^T$ is absolutely continuous with some joint density $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$. As a consequence of (3.2) one has that the distribution of $\mathbf{Z} = [Z_1, \ldots, Z_m]^T$, where $Z_j = \frac{1}{\Delta}\tau_\Delta(Y_j) = \left\{\frac{Y_j}{\Delta} + \frac{1}{2}\right\} - \frac{1}{2}$, is

$$f_\Delta(\mathbf{x}) := \Delta^m \sum_{\mathbf{a} \in \mathbb{Z}^m} f(\Delta\mathbf{x} - \Delta\mathbf{a}). \tag{4.2}$$

for $\mathbf{x} \in [-1/2, 1/2]^m$. Again denote $\mathcal{I}_1 := [-1/2, 1/2]$. It is easy to see that $\|f_\Delta\|_{L^1(\mathcal{I}_1^m)} \leq \|f\|_{L^1(\mathbb{R}^m)}$, for

$$\begin{aligned}
\|f_\Delta\|_{L^1(\mathcal{I}_1^m)} &= \int_{\mathcal{I}_1^m} |f_\Delta(\mathbf{x})| \, d\mathbf{x} \\
&\leq \sum_{\mathbf{a} \in \mathbb{Z}^m} \int_{\mathcal{I}_1^m} \Delta^m \left| f(\Delta\mathbf{x} - \Delta\mathbf{a}) \right| \, d\mathbf{x} \\
&= \sum_{\mathbf{a} \in \mathbb{Z}^m} \int_{\Delta\mathcal{I}_1{}^m + \Delta\mathbf{a}} \left| f\left(\mathbf{y}\right) \right| \, d\mathbf{y} \\
&= \int_{\mathbb{R}^m} \left| f\left(\mathbf{y}\right) \right| \, d\mathbf{y} \\
&= \|f\|_{L^1(\mathbb{R}^m)}.
\end{aligned}$$

Now, if $\Omega = [a_1, b_1] \times \cdots \times [a_m, b_m]$ and $f(\mathbf{x}) = \mathbf{1}_\Omega(\mathbf{x})$, then for $\mathbf{x} \in \mathcal{I}_1^m$ observe that $f_\Delta(\mathbf{x}) = \Delta^m K_\Delta$ where $K_\Delta(\mathbf{x}) = \#\{\mathbf{a} \in \mathbb{Z}^m : \Delta\mathbf{x} + \Delta\mathbf{a} \in \Omega\}$. Obviously, $K_\Delta(\mathbf{x}) = s/\Delta^m + O(\Delta^{-m+1})$ where $s = \mathcal{L}(\Omega)$ is the Lebesgue measure of Ω. Then $f_\Delta \to s\mathbf{1}_{\mathcal{I}_1^m}$ in $L^1(\mathcal{I}_1^m)$ as $\Delta \to 0^+$.

Coming back to the case when $f(\mathbf{x})$ is the density of $\mathbf{Y}$. For any $\varepsilon > 0$ it is possible to choose a $g(\mathbf{x}) \in L^1(\mathbb{R}^m)$ such that $\|f - g\|_{L^1} < \frac{\varepsilon}{3}$, and furthermore, $g(\mathbf{x}) = \sum_{j=1}^N c_j \mathbf{1}_{E_j}(\mathbf{x})$ is a simple function where $c_j \in \mathbb{R}$ and each E_j is a product of finite intervals. Observe that $\int_{\mathbb{R}^m} g = \sum_{j=1}^N c_j \mathcal{L}(E_j)$. Since $(\mathbf{1}_{E_j})_\Delta \to \mathcal{L}(E_j)\mathbf{1}_{\mathcal{I}_1^m}$ in L^1 we have $g_\Delta \to \left(\int_{\mathbb{R}^m} g\right) \mathbf{1}_{\mathcal{I}_1^m}$ as $\Delta \to 0$. Hence there exists a $\delta > 0$ such that $\left\| g_\Delta - \left(\int_{\mathbb{R}^m} g\right) \mathbf{1}_{\mathcal{I}_1^m} \right\|_{L^1} < \varepsilon/3$ whenever $\Delta < \delta$. Now, for $\Delta < \delta$,

$$\begin{aligned}
\left\| f_\Delta - \mathbf{1}_{\mathcal{I}_1^m} \right\|_{L^1(\mathcal{I}_1^m)} &= \|f_\Delta - g_\Delta\|_{L^1(\mathcal{I}_1^m)} + \left\| g_\Delta - \left(\int_{\mathbb{R}^m} g\right) \mathbf{1}_{\mathcal{I}_1^m} \right\|_{L^1(\mathcal{I}_1^m)} \\
&\quad + \left| 1 - \left(\int_{\mathbb{R}^m} g\right) \right| \|\mathbf{1}_{\mathcal{I}_1^m}\|_{L^1(\mathcal{I}_1^m)} \\
&< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \left| 1 - \left(\int_{\mathbb{R}^m} g\right) \right| \\
&= \frac{2\varepsilon}{3} + \left| \left(\int_{\mathbb{R}^m} f\right) - \left(\int_{\mathbb{R}^m} g\right) \right| \\
&< \varepsilon.
\end{aligned}$$

■

Remark: We in fact proved a stronger result, namely the densities converge in L^1. Applying the above theorem to the **MSE**, if $\{\mathbf{v}_j\}_{j=1}^N$ are pairwise linearly independent then the error components $\{\tau_\Delta(\mathbf{X}\cdot\mathbf{v}_j)\}_{j=1}^N$ become asymptotically pairwise independent and each uniformly distributed in $[-\frac{\Delta}{2},\frac{\Delta}{2}]$.

Corollary 4.2. *Let $\mathbf{X}\in\mathbb{R}^d$ be an absolutely continuous random vector. If $\{\mathbf{v}_j\}_{j=1}^N$ are pairwise linearly independent, then as $\Delta\to 0^+$ we have*

$$\mathcal{E}\left(\|\mathbf{X}-\tilde{\mathbf{X}}\|^2\right)=\frac{\Delta^2}{12}\sum_{j=1}^d\lambda_j^{-1}+o(\Delta^2)=\frac{\Delta^2}{12}\sum_{j=1}^N\|\mathbf{u}_j\|^2+o(\Delta^2). \tag{4.3}$$

Proof. As usual denote by F and G the frame matrices associated with the frame $\{\mathbf{v}_j\}_{j=1}^N$ and the dual frame $\{\mathbf{u}_j\}_{j=1}^N$, respectively. Let $H=G^TG$, $Y_j=\mathbf{X}\cdot\mathbf{v}_j$, $Z_j=\left\{\frac{Y_j}{\Delta}+\frac{1}{2}\right\}-\frac{1}{2}$, and $\mathbf{Z}=[Z_1,\ldots,Z_N]^T$. By Theorem 4.1, $\mathcal{E}(Z_i)\to 0$ and $\mathcal{E}(Z_iZ_j)\to\frac{1}{12}\delta_{ij}$ as $\Delta\to 0^+$. Now $\mathbf{X}-\tilde{\mathbf{X}}=G\mathbf{Z}$. It follows from the proof of Proposition 2.2 that

$$\begin{aligned}\frac{1}{\Delta^2}\mathcal{E}(\|\mathbf{X}-\tilde{\mathbf{X}}\|^2) &= \mathcal{E}(\mathbf{Z}^TH\mathbf{Z})\\ &= \mathcal{E}\left(\sum_{i,j=1}^N Z_iZ_jh_{ij}\right)\\ &= \sum_{i,j=1}^N h_{ij}\mathcal{E}(Z_iZ_j)\\ &= \frac{1}{12}\sum_{i=1}^N h_{ii}+o(1)\\ &= \frac{1}{12}\sum_{j=1}^d\lambda_j^{-1}+o(1),\end{aligned}$$

and hence

$$\mathcal{E}\left(\|\mathbf{X}-\tilde{\mathbf{X}}\|^2\right)=\frac{\Delta^2}{12}\sum_{j=1}^d\lambda_j^{-1}+o(\Delta^2)=\frac{\Delta^2}{12}\sum_{j=1}^N\|\mathbf{u}_j\|^2+o(\Delta^2).$$

■

5. Asymptotic Behavior of Errors: Linear Dependence Case

In this section we consider the case in which some vectors in the frame may be parallel. This can happen, for example, if the frame contains redundant elements. Mathematically it would be interesting to understand how the **MSE** behaves as $\Delta\to 0^+$. We return to previous calculations and note that

$$\mathcal{E}(\|\mathbf{X}-\tilde{\mathbf{X}}\|^2) = \sum_{i,j=1}^{N} h_{ij}\mathcal{E}\left(\tau_\Delta(\mathbf{X}\cdot\mathbf{v}_i)\tau_\Delta(\mathbf{X}\cdot\mathbf{v}_j)\right).$$

Our main result in this section is:

Theorem 5.1. *Let X be an absolutely continuous real random variable. Let $\alpha \in \mathbb{R}\setminus\{0\}$. Then*

$$\lim_{\Delta\to 0^+}\frac{1}{\Delta^2}\mathcal{E}\left(\tau_\Delta(X)\tau_\Delta(\alpha X)\right) = \begin{cases} 0, & \alpha\notin\mathbb{Q}, \\ \dfrac{1}{12pq}, & \alpha=\dfrac{p}{q} \text{ and } p+q \text{ is even,} \\ -\dfrac{1}{24pq}, & \alpha=\dfrac{p}{q} \text{ and } p+q \text{ is odd,} \end{cases} \tag{5.1}$$

where p, q are coprime integers.

Proof. Denote $g(x) := \{x+\frac{1}{2}\} - \frac{1}{2}$. Let $\phi(x)\geq 0$ be an even C^∞ function such that $\operatorname{supp}(\phi)\subseteq[-1,1]$ and $\int_{\mathbb{R}}\phi = 1$. Let $g_n(x) = g*\phi_n$ where $\phi_n(x) = n\phi(nx)$. It is standard to check that

(a) $|g_n(x)|\leq 1/2$;
(b) $\operatorname{supp}(g(x)-g_n(x))\subseteq[\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n}]+\mathbb{Z}$;
(c) $g_n(x)\in C^\infty$, and is $\mathbb{Z}$-periodic;
(d) $\int_{\mathbb{R}} g_n(x)\,dx = 0$.

$g_n(x)$ represents a small perturbation of $g(x)$ that “smoothes out” the discontinuities of $g(x)$. Now, set

$$\begin{aligned} E(\Delta) &:= \mathcal{E}\left(\frac{1}{\Delta^2}\tau_\Delta(X)\tau_\Delta(\alpha X)\right) \\ &= \mathcal{E}\left(g\left(\frac{X}{\Delta}\right)g\left(\frac{\alpha X}{\Delta}\right)\right) \\ &= \int_{\mathbb{R}} g\left(\frac{x}{\Delta}\right)g\left(\frac{\alpha x}{\Delta}\right)f(x)\,dx, \end{aligned}$$

and

$$E_n(\Delta) := \int_{\mathbb{R}} g_n\left(\frac{x}{\Delta}\right)g_n\left(\frac{\alpha x}{\Delta}\right)f(x)\,dx.$$

Claim: $E_n(\Delta)\to E(\Delta)$ *as $n\to\infty$ uniformly for all $\Delta>0$.*

Proof of the Claim. Let f be the density of $\mathbf{X}$. For any $\varepsilon > 0$,

$$\begin{aligned} |E_n(\Delta) - E(\Delta)| &= \left|\int_{\mathbb{R}} \left[g_n\left(\frac{x}{\Delta}\right) g_n\left(\frac{\alpha x}{\Delta}\right) - g\left(\frac{x}{\Delta}\right) g\left(\frac{\alpha x}{\Delta}\right)\right] f(x)\, dx\right| \\ &\le \frac{1}{2}\int_{\mathbb{R}} \left|g_n\left(\frac{x}{\Delta}\right) - g\left(\frac{x}{\Delta}\right)\right| f(x)\, dx + \frac{1}{2}\int_{\mathbb{R}} \left|g_n\left(\frac{\alpha x}{\Delta}\right) - g\left(\frac{\alpha x}{\Delta}\right)\right| f(x)\, dx. \end{aligned}$$

Now there exists an $M > 0$ such that $\int_{[-M,M]^c} f(x)\, dx < \frac{\varepsilon}{2}$. So

$$\int_{\mathbb{R}} \left|g_n\left(\frac{x}{\Delta}\right) - g\left(\frac{x}{\Delta}\right)\right| f(x)\, dx \le \int_{-M}^{M} \left|g_n\left(\frac{x}{\Delta}\right) - g\left(\frac{x}{\Delta}\right)\right| f(x)\, dx + \frac{\varepsilon}{2}.$$

Furthermore, let $A_n(\Delta, M) := \operatorname{supp}(g_n(x/\Delta) - g(x/\Delta)) \cap [-M, M]$. Then we have

$$A_n(\Delta, M) \subseteq \Delta\Big([\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}] + \mathbb{Z}\Big) \cap [-M, M].$$

Hence $\mathcal{L}(A_n(\Delta, M)) \le \frac{2M}{\Delta} \cdot \frac{2\Delta}{n} = \frac{4M}{n}$, and thus

$$\int_{-M}^{M} \left|g_n\left(\frac{x}{\Delta}\right) - g\left(\frac{x}{\Delta}\right)\right| f(x)\, dx \le \int\limits_{A_n(\Delta,M)} f(x)\, dx < \frac{\varepsilon}{2}$$

for $n > 4M/\varepsilon$ (which is independent of Δ). This yields

$$\int_{\mathbb{R}} \left|g_n\left(\frac{x}{\Delta}\right) - g\left(\frac{x}{\Delta}\right)\right| f(x)\, dx < \varepsilon.$$

Similarly we have

$$\int_{\mathbb{R}} \left|g_n\left(\frac{\alpha x}{\Delta}\right) - g\left(\frac{\alpha x}{\Delta}\right)\right| f(x)\, dx < \varepsilon$$

for sufficiently large n, proving the Claim. □

Now consider the Fourier series of $g_n(t)$,

$$g_n(t) = \sum_{k \in \mathbb{Z}} c_k^{(n)} e^{2\pi i k t}.$$

It is well known that the Fourier series converges to $g_n(t)$ uniformly for all t, see e.g. [24]. Furthermore, since $g_n(t)$ is C^∞ we have $|c_k^{(n)}| = o\left((|k| + 1)^{-L}\right)$ for all $L > 0$, giving absolute convergence of the Fourier series. Thus

$$\begin{aligned} E_n(\Delta) &= \lim_{K \to \infty} \int_{\mathbb{R}} \Big(\sum_{|k| \le K} c_k^{(n)} e^{2\pi i k t \Delta^{-1}}\Big)\Big(\sum_{|k| \le K} c_k^{(n)} e^{2\pi i k \alpha t \Delta^{-1}}\Big) f(t)\, dt \\ &= \lim_{K \to \infty} \sum_{|k|, |\ell| \le K} c_k^{(n)} c_\ell^{(n)} \widehat{f}\Big(-\frac{k + \alpha \ell}{\Delta}\Big). \end{aligned}$$

Observe that $|\widehat{f}(\xi)| \leq \|f\|_{L^1} = 1$, and $|c_k^{(n)}| = o\big((|k|+1)^{-L}\big)$ for any $L > 0$. So the series converges absolutely and uniformly in Δ. Thus

$$E_n(\Delta) = \sum_{k,\ell\in\mathbb{Z}} c_k^{(n)} c_\ell^{(n)} \widehat{f}\Big(-\frac{k+\alpha\ell}{\Delta}\Big). \tag{5.2}$$

For any $n > 0$ we have

$$\lim_{\Delta\to 0^+} E_n(\Delta) = \sum_{k,\ell\in\mathbb{Z}} c_k^{(n)} c_\ell^{(n)} \lim_{\Delta\to 0^+} \widehat{f}\Big(-\frac{k+\alpha\ell}{\Delta}\Big)$$

because the series converges absolutely and uniformly. Suppose $\alpha \notin \mathbb{Q}$. Then $k + \alpha\ell \neq 0$ if either $k \neq 0$ or $\ell \neq 0$. Thus $\big|-\frac{k+\alpha\ell}{\Delta}\big| \to \infty$ as $\Delta \to \infty$, and hence $\lim_{\Delta\to 0^+} \widehat{f}\big(-\frac{k+\alpha\ell}{\Delta}\big) = 0$ as $f \in L^1(\mathbb{R})$. Note also that $c_0^{(n)} = \int_{\mathbb{R}} g_n = 0$. It follows that

$$\lim_{\Delta\to 0^+} E_n(\Delta) = 0.$$

But $E_n(\Delta) \to E(\Delta)$ as $n \to \infty$ uniformly in Δ, which yields $E(\Delta) \to 0$ as $\Delta \to 0^+$.

Next, suppose $\alpha = \frac{p}{q}$ where $p, q \in \mathbb{Z}$, $(p,q) = 1$. We observe that $k + \alpha\ell = 0$ if and only if $k = pm$ and $\ell = -qm$ for some $m \in \mathbb{Z}$. In such a case

$$\widehat{f}\Big(-\frac{k+\alpha\ell}{\Delta}\Big) = \widehat{f}(0) = \int_{\mathbb{R}} f = 1.$$

It follows that

$$\lim_{\Delta\to 0^+} E_n(\Delta) = \sum_{m\in\mathbb{Z}} c_{pm}^{(n)} c_{-qm}^{(n)} \widehat{f}(0) = \sum_{m\in\mathbb{Z}} c_{pm}^{(n)} c_{-qm}^{(n)} = \sum_{m\in\mathbb{Z}} c_{pm}^{(n)} \overline{c_{qm}^{(n)}}.$$

For $r \in \mathbb{Z}, r \neq 0$ set

$$G_r^{(n)}(x) := \sum_{m\in\mathbb{Z}} c_{rm}^{(n)} e^{2\pi i m x}.$$

By Parseval we have

$$\lim_{\Delta\to 0} E_n(\Delta) = \Big\langle G_q^{(n)}, G_p^{(n)} \Big\rangle_{L^2([0,1])}.$$

It is easy to check that

$$G_r^{(n)} = \frac{1}{|r|} \sum_{j=0}^{|r|-1} g_n\Big(\frac{x+j}{r}\Big).$$

Hence $G_r^{(n)}$ converges in $L^2([0,1])$ to $G_r(x) := \frac{1}{|r|}\sum_{j=0}^{|r|-1} g\big(\frac{x+j}{r}\big)$, which has Fourier series $G_r(x) = \sum_{m\in\mathbb{Z}} c_{rm} e^{2\pi i m x}$ with $c_0 = 0$ and $c_k = \frac{(-1)^{k-1}}{2\pi i k}$ for $k \neq 0$. This yields

$$\lim_{n\to\infty} \lim_{\Delta\to 0^+} E_n(\Delta) = \lim_{n\to\infty} \Big\langle G_q^{(n)}, G_p^{(n)} \Big\rangle = \langle G_q, G_p \rangle = \sum_{m\in\mathbb{Z}} c_{qm} \overline{c_{pm}}.$$

Finally

$$\begin{aligned}\sum_{m\in\mathbb{Z}} c_{qm}\overline{c_{pm}} &= \sum_{m\in\mathbb{Z}\setminus\{0\}} \Big(\frac{(-1)^{qm-1}}{2\pi imq}\Big)\overline{\Big(\frac{(-1)^{pm-1}}{2\pi imp}\Big)} \\ &= \frac{1}{2pq\pi^2}\sum_{m=1}^{\infty}\frac{(-1)^{(p+q)m}}{m^2}.\end{aligned}$$

Note that if $p+q$ is even then $\sum_{m=1}^{\infty}\frac{(-1)^{(p+q)m}}{m^2}=\sum_{m=1}^{\infty}\frac{1}{m^2}=\frac{\pi^2}{6}$. On the other hand, if $p+q$ is odd then $\sum_{m=1}^{\infty}\frac{(-1)^{(p+q)m}}{m^2}=\sum_{m=1}^{\infty}\frac{(-1)^m}{m^2}=-\frac{\pi^2}{12}$. The theorem follows. ■

Corollary 5.2. *Let* $\mathbf{X}$ *be an absolutely continuous random vector in* $\mathbb{R}^d$, $\mathbf{w}\neq 0$, $\mathbf{w}\in\mathbb{R}^d$ *and* $\alpha\in\mathbb{R}\setminus\{0\}$. *Then*

$$\lim_{\Delta\to 0^+}\frac{1}{\Delta^2}\mathcal{E}\left(\tau_\Delta(\mathbf{w}\cdot\mathbf{X})\tau_\Delta(\alpha\mathbf{w}\cdot\mathbf{X})\right)=\begin{cases}0, & \alpha\notin\mathbb{Q},\\ \dfrac{1}{12pq}, & \alpha=\dfrac{p}{q}\ \textit{and } p+q \textit{ is even},\\ -\dfrac{1}{24pq}, & \alpha=\dfrac{p}{q}\ \textit{and } p+q \textit{ is odd},\end{cases}\tag{5.3}$$

where p,q *are coprime integers.*

Proof. We only need to note that $\mathbf{w}\cdot\mathbf{X}$ is an absolutely continuous random variable. The corollary follows immediately from Theorem 5.1. ■

We can now characterize completely the asymptotic bahavior of the MSE in all cases. For any two vectors $\mathbf{w}_1,\mathbf{w}_2\in\mathbb{R}^d$ define $r(\mathbf{w}_1,\mathbf{w}_2)$ by

$$r(\mathbf{w}_1,\mathbf{w}_2)=\begin{cases}\dfrac{1}{pq}\mathbf{w}_1\cdot\mathbf{w}_2, & \mathbf{w}_1=\dfrac{p}{q}\mathbf{w}_2,\ \text{and } p+q \text{ is even},\\ -\dfrac{1}{2pq}\mathbf{w}_1\cdot\mathbf{w}_2, & \mathbf{w}_1=\dfrac{p}{q}\mathbf{w}_2,\ \text{and } p+q \text{ is odd},\\ 0, & \text{otherwise},\end{cases}$$

where p,q are coprime integers.

Corollary 5.3. *Let* $\mathbf{X}\in\mathbb{R}^d$ *be an absolutely continuous random vector. Then as* $\Delta\longrightarrow 0^+$ *the* **MSE** *satisfies*

$$\mathcal{E}\left(\|\mathbf{X}-\tilde{\mathbf{X}}\|^2\right)=\frac{\Delta^2}{12}\sum_{j=1}^{d}\lambda_j^{-1}+\frac{\Delta^2}{6}\sum_{1\le i<j\le N}r(\mathbf{u}_i,\mathbf{u}_j)+o(\Delta^2),\tag{5.4}$$

Proof. In the proof of (4.2) we showed that

$$\lim_{\Delta\to 0^+}\frac{1}{\Delta^2}\mathcal{E}\left(\|\mathbf{X}-\tilde{\mathbf{X}}\|^2\right)=\sum_{i,j}h_{ij}\mathcal{E}\left(Z_iZ_j\right)$$

with the notations there. Observe that $h_{ij}=\mathbf{u}_i\cdot\mathbf{u}_j$. The result is immediate from Corollary 5.2. ■

For fixed quantization step $\Delta>0$ we shall denote

$$\mathbf{MSE}_{ideal}=\frac{\Delta^2}{12}\sum_{j=1}^{d}\lambda_j^{-1}+\frac{\Delta^2}{6}\sum_{1\le i<j\le N}r(\mathbf{u}_i,\mathbf{u}_j), \tag{5.5}$$

and call it the *ideal **MSE***. If $\{\mathbf{v}_j\}_{j=1}^N$ are pairwise linearly independent, then the $\mathbf{MSE}_{ideal}$ is simply $\frac{\Delta^2}{12}\sum_{j=1}^d\lambda_j^{-1}$, the **MSE** under the **WNH**.

We should point out that even though the **WNH** is not true aysmpototically if some vectors in a frame are parallel, the contribution from the second part of (5.5) is often small enough that the **MSE** under the **WNH** is close enough to the ideal **MSE**. In the next section we shall show some numerical data, comparing the actual **MSE** with the ideal **MSE**.

Appendix. Numerical Results

Here we present data from our computer experiments comparing the ideal **MSE** to the actual **MSE**. We have performed Monte Carlo simulations for several different sets of frames. We also experimented with various distributions for $\mathbf{X}\in\mathbb{R}^d$. As it turns out, we get very similar results for the distributions we used for most of the frames we tried. In the examples shown, the random vectors $\mathbf{X}$ are all chosen to be uniformly distributed in $[-5,5]^d$.

Example 5.1. *Let $\{\mathbf{v}_j\}_{j=1}^N$ be the harmonic frame in $\mathbb{R}^2$, with $\mathbf{v}_j=\left[\cos\frac{2\pi j}{N},\sin\frac{2\pi j}{N}\right]^T$. This is a tight frame with frame constant $\lambda=\frac{N}{2}$. The ideal **MSE** is $\frac{\Delta^2}{3N}$ for N odd. Taking $\Delta=\frac{1}{2}$, Table 1 displays the actual **MSE**, the ideal **MSE** and the ratio between them. It shows that as N gets larger than 129, the actual **MSE** does not improve, which shows that the WNH is invalid for large Δ.*

N	Actual **MSE**	Ideal **MSE**	*Ratio*
9	0.00934342	0.00925926	1.009090
17	0.00479521	0.00252525	0.976808
33	0.00246669	0.00490196	0.978223
65	0.00122499	0.00128205	0.955496
129	0.00065858	0.000645995	1.019480
257	0.00057971	0.00032425	1.787810
513	0.00056039	0.00016244	3.449740
1025	0.00052914	0.00008130	6.508450
2049	0.00053895	0.00004067	13.25180
4097	0.00058846	0.00002034	28.93090

TABLE 1. The Harmonic frame in $\mathbb{R}^2$

Example 5.2. *Let* $\{\mathbf{v}_j\}_{j=1}^N$ *be* N *independently and randomly generated vectors uniformly distributed on the unit sphere in* $\mathbb{R}^4$*. Table 2 shows the ratio between the actual **MSE** and the ideal **MSE**, where* $\mathbf{MSE}_{ideal} = \frac{\Delta^2}{12}(\sum_{j=1}^d \lambda_j^{-1})$*, with* $\Delta = 2^{-k}$*.*

k/N	$N = 64$	$N = 128$	$N = 256$	$N = 512$	$N = 1024$
k= 0	1.581960	2.232260	3.697160	6.497800	12.20670
k= 1	1.076590	1.130510	1.397840	1.649530	2.480920
k= 2	1.003680	0.995214	1.008370	1.033280	1.196680
k= 3	0.967138	0.990876	0.999648	0.981633	1.010090
k= 4	0.989295	1.009840	1.032110	1.002630	1.002260
k= 5	1.011720	1.035590	1.020870	1.002350	1.022250
k= 6	0.978712	1.006760	0.992207	1.001490	0.979342
k= 7	0.997524	1.017840	0.995852	0.972120	0.976273
k= 8	0.998725	1.011380	1.040270	0.978204	0.973284
k= 9	0.982450	1.038580	0.994463	1.021580	1.037800
k=10	0.993099	1.002340	1.009930	1.009870	0.974017
k=11	0.981428	0.998280	0.975881	1.049010	1.009570

TABLE 2. The randomly generated frame in $\mathbb{R}^4$

Example 5.3. *Let* $\{\mathbf{v}_j\}_{j=0}^{N-1}$ *be the harmonic frame in* $\mathbb{R}^4$*, with*

$$\mathbf{v}_j = \sqrt{\frac{1}{2}}\left[\cos\frac{2\pi j}{N}, \sin\frac{2\pi j}{N}, \cos\frac{4\pi j}{N}, \sin\frac{4\pi j}{N}\right]^T.$$

This is a tight frame with frame constant $\lambda = \dfrac{N}{4}$*, and the ideal **MSE** is* $\dfrac{4\Delta^2}{3N}$*. Table 3 shows the ratio between the actual **MSE** and the ideal **MSE** where* $\Delta = 2^{-k}$*.*

k/N	$N = 64$	$N = 128$	$N = 256$	$N = 512$	$N = 1024$
k= 0	0.997218	0.928318	1.287990	2.312710	4.497050
k= 1	1.005460	1.004720	0.950783	1.339810	2.395180
k= 2	0.990253	1.001070	0.977474	0.960994	1.354320
k= 3	0.995848	0.993963	0.981683	0.992655	0.955345
k= 4	0.987371	1.007310	1.028120	1.016760	1.002570
k= 5	0.993840	1.015230	1.026680	1.003770	1.023820
k= 6	1.012230	1.012280	0.996363	0.999742	1.004120
k= 7	1.020450	1.025820	1.031120	1.003770	1.004770
k= 8	1.004710	1.010820	0.999289	0.973596	0.970415
k= 9	0.993542	1.003380	0.981550	0.984594	0.981001
k=10	1.015610	1.008740	0.997469	0.986705	1.004360
k=11	1.010690	1.009080	0.994975	1.010510	0.998485

Table 3. The Harmonic frame in $\mathbb{R}^4$

Example 5.4. *Let $\{\mathbf{v}_j\}_{j=1}^5$ be a frame in $\mathbb{R}^3$, with the corresponding matrix*

$$F = \begin{pmatrix} 1 & 1 & 0 & 1 & 3 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Note that the set contains many parallel vectors. The dual frame matrix is

$$\begin{pmatrix} \frac{1}{11} & 0 & 0 & \frac{1}{11} & \frac{3}{11} \\ \frac{1}{11} & -1 & 0 & \frac{1}{11} & \frac{3}{11} \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

The ***MSE*** *under the WNH is* $0.181818\Delta^2$ *and by our result, the ideal* ***MSE*** *is* $0.190083\Delta^2$ *which is closer to the actual* ***MSE****. The difference between the two estimates comes from the second part in (5.5). Table 4 shows the actual* ***MSE****, the ideal* ***MSE****, and the* ***MSE*** *under the WNH, where $\Delta = 2^{-k}$.*

References

[1] J. Benedetto and M. Fickus, Finite normalized tight frames, *Advances in Computational Mathematics*, **18**, (2003) 357–385.

[2] J. Benedetto, A. M. Powell, and Ö. Yılmaz, Sigma-Delta ($\Sigma\Delta$) quantization and finite frames, *IEEE Trans. Inform. Theory*, **52** (2006), no. 5, 1990–2005.

[3] J. Benedetto, A. M. Powell, and Ö. Yılmaz, Second order sigma-delta ($\Sigma\Delta$) quantization of finite frame expansions, *Appl. Comput. Harmon. Anal.*, **20**, (2006), no. 1, 126–148.

[4] W. Bennett, Spectra of quantized signals, *Bell Syst.Tech.J* **27** (1948) 446–472.

[5] S. Bochner and K. Chandrasekharan, *Fourier Transforms*, Princeton University Press, 1949.

[6] P. G. Casazza and J. Kovačević, Uniform tight frames with erasures, *Advances in Computational Mathematics*, **18**, (2003) 387–430.

k	Actual **MSE**	Ideal **MSE**	**MSE** Under WNH
2	0.012234100000	0.011880200000	0.011363600000
3	0.002935150000	0.002970040000	0.002840910000
4	0.000732567000	0.000742510000	0.000710227000
5	0.000188331000	0.000185628000	0.000177557000
6	0.000046664900	0.000046406900	0.000044389200
7	0.000011626300	0.000001160170	0.000011097300
8	0.000002953720	0.000002900430	0.000002774330
9	0.000000724800	0.000000725108	0.000000693581
10	0.000000180127	0.000000181277	0.000000173395
11	0.000000045856	0.000000045319	0.000000043349

TABLE 4. The frame of Example 5.4 in $\mathbb{R}^3$

[7] 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 Math.*, **158** (2003), no. 2, 679–710.

[8] R. J. Duffin and A. C. Schaeffer, A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, **72** (1952), 341–366.

[9] Y. Elder and G. D. Forney, Optimal tight frames and quantum measurement, *IEEE Trans. Inform. Theory*, **48** (2002), no. 3, 599–610.

[10] D. J. Feng, L. Wang, and Y. Wang, Generation of finite tight frames by Householder transformations, *Advances in Computational Mathematics*, **24** (2006), 297–309.

[11] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer, Boston (1992).

[12] V. K. Goyal, J. Kovačcević, and J. Kelner, Quantized frame expansions with erasures, *Appl. Comput. Harmon. Anal.*, **10**, (2001) 203–233.

[13] V. K. Goyal, M. Vetterli and N. T. Thao, Quantized overcomplete expansions in $\mathbb{R}^N$: analysis, synthesis, and algorithms, *IEEE Trans. Inform. Theory*, **44** (1998), 16–31.

[14] R. M. Gray and D. L. Neuhoff, Quantization, *IEEE Trans. Inform. Theory*, **44** (1998), 2325–2383.

[15] R. Gray, Quantized noise spectra, *IEEE Trans. Inform. Theory*, **36** (1990), no. 6, 1220–1244.

[16] S. Güntürk, Approximating a bandlimited function using very coarsely quantized data, *J. Amer. Math. Soc.*, **17** (2004), no. 1, 229–242.

[17] Y. Katznelson, *Harmonic Analysis*, Wiley Inc, New York, 1968.

[18] T. Linder, R. Zamir and K. Zeger, High resolution source coding for nondifference distortion measure: multidimensional companding, *IEEE Trans. Inform. Theory*, **45** (1999), no. 2, 548–561.

[19] S. Na and D. L. Neuhoff, Bennett's integral for vector quantizers, *IEEE Trans. Inform. Theory*, **41** (1995), no. 4, 886–900.

[20] G. Rath and C. Guillemot, Recent advances in DFT codes based on quantized finite frame expansions for erasure channels, *Digital Signal Processing*, **14** (4) (2004) 332-354.

[21] N. Thao and M. Vetterli, Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimate, *IEEE Trans. Signal Proc.*, **42** (1994), no. 3, 519–531.

[22] H. Viswanathan and R. Zamir, On the whiteness of high-resolution quantization errors, *IEEE Trans. Inform. Theory*, **47** (2001), 2029–2038.

[23] R. Zamir and M. Feder, On lattice quantization noise, *IEEE Trans. Inform. Theory*, **42** (1996), 1152–1159.

[24] A. Zygmund, *Trigonometrical Series*, Chelsea Publishing Company, New York, 1950.

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.
E-mail address: djimenez@math.gatech.edu

Department of Mathematics, Southern Polytechnic State University, Marietta, GA 30065
E-mail address: lwang@spsu.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.
E-mail address: wang@math.gatech.edu