FINITE REPRESENTATION OF FINITE ENERGY SIGNALS

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND

ELECTRONICS ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCES

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Talha Cihad Gülcü

July 2011

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Haldun M. Özaktaş(Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Erdal Arıkan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Metin Gürses

Approved for the Graduate School of Engineering and Sciences:

Prof. Dr. Levent Onural
Director of Graduate School of Engineering and Sciences

ABSTRACT

FINITE REPRESENTATION OF FINITE ENERGY SIGNALS

Talha Cihad Gülcü
M.S. in Electrical and Electronics Engineering
Supervisor: Prof. Dr. Haldun M. Özaktaş
July 2011

In this thesis, we study how to encode finite energy signals by finitely many bits. Since such an encoding is bound to be lossy, there is an inevitable reconstruction error in the recovery of the original signal. We also analyze this reconstruction error. In our work, we not only verify the intuition that finiteness of the energy for a signal implies finite degree of freedom, but also optimize the reconstruction parameters to get the minimum possible reconstruction error by using a given number of bits and to achieve a given reconstruction error by using minimum number of bits. This optimization leads to a number of bits vs reconstruction error curve consisting of the best achievable points, which reminds us the rate distortion curve in information theory. However, the rate distortion theorem are not concerned with sampling, whereas we need to take sampling into consideration in order to reduce the finite energy signal we deal with to finitely many variables to be quantized. Therefore, we first propose a finite sample representation scheme and question the optimality of it. Then, after representing the signal of interest by finite number of samples at the expense of a certain error, we discuss several quantization methods for these finitely many samples and compare their performances.

Keywords: Finite Energy Signals, Sampling, Finite Sample Representation, Degree of Freedom (DOF), Space Bandwidth Product, Reconstruction Error, Uniform Quantization, Vector Quantization, Quantization Error, Rate Distortion Theory

ÖZET

SONLU ENERJİLİ SİNYALLERİN SONLU GÖSTERİMİ

Talha Cihad Gülcü
Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans
Tez Yöneticisi: Prof. Dr. Haldun M. Özaktaş
Temmuz 2011

Bu tezde, sonlu enerjili sinyallerin sonlu sayıda ikil(bit) ile nasıl kodlanılacağı çalışılmaktadır. Böyle bir kodlama kayıpsız olamayacağı için, asıl sinyalin yeniden elde edilmesinde kaçınılmaz bir yeniden kurma hatası olmaktadır. Bu yeniden kurma hatası da burada analiz edilmektedir. Bu çalışmada, sadece bir sinyal için enerji sonluluğunun sonlu erkinlik derecesine işaret edeceği sezgisi doğrulanmamakta, ayrıca belli sayıda ikil kullanarak mümkün olan en az yeniden kurma hatasını elde etmek ve en az ikil kullanarak belli bir yeniden kurma hatasını başarmak için yeniden kurma değiştirgeleri de eniyileştirilmektedir. Bu en iyileme, bilgi kuramındaki oran bozulma eğrisini anımsatan, en iyi elde edilebilir noktalardan oluşan bir ikil sayısına karşı yeniden kurma hatası eğrisi getirmektedir. Ancak, oran bozulma teoremi örneklemeyi konu edinmemektedir, oysa ki bu çalışmada sözkonusu sonlu enerjili sinyalin nicemlenecek sonlu sayıda değişkene indirgenmesi adına örneklemenin dikkate alınması gerekmektedir. Bu nedenle, ilk olarak, bir sonlu örnek gösterim tasarısı önerilmekte ve bunun eniyiliği sorgulanmaktadır. Belli bir hata karşılığında, sözkonusu sinyali sonlu sayıda örnek ile temsil ettikten sonra, bu sonlu sayıda örnek için, değişik nicemleme yöntemleri tartışılmakta ve performansları karşılaştırılmaktadır.

Anahtar Kelimeler: Sonlu Enerjili Sinyaller, Örnekleme, Sonlu Örnek Gösterimi, Erkinlik Derecesi, Uzam Bant Genişliği Çarpımı, Yeniden Kurma Hatası, Tekbiçimli Nicemleme, Yöney Nicemlemesi, Nicemleme Hatası, Oran Bozulma Kuramı

ACKNOWLEDGMENTS

I would like to thank Prof. Dr. Haldun M. Özaktaş for his valuable guidance and contributions throughout this study. In addition, I would like to thank Prof. Dr. Erdal Arıkan and Prof. Dr. Metin Gürses for their constructive comments and advices. Moreover, I acknowledge the support of TUBITAK through a graduate scholarship. Finally, I would like to thank my parents and my brother for their invaluable support throughout my life.

Contents

List of Figures

List of Tables

Dedicated to my family

Chapter 1

INTRODUCTION

In this thesis, we are concerned with the problem of encoding finite energy signals by finite number of bits, which was originated from [1, 2]. This problem has two main parts: Sampling and quantization.

Sampling is a well established topic of signal processing. Nyquist [3] and Shannon [4] set the foundations of sampling by proving the classic well-known uniform sampling theorem for bandlimited signals. Actually, this theorem was previously introduced in several works [5,6]. Sampling theorem for bandlimited processes is considered in [7]. Various extensions of Shannon-Nyquist sampling theorem, such as sampling for functions of more than one variable, random processes, nonuniform sampling, nonbandlimited signals, are presented in [8]. Sampling theory of nonbandlimited signals is reviewed in [9]. An error analysis for nonuniform sampling of nonbandlimited signals is provided in [10]. Reconstruction error for the uniform sampling of nonbandlimited signals is considered in [11].

More recent review articles on sampling are [12, 13]. The main focus of [12] is uniform(regular) sampling. In [13], the topics such as reconstruction of nonbandlimited signals and stability of reconstruction are reviewed.

[14–32] are some of the works in which nonuniform(irregular) sampling is taken into account. Instead of sinc function in reconstrunction, wavelets [33–46] and splines [47–59] are considered in numerous works. We use regular sampling and the usual sinc interpolation of samples in this work, because the expression of the resultant reconstruction error provides us useful interpretations in this case. An error analysis for the reconstruction method we cover is given in [60]. The formulation of bandlimited signal interpolation as a linear estimation problem is given in [61].

Quantization is a fundamental subject of signal processing as well. In earlier works, fixed rate scalar quantization [62–66] and scalar quantization with memory [67–71] are considered. Shannon's well known 1948 paper [72] paved the way for variable rate quantization. Later on, in his landmark paper [73] published in 1959, Shannon introduced rate distortion theory and motivated vector quantization. After Shannon's 1959 paper, different kinds of vector quantizers are proposed [74–79]. Lattice quantizers [79–82], product quantizers [83–85], tree structured quantization [86, 87], multistage vector quantization [88, 89] and feedback vector quantization [90–92] are some of the quantization methods available in the literature. [93–96] are some of the more recent works on quantization. The whole history of quantization is reviewed in [97] in detail. We employ both uniform scalar quantization and vector quantization in this work.

Before encoding finite energy signals, we represent them with finitely many samples as an intermediate step. The finite sample representation subject we cover here is closely related to the concepts such as degree of freedom (DOF) and space-bandwidth product. The number-of-degrees-of-freedom concept is considered in different contexts in the literature [98–108].

Actually, signal encoding is covered in a couple of books [109, 110]. In [109], time-continuous stationary source encoding is considered. But, we focus on finite energy time-continuous sources in this thesis, and a finite energy signal cannot

be stationary. Autoregressive nonstationary source encoding is also discussed in [109]. However, for signal encoding, the units of rate and distortion are always taken as per second in [109], whereas in this work, we aim to encode time-continuous sources by finitely many bits at the expense of a finite overall error. On the other hand, in [110], different waveform coding techniques, such as delayed decision coding, subband coding, transform coding, are treated. However, similar to [109], in [110], rate is always taken as bits per second or bits per sample, and error variance or SNR is considered as the quantity to be minimized . In this work, we are not interested in the error variance at a certain sample or the number of bits used per sample. What we are interested in is the number of bits used to encode the whole signal, and the associated error in reconstructing it. Thus, our problem formulation is quite different from [109, 110].

Throughout our work, we will first consider a single deterministic complex function(signal) having finite energy, i.e.,

$$\int_{-\infty}^{\infty} |f(u)|^2 \, du < \infty \tag{1.1}$$

and extend our results wherever applicable to a class of signals which will be denoted by $\mathcal{F}$. Once the signal to be represented by finitely many samples or bits is known, there is no point in representing it. Therefore, we need to generalize our results to the case when there is more than one signal possible to be encountered.

By assigning a probability to each member of a signal class $\mathcal{F}$, we can model $\mathcal{F}$ as a random process. Some of our results will require the energy of the signals in $\mathcal{F}$ to be upperbounded. Whereas our other results will simply require that the expectation of energy (average energy), namely

$$E\left[\int_{-\infty}^{\infty} |f(u)|^2 \, du\right] = \int_{-\infty}^{\infty} E[|f(u)|^2] \, du \tag{1.2}$$

is finite. Note that we are able to change the order of the integration and expectation in (1.2) thanks to Fubini's theorem [111], since the integrand $|f(u)|^2$

is nonnegative. In this work, we have changed the order of the integration and expectation several times, and this justification is applicable to all those changes of order.

In Chapter 2, we first propose a method based on Δu truncation in space domain and $\Delta\mu$ truncation in frequency domain to reconstruct any finite energy signal by using only finitely many samples of it and analyze the corresponding finite sample reconstruction error. Then, we simplify the finite sample reconstruction error expression and choose the finite sample reconstruction parameters Δu and $\Delta\mu$ optimally to minimize it and to obtain the number of samples vs finite sample reconstruction error Pareto optimal curve. Moreover, the form that error takes when antialiasing filter is not used is also investigated. Lastly, the connections between our work and the results on prolate spheroidal functions in the literature are discussed.

In Chapter 3, different quantization techniques on the finitely many samples that the finite energy signal is reduced to are considered. Firstly, the scalar K level uniform quantization of as many as $\Delta u\Delta\mu$ samples is discussed, and a vector quantization method is proposed to improve the quantization performance. Then, for the vector quantization, the parameters that the number of bits and finite bit reconstruction error depend on, namely Δu, $\Delta\mu$ and K, are optimized, which makes it possible to get the number of bits vs error Pareto optimal curve. Another quantization technique outperforming this vector quantization is also considered in Chapter 3. Finally, rate distortion theorem is adapted to our setup to obtain the best achievable performance. The conclusions and future works are listed in Chapter 4.

In this thesis, the domain of the signals can be taken as space or time. In other words, for the signals $f(u)$ considered throughout this work, the unit of u can be taken as second or meter. We will denote the unit of u as s wherever needed. Throughout our work, the terminology of space domain (the words such

as space limited, space-bandwidth product, spatial truncation, spatial width etc.) is preferred instead of that of time domain. Moreover, the unit of the values that signals take can be volts or volts per meter. We will denote the unit of $f(u)$ as Φ wherever needed.

Integrals whose limits are not given will signify integrals from minus to plus infinity. Throughout this work, signals will be denoted by f and their Fourier transforms will be denoted by F. Moreover, vectors and matrices will be denoted by boldface letters.

List of symbols is given in Table 1.1 and list of operator and function notations is given in Table 1.2.

Symbol	Explanation
$\mathbb{Z}$	the set of integers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^+$	the set of nonnegative real numbers
$\mathbb{C}$	the set of complex numbers
$f : A \to B$	f is a function with domain A, range B
$A \times B$	the set of pairs (a, b) such that $a \in A$, $b \in B$
$[a, b]$	the set of real numbers r satisfying $a \leq r \leq b$
j	the imaginary number $\sqrt{-1}$
e	the natural number $2.7183\ldots$
π	the pi number $3.14159\ldots$
δ_{mn}	Kronecker delta
$n!$	n factorial, i.e., $1 \times 2 \times \cdots \times n$
$\min\{a, b\}$	the smaller one of the real numbers a and b
$\min_S g$	the minimum value that g takes on the set S
$\mathrm{diag}\{a_1, \ldots, a_n\}$	diagonal matrix having $\{a_1, \ldots, a_n\}$ on its diagonal

Table 1.1: List of symbols

Operator&Function	Explanation
$\mathrm{Re}\{.\}$	real part of
$\mathrm{Im}\{.\}$	imaginary part of
$\lvert . \rvert$	absolute value
$E[.]$	expectation value
$\lfloor r \rfloor$	largest integer less than or equal to r
$\langle ., . \rangle$	inner product
$(.)^*$	conjugate
$(.)^{\mathrm{T}}$	matrix transpose
$\mathrm{tr}(.)$	trace of the matrix
$\lvert\lvert . \rvert\rvert_2^2$	square of the Euclidean norm of the vector
$\ln$	natural logarithm
$\log_2$	base 2 logarithm
$\mathrm{sinc}(x)$	$\sin(\pi x)/(\pi x)$
$\mathrm{rect}(x)$	rectangle function
$Q(x)$	$\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt$

Table 1.2: List of operator and function notations

Chapter 2

FINITE SAMPLE REPRESENTATION

In this chapter, we present a method to represent any finite energy signal by finite number of samples. Then, we show that the reconstruction error can be made arbitrarily small by choosing the number of samples large enough. After proving that the finite sample reconstruction error can be made as small as desired, we approximate this error by a suitable term, and optimize the spatial width Δu and the spectral width $\Delta \mu$ so that the number of bits vs reconstruction error curve consisting of Pareto optimal points is obtained. The Pareto optimal curves corresponding to certain autocorrelation functions are also provided. Moreover, the reconstruction error for the case when antialiasing filter is not used is analyzed as well. Finally, some topics about our finite sample reconstruction error are discussed in the light of the works on prolate spheroidal functions.

We begin our discussion by analyzing the spatial and spectral truncation error for a finite energy signal. In this analysis, the only assumption we have is that the energy of the signal of interest is finite. The results we obtain will be used

later to show that the reconstruction error corresponding to the finite sample representation we suggest can be made arbitrarily small.

2.1 Spatial and Spectral Truncation Error

Let $f(u)$ be a single finite energy signal, i.e. a signal satisfying (1.1). Although it is very natural to say "Let the spatial width of $f(u)$ be Δu and the frequency(spectral) width of $f(u)$ be $\Delta\mu$", there is something hidden in this statement: Truncation error. A signal cannot be both space limited and frequency limited at the same time. Therefore, in either spatial or spectral truncation, there is a deviation from the original signal. However, both spatial and spectral truncation errors can be made arbitrarily small by selecting the truncation interval sufficiently large, as we will show.

Let $\tilde{f}_{\Delta u}(u)$ denote the result of spatial truncation to the interval $[-\Delta u/2, \Delta u/2]$, namely

$$\tilde{f}_{\Delta u}(u) = \begin{cases} f(u) & \text{if } |u| \leq \Delta u/2, \\ 0 & \text{else .} \end{cases} \tag{2.1}$$

Then, the spatial truncation error $\int |f(u) - \tilde{f}_{\Delta u}(u)|^2 \, du$ can be expressed as

$$\begin{aligned} \int |f(u) - \tilde{f}_{\Delta u}(u)|^2 \, du &= \int_{-\infty}^{-\Delta u/2} |f(u)|^2 \, du + \int_{\Delta u/2}^{\infty} |f(u)|^2 \, du & (2.2) \\ &= \int |f(u)|^2 \, du - \int_{-\Delta u/2}^{\Delta u/2} |f(u)|^2 \, du & (2.3) \\ &= \int |f(u)|^2 \, du - \int |\tilde{f}_{\Delta u}(u)|^2 \, du & (2.4) \end{aligned}$$

From Lebesgue monotone convergence theorem [111], we have

$$\lim_{\Delta u \to \infty} \int |\tilde{f}_{\Delta u}(u)|^2 \, du = \int |f(u)|^2 \, du \tag{2.5}$$

Using (2.5) with (2.4), we obtain

$$\lim_{\Delta u \to \infty} \int |f(u) - \tilde{f}_{\Delta u}(u)|^2 \, du = 0 \tag{2.6}$$

Therefore, the spatial truncation error can be made as small as desired by selecting Δu large enough. A similar fact is also valid for spectral truncation error, as we will explain.

If the original function $f(u)$ is truncated to the frequency band $[-\Delta\mu/2,\, \Delta\mu/2]$, denoting the output of bandlimiting operation as $\breve{f}_{\Delta\mu}(u)$, from Parseval's theorem, we have

$$\int |f(u) - \breve{f}_{\Delta\mu}(u)|^2 \, du = \int |F(\mu) - \breve{F}_{\Delta\mu}(\mu)|^2 \, d\mu \tag{2.7}$$

where F and $\breve{F}_{\Delta\mu}$ refer to the Fourier transforms of f and $\breve{f}_{\Delta\mu}$, respectively. Then, we obtain

$$\begin{aligned} \int |f(u) - \breve{f}_{\Delta\mu}(u)|^2 \, du &= \int |F(\mu)|^2 \, d\mu - \int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu)|^2 \, d\mu & (2.8) \\ &= \int |F(\mu)|^2 \, d\mu - \int |\breve{F}_{\Delta\mu}(\mu)|^2 \, d\mu & (2.9) \end{aligned}$$

Using Lebesgue monotone convergence theorem once again, we get

$$\lim_{\Delta\mu\to\infty} \int |\breve{F}_{\Delta\mu}(\mu)|^2 \, d\mu = \int |F(\mu)|^2 \, d\mu \tag{2.10}$$

From (2.9) and (2.10), similar to spatial truncation case considered above, we conclude

$$\lim_{\Delta\mu\to\infty} \int |f(u) - \breve{f}_{\Delta\mu}(u)|^2 \, du = 0 \tag{2.11}$$

Hence the spectral truncation error $\int |f(u) - \breve{f}_{\Delta\mu}(u)|^2 \, du$ can be made arbitrarily small by choosing $\Delta\mu$ sufficiently large.

Now, if $f(u)$ is a random process having finite expectation of energy, similarly we have

$$\begin{aligned} &\lim_{\Delta u\to\infty} E\left[\int |f(u) - \tilde{f}_{\Delta u}(u)|^2 \, du\right] \\ &\quad = \lim_{\Delta u\to\infty} \left[\int E[|f(u)|^2] \, du - \int E[|\tilde{f}_{\Delta u}(u)|^2] \, du\right] = 0 \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} &\lim_{\Delta\mu\to\infty} E\left[\int |f(u) - \breve{f}_{\Delta\mu}(u)|^2 \, du\right] \\ &\quad = \lim_{\Delta\mu\to\infty} \left[\int E[|F(\mu)|^2] \, d\mu - \int E[|\breve{F}_{\Delta\mu}(\mu)|^2] \, d\mu\right] = 0 \end{aligned} \tag{2.13}$$

as the stochastic counterparts of (2.6) and (2.11), respectively. Thus, in this case, the spatial truncation error $E[\int |f(u) - \tilde{f}_{\Delta u}(u)|^2 \, du]$ and the spectral truncation error $E[\int |f(u) - \breve{f}_{\Delta \mu}(u)|^2 \, du]$ can be made arbitrarily small by choosing Δu and $\Delta \mu$ large enough, respectively.

The results given up to here will be used to analyze the reconstruction error of the finite sample representation scheme we will cover now.

2.2 Finite Sample Reconstruction and its Error Analysis

In this section, we will propose an approach to represent a finite energy signal $f(u)$ by finite number of samples and analyze the associated finite sample reconstruction error.

As commonly known, $\mathbb{R}$ and any interval $[a, b]$ in it consists of uncountably many elements. Therefore, even if the signal $f(u)$ can be truncated in spatial or spectral domain, still there will be uncountably many number of points belonging to the support of the signal. We cannot use all of the uncountable number of data if we want to eventually get a finite sample representation. Thus, *sampling* is a required part of the job. Sampling can be performed either in spatial or spectral domain.

Secondly, there is no assumption on (spatial or spectral) bandwidth of $f(u)$. Therefore, sampling is expected to result in *aliasing* problem, which may cause extra error. Hence, we may need an *antialiasing filter* to have a more accurate reconstruction. Thus, we have two options:

1. Filtering in spectral domain first, then taking samples in spatial domain.

2. Filtering in spatial domain first, then taking samples in spectral domain.

The second option for finite sample representation can be analyzed similar to the first option and will be mentioned briefly wherever applicable throughout our work. Moreover, the finite sample representation without antialiasing filter is analyzed in Section 2.4.

Now, we begin to explain our finite sample representation (will be abbreviated as FSR from now on) scheme by taking the first option described above into consideration. After truncating $f(u)$ to a two-sided bandwidth of $\Delta\mu$ in spectral domain, from Nyquist and Shannon's sampling theorem, the resultant bandlimited signal can be expressed as

$$\breve{f}_{\Delta\mu}(u) = \sum_{n=-\infty}^{\infty} \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{2.14}$$

To have a FSR, we discard all the samples except for the ones lying in the interval $[-\Delta u/2,\, \Delta u/2]$ and obtain the signal

$$\hat{f}_{\Delta u,\Delta\mu}(u) = \sum_{n=-\lfloor \Delta u\Delta\mu/2 \rfloor}^{\lfloor \Delta u\Delta\mu/2 \rfloor} \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{2.15}$$

which can be characterized completely by

$$2\left\lfloor \frac{\Delta u\Delta\mu}{2} \right\rfloor + 1 \approx \Delta u\Delta\mu \tag{2.16}$$

number of samples. These samples constitute the vector

$$\mathbf{f} = \left(\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \middle| -\left\lfloor \frac{\Delta u\Delta\mu}{2} \right\rfloor \le n \le \left\lfloor \frac{\Delta u\Delta\mu}{2} \right\rfloor \right) \tag{2.17}$$

denoting the FSR of $f(u)$.

The finite sample reconstruction signal $\hat{f}_{\Delta u,\Delta\mu}(u)$ has a bandwidth $\Delta\mu$ and an approximate spatial width Δu. Note that we have $\Delta u\Delta\mu \gg 1$ in practice, thus the approximation made in (2.16) is reasonable. Thus, the degree-of-freedom (will be abbreviated as DOF from now on) for $\hat{f}_{\Delta u,\Delta\mu}(u)$ is approximately its space-bandwidth product $\Delta u\Delta\mu$.

Now, we analyze the error in reconstructing $f(u)$ as $\hat{f}_{\Delta u,\Delta\mu}(u)$. As an intermediate step, we first calculate the truncation error $e_{tr}(\Delta u, \Delta\mu)$ made by discarding

all but $2\lfloor \Delta u \Delta\mu/2 \rfloor + 1$ samples to get $\hat{f}_{\Delta u,\Delta\mu}(u)$ from $\breve{f}_{\Delta\mu}(u)$. Since the set

$$\{\mathrm{sinc}(\Delta\mu\, u - n) \,|\, n \in \mathbb{Z}\} \tag{2.18}$$

consists of orthogonal functions each having an energy of $1/\Delta\mu$ (can be seen very easily using the fact that Fourier transform preserves the inner product, that is

$$\begin{aligned}
&\langle \mathrm{sinc}(\Delta\mu\, u - n), \mathrm{sinc}(\Delta\mu\, u - m)\rangle \\
&\quad = \left\langle \frac{1}{\Delta\mu}\,\mathrm{rect}\left(\frac{\mu}{\Delta\mu}\right) e^{-j\,2\pi\frac{n}{\Delta\mu}\mu}, \frac{1}{\Delta\mu}\,\mathrm{rect}\left(\frac{\mu}{\Delta\mu}\right) e^{-j\,2\pi\frac{m}{\Delta\mu}\mu} \right\rangle \\
&\quad = \frac{1}{\Delta\mu}\delta_{mn}
\end{aligned} \tag{2.19}$$

and the result follows.), we have

$$e_{tr}(\Delta u, \Delta\mu) = \int |\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du = \frac{1}{\Delta\mu} \sum_{|n|>\lfloor \Delta u\Delta\mu/2 \rfloor} \left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2 \tag{2.20}$$

Note that the energy of $\breve{f}_{\Delta\mu}$ cannot exceed that of f, which is finite by assumption. Thus, using the orthogonality of the sincs again, we conclude

$$\int |\breve{f}_{\Delta\mu}(u)|^2\, du = \frac{1}{\Delta\mu} \sum_{n=-\infty}^{\infty} \left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2 < \infty \tag{2.21}$$

Then, from (2.20) and (2.21), we get

$$\lim_{\Delta u\to\infty} e_{tr}(\Delta u, \Delta\mu) = 0 \tag{2.22}$$

On the other hand, in order to express the finite sample reconstruction error in a more explicit form, we first write

$$\begin{aligned}
|f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2 &= |(f(u) - \breve{f}_{\Delta\mu}(u)) + (\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u))|^2 \\
&= |f(u) - \breve{f}_{\Delta\mu}(u)|^2 \\
&\quad + 2\,\mathrm{Re}\{(f(u) - \breve{f}_{\Delta\mu}(u))(\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u))^*\} \\
&\quad + |\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2
\end{aligned} \tag{2.23}$$

Then, from (2.23), the finite sample reconstruction error can be expressed as

$$\begin{aligned}\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du &= \int |f(u) - \breve{f}_{\Delta\mu}(u)|^2\, du \\ &\quad + 2\,\mathrm{Re}\{\langle f(u) - \breve{f}_{\Delta\mu}(u),\, \breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)\rangle\} \\ &\quad + \int |\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du \qquad (2.24)\end{aligned}$$

Since Fourier transform preserves inner product, we have

$$\langle f(u) - \breve{f}_{\Delta\mu}(u),\, \breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)\rangle = \langle F(\mu) - \breve{F}_{\Delta\mu}(\mu),\, \breve{F}_{\Delta\mu}(\mu) - \hat{F}_{\Delta u,\Delta\mu}(\mu)\rangle \qquad (2.25)$$

By definition, $\breve{F}_{\Delta\mu}(\mu)$ is identical to $F(\mu)$ at $[-\Delta\mu/2, \Delta\mu/2]$, thus $F(\mu) - \breve{F}_{\Delta\mu}(\mu)$ is zero in this frequency band. On the other hand, as (2.15) implies, $\hat{F}_{\Delta u,\Delta\mu}(\mu)$ is zero outside $[-\Delta\mu/2, \Delta\mu/2]$ as well as $\breve{F}_{\Delta\mu}(\mu)$. Hence, $\breve{F}_{\Delta\mu}(\mu) - \hat{F}_{\Delta u,\Delta\mu}(\mu)$ is nonzero only at $[-\Delta\mu/2, \Delta\mu/2]$. Then, we conclude

$$\begin{aligned}&\langle F(\mu) - \breve{F}_{\Delta\mu}(\mu),\, \breve{F}_{\Delta\mu}(\mu) - \hat{F}_{\Delta u,\Delta\mu}(\mu)\rangle \\ &= \int_{-\Delta\mu/2}^{\Delta\mu/2} (F(\mu) - \breve{F}_{\Delta\mu}(\mu))(\breve{F}_{\Delta\mu}(\mu) - \hat{F}_{\Delta u,\Delta\mu}(\mu))^*\, d\mu \\ &\quad + \int_{|\mu|>\Delta\mu/2} (F(\mu) - \breve{F}_{\Delta\mu}(\mu))(\breve{F}_{\Delta\mu}(\mu) - \hat{F}_{\Delta u,\Delta\mu}(\mu))^*\, d\mu \\ &= 0 + 0 = 0 \qquad (2.26)\end{aligned}$$

Therefore, (2.24) can be simplified as

$$\begin{aligned}\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du &= \int |f(u) - \breve{f}_{\Delta\mu}(u)|^2\, du \\ &\quad + \int |\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du \qquad (2.27) \\ &= \int |f(u) - \breve{f}_{\Delta\mu}(u)|^2\, du + e_{tr}(\Delta u, \Delta\mu) \qquad (2.28)\end{aligned}$$

Then, combining (2.28) with (2.11) and (2.22), we conclude

$$\lim_{\Delta u,\Delta\mu\to\infty} \int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du = 0 \qquad (2.29)$$

Therefore, the reconstruction error of the FSR we propose can be made as small as desired by selecting Δu and $\Delta\mu$, namely the two parameters product of which give the number of DOF for the reconstruction signal $\hat{f}_{\Delta u,\Delta\mu}(u)$, large enough.

To obtain an alternative FSR, one can consider confining $\hat{f}_{\Delta u,\Delta\mu}(u)$ to the interval $[-\Delta u/2, \Delta u/2]$ in space domain. However, the analysis of the finite sample reconstruction error as carried out here seems to be difficult to handle in this case.

On the other hand, as mentioned at the beginning of this section, there is a second option to obtain a FSR. In this option, we first truncate $f(u)$ to the space interval $[-\Delta u/2, \Delta u/2]$, and from Nyquist and Shannon's sampling theorem, we express the Fourier transform of the resultant spacelimited signal $\tilde{f}_{\Delta u}(u)$ as

$$\tilde{F}_{\Delta u}(\mu) = \sum_{n=-\infty}^{\infty} \tilde{F}_{\Delta u}\left(\frac{n}{\Delta u}\right) \operatorname{sinc}(\Delta u\, \mu - n) \tag{2.30}$$

Then, we only keep the samples in the frequency band $[-\Delta\mu/2, \Delta\mu/2]$ and obtain the signal

$$\hat{F}_{\Delta u,\Delta\mu}(\mu) = \sum_{n=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \tilde{F}_{\Delta u}\left(\frac{n}{\Delta u}\right) \operatorname{sinc}(\Delta u\, \mu - n) \tag{2.31}$$

the inverse Fourier transform $\hat{f}_{\Delta u,\Delta\mu}(u)$ of which is the FSR signal of the second option, having a spatial width Δu, an approximate bandwidth $\Delta\mu$, an approximate space-bandwidth product and the number of DOF $\Delta u \Delta\mu$. Here, please note that $\hat{f}_{\Delta u,\Delta\mu}(u)$ we mention here is different from $\hat{f}_{\Delta u,\Delta\mu}(u)$ defined in (2.15) and used up to this point. $\hat{f}_{\Delta u,\Delta\mu}(u)$ of the second option is spacelimited, whereas $\hat{f}_{\Delta u,\Delta\mu}(u)$ of the first option is bandlimited. On the other hand, these two functions are close to each other as much as Uncertainty Principle permits, and the samples used to construct them are not the exact DFT of each other.

For this second option, we define $e_{tr}(\Delta u, \Delta\mu)$ as

$$e_{tr}(\Delta u, \Delta\mu) = \int |\tilde{f}_{\Delta u}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du = \int |\tilde{F}_{\Delta u}(\mu) - \hat{F}_{\Delta u,\Delta\mu}(\mu)|^2\, d\mu \tag{2.32}$$

By following the same argument that leads to (2.20), one can show that

$$e_{tr}(\Delta u, \Delta\mu) = \frac{1}{\Delta u} \sum_{|n|>\lfloor \Delta u \Delta\mu/2 \rfloor} \left|\tilde{F}_{\Delta u}\left(\frac{n}{\Delta u}\right)\right|^2 \tag{2.33}$$

and conclude

$$\lim_{\Delta\mu\to\infty} e_{tr}(\Delta u, \Delta\mu) = 0 \tag{2.34}$$

Moreover, the counterpart of (2.28), namely the equation

$$\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du = \int |f(u) - \tilde{f}_{\Delta u}(u)|^2\, du + e_{tr}(\Delta u, \Delta\mu) \tag{2.35}$$

can be derived similarly. Then, from (2.6), (2.34), and (2.35), we find that (2.29) is also valid for the second option. Therefore, this option makes it possible as well to obtain arbitrarily small finite sample reconstruction errors by choosing Δu and $\Delta\mu$ sufficiently large.

Now, consider a class of signals $\mathcal{F}$ each member of which has finite energy. Then, as (2.11) implies, for any fixed $\epsilon_1 > 0$, and for any chosen $f(u) \in \mathcal{F}$, there exists some bandwidth $\Delta\mu$ depending on the chosen signal $f(u)$ such that $\int |f(u) - \breve{f}_{\Delta\mu}(u)|^2\, du < \epsilon_1$. If the maximum of all these $\Delta\mu$ values exist, then for all $f(u) \in \mathcal{F}$, and for this maximum $\Delta\mu$, we have $\int |f(u) - \breve{f}_{\Delta\mu}(u)|^2\, du < \epsilon_1$. Similarly, as (2.22) implies, for any fixed $\epsilon_2 > 0$ and $\Delta\mu$ (in particular for the maximum $\Delta\mu$ we defined), for any chosen $f(u) \in \mathcal{F}$, there exists another Δu depending on the chosen signal $f(u)$ such that $e_{tr}(\Delta u, \Delta\mu) < \epsilon_2$. If the maximum of all these Δu values exist, then for all $f(u) \in \mathcal{F}$, and for this maximum Δu, we have $e_{tr}(\Delta u, \Delta\mu) < \epsilon_2$. Hence, from (2.28), we see that the worst case finite sample reconstruction error for $\mathcal{F}$ is $\epsilon_1 + \epsilon_2$, and thus can be made arbitrarily small, provided that the maximum $\Delta\mu$ and Δu described above exists for all $\epsilon_1, \epsilon_2 > 0$. A similar argument is obviously valid for the FSR of the second option. However, the condition that we require here to make sure that worst case error can be made as small as desired is difficult to be satisfied. Because, even if either the maximum Δu or maximum $\Delta\mu$ does not exist for a single nonzero ϵ_1 and ϵ_2, the condition is violated.

There is no need to make any assumptions on the existence of the maximum Δu or $\Delta\mu$ if *average error* is considered instead of *worst case error*, as we will show. Now, we define the signal class $\mathcal{F}$ we deal with as a random process $f(u)$,

and instead of requiring all the signals in $\mathcal{F}$ (all the realizations of $f(u)$, in the language of random processes) to have finite energy, we only assume that the average energy as given in (1.2) is finite. Then, taking the expectation of both sides in (2.28) and using (2.20), we get

$$\begin{aligned} E\left[\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right] &= E\left[\int |f(u)-\breve{f}_{\Delta\mu}(u)|^2\,du\right] \\ &\quad + \frac{1}{\Delta\mu}\sum_{|n|>\lfloor \Delta u\Delta\mu/2\rfloor} E\left[\left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2\right] \end{aligned} \tag{2.36}$$

Since the average energy of $\breve{f}_{\Delta\mu}(u)$ cannot exceed that of $f(u)$, which we assume to be finite, similar to (2.21), we have

$$E\left[\int |\breve{f}_{\Delta\mu}(u)|^2\,du\right] = \frac{1}{\Delta\mu}\sum_{n=-\infty}^{\infty} E\left[\left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2\right] < \infty \tag{2.37}$$

From (2.37), we obtain

$$\lim_{\Delta u\to\infty}\left\{\frac{1}{\Delta\mu}\sum_{|n|>\lfloor \Delta u\Delta\mu/2\rfloor} E\left[\left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2\right]\right\} = 0 \tag{2.38}$$

Using (2.13) and (2.38) in (2.36), we conclude

$$\lim_{\Delta u,\Delta\mu\to\infty} E\left[\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right] = 0 \tag{2.39}$$

which completes the proof of the fact that the average finite sample reconstruction error $E[\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du]$ can be made arbitrarily small by choosing Δu and $\Delta\mu$ sufficiently large.

Now, if the second option is considered for FSR, similar to (2.36) and (2.38), we have

$$\begin{aligned} E\left[\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right] &= E\left[\int |f(u)-\tilde{f}_{\Delta u}(u)|^2\,du\right] \\ &\quad + \frac{1}{\Delta u}\sum_{|n|>\lfloor \Delta u\Delta\mu/2\rfloor} E\left[\left|\tilde{F}_{\Delta u}\left(\frac{n}{\Delta u}\right)\right|^2\right] \end{aligned} \tag{2.40}$$

and

$$\lim_{\Delta\mu\to\infty}\left\{\frac{1}{\Delta u}\sum_{|n|>\lfloor \Delta u\Delta\mu/2\rfloor} E\left[\left|\tilde{F}_{\Delta u}\left(\frac{n}{\Delta u}\right)\right|^2\right]\right\} = 0 \tag{2.41}$$

respectively. Using (2.12) and (2.41) in (2.40), we conclude that (2.39) is also true for this option. Therefore, the second option for FSR makes it possible as well to obtain arbitrarily small average finite sample reconstruction errors by choosing Δu and $\Delta\mu$ large enough.

2.3 A Useful Approximation of Finite Sample Reconstruction Error

In Section 2.2, we found that finite sample reconstruction error can be written as (2.28) for the first FSR option and as (2.35) for the second FSR option. In this section, we will focus on the term $e_{tr}(\Delta u, \Delta\mu)$ which denotes the error made by discarding all the samples except for finitely many of them. At the end, we will show that, for both of the FSR options, finite sample reconstruction error can be approximated as the sum of the spatial truncation error (2.3) and the spectral truncation error (2.8).

As given in (2.20), for the first FSR option, the error made by ignoring the samples outside the interval $[-\Delta u/2, \Delta u/2]$ can be expressed as

$$e_{tr}(\Delta u, \Delta\mu) = \int |\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2 \, du = \frac{1}{\Delta\mu} \sum_{|n|>\lfloor \Delta u \Delta\mu/2 \rfloor} \left| \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \right|^2 \tag{2.42}$$

Since $\breve{f}_{\Delta\mu}(u)$ is bandlimited to $[-\Delta\mu/2, \Delta\mu/2]$, it does not increase or decrease significantly during a length of $1/\Delta\mu$. Thus, we have

$$\frac{1}{\Delta\mu} \sum_{|n|>\lfloor \Delta u \Delta\mu/2 \rfloor} \left| \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \right|^2 \approx \int_{|u|>\frac{\lfloor \Delta u \Delta\mu/2 \rfloor}{\Delta\mu}} |\breve{f}_{\Delta\mu}(u)|^2 \, du \tag{2.43}$$

$$\approx \int_{|u|>\Delta u/2} |\breve{f}_{\Delta\mu}(u)|^2 \, du \tag{2.44}$$

The approximation (2.44) can also be justified as follows: In practice, Δu is expected to be large enough so that $|\breve{f}_{\Delta\mu}(u)|^2$ is decreasing when $u > \frac{\lfloor \Delta u \Delta\mu/2 \rfloor}{\Delta\mu}$

and increasing when $u < -\frac{\lfloor \Delta u \Delta \mu/2 \rfloor}{\Delta \mu}$. Thus, we can write

$$\int_{|u|>\frac{\lfloor \Delta u \Delta \mu/2 \rfloor+1}{\Delta \mu}} |\breve{f}_{\Delta \mu}(u)|^2 \, du < e_{tr}(\Delta u, \Delta \mu) < \int_{|u|>\frac{\lfloor \Delta u \Delta \mu/2 \rfloor}{\Delta \mu}} |\breve{f}_{\Delta \mu}(u)|^2 \, du \tag{2.45}$$

Moreover, since $\Delta u \Delta \mu \gg 1$ in practice, we have

$$\frac{\lfloor \Delta u \Delta \mu/2 \rfloor + 1}{\Delta \mu} \approx \frac{\lfloor \Delta u \Delta \mu/2 \rfloor}{\Delta \mu} \approx \frac{\Delta u}{2} \tag{2.46}$$

and the result follows. Actually, it is proven in [117] that there exists some functions for which the approximation (2.44) is not valid. Nevertheless, (2.44) is a plausible approximation. For more details about this topic, see the discussion after Theorem 5 in Section 2.6.

Now, inserting (2.8) and (2.44) in (2.28), we get

$$\int |f(u) - \hat{f}_{\Delta u, \Delta \mu}(u)|^2 \, du \approx \int_{|\mu|>\Delta \mu/2} |F(\mu)|^2 \, d\mu + \int_{|u|>\Delta u/2} |\breve{f}_{\Delta \mu}(u)|^2 \, du \tag{2.47}$$

For the FSR of the second option, similarly we have

$$e_{tr}(\Delta u, \Delta \mu) = \frac{1}{\Delta u} \sum_{|n|>\lfloor \Delta u \Delta \mu/2 \rfloor} \left| \tilde{F}_{\Delta u} \left(\frac{n}{\Delta u} \right) \right|^2 \approx \int_{|\mu|>\Delta \mu/2} |\tilde{F}_{\Delta u}(\mu)|^2 \, d\mu \tag{2.48}$$

Then, combining (2.3) and (2.48) with (2.35), we obtain

$$\int |f(u) - \hat{f}_{\Delta u, \Delta \mu}(u)|^2 \, du \approx \int_{|u|>\Delta u/2} |f(u)|^2 \, du + \int_{|\mu|>\Delta \mu/2} |\tilde{F}_{\Delta u}(\mu)|^2 \, d\mu \tag{2.49}$$

For large enough Δu and $\Delta \mu$, we have

$$\int_{|u|>\Delta u/2} |\breve{f}_{\Delta \mu}(u)|^2 \, du \approx \int_{|u|>\Delta u/2} |f(u)|^2 \, du \tag{2.50}$$

$$\int_{|\mu|>\Delta \mu/2} |\tilde{F}_{\Delta u}(\mu)|^2 \, d\mu \approx \int_{|\mu|>\Delta \mu/2} |F(\mu)|^2 \, d\mu \tag{2.51}$$

Using (2.50) in (2.47) and using (2.51) in (2.49), for FSR of both first and second options, we obtain the following approximation

$$\int |f(u) - \hat{f}_{\Delta u, \Delta \mu}(u)|^2 \, du \approx \int_{|u|>\Delta u/2} |f(u)|^2 \, du + \int_{|\mu|>\Delta \mu/2} |F(\mu)|^2 \, d\mu \tag{2.52}$$

the right hand side (will be abbreviated as RHS from now on) of which is simply the sum of spatial and spectral truncation errors covered in the beginning of our work.

It is important to observe that the truncation made in the space and frequency domain directly appear in the approximate error expression (2.52) without any cross terms or amplification. This result is similar to the one obtined in [112], in which it was shown that the approximation error for the linear canonical transform computation algorithms proposed is basically determined by the error in approximating continuous Fourier transform by discrete Fourier transform (DFT), namely the error coming from the amount of energy contained outside the time-frequency region corresponding to the DFT applied.

From (2.52), we also conclude that, although $\hat{f}_{\Delta u,\Delta\mu}(u)$ of first and second options are different as explained previously, the finite sample reconstruction errors they result in are approximately the same and equal to the sum of spatial and spectral truncation errors if the FSR parameters Δu and $\Delta\mu$ are taken large enough.

For a random process $f(u)$, taking the expectation of both sides of (2.52), we get

$$E\left[\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right] \approx \int_{|u|>\Delta u/2} E[|f(u)|^2]\,du + \int_{|\mu|>\Delta\mu/2} E[|F(\mu)|^2]\,d\mu \tag{2.53}$$

In terms of the autocorrelation function of $f(u)$

$$R(u_1,u_2) = E[f(u_1)f^*(u_2)] \tag{2.54}$$

and the autocorrelation of the Fourier transform of $f(u)$

$$S(\mu_1,\mu_2) = \iint R(u_1,u_2)e^{-j2\pi\mu_1u_1}e^{j2\pi\mu_2u_2}\,du_1\,du_2 = E\left[F(\mu_1)F^*(\mu_2)\right] \tag{2.55}$$

(2.53) can be rewritten as

$$E\left[\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right] \approx \int_{|u|>\Delta u/2} R(u,u)\,du + \int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\,d\mu \tag{2.56}$$

Therefore, for a random process $f(u)$, the average finite sample reconstruction error can be approximated by the sum of the truncation errors of the diagonal of

its autocorrelation function and the diagonal of the autocorrelation of its Fourier transform.

2.4 Error Analysis for the Reconstruction Without Prefiltering

In this section, we will consider the case when the antialiasing filter is not used and the signal $f(u)$ is directly sampled and sinc interpolated. We will analyze the associated finite sample reconstruction error as done in Section 2.2 and derive an upperbound for it. This upperbound will be larger than (2.52). Note that, as found in Section 2.3, (2.52) is the form that reconstruction error for FSR with prefiltering takes when Δu and $\Delta\mu$ are large enough. The remaining part of this section is devoted to the details of the error upperbound derivation and can be omitted without loss of continuity.

Here, $f(u)$ is to be reconstructed as

$$\hat{f}_{\Delta u,\Delta\mu}(u) = \sum_{n=-\lfloor \Delta u\Delta\mu/2 \rfloor}^{\lfloor \Delta u\Delta\mu/2 \rfloor} f\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{2.57}$$

Note that, contrary to (2.15), the samples of the original signal $f(u)$ is used for sinc interpolation in (2.57) because prefiltering is not carried out for the reconstruction considered here.

The "second option" counterpart of this reconstruction signal would be the inverse Fourier transform of

$$\hat{F}_{\Delta u,\Delta\mu}(\mu) = \sum_{n=-\lfloor \Delta u\Delta\mu/2 \rfloor}^{\lfloor \Delta u\Delta\mu/2 \rfloor} F\left(\frac{n}{\Delta u}\right) \operatorname{sinc}(\Delta u\, \mu - n) \tag{2.58}$$

The analysis of the reconstructions described by (2.57) and (2.58) are nearly identical, therefore we continue our discussion from (2.57). Before proceeding,

we define another signal $\check{f}(u)$ as

$$\check{f}_{\Delta\mu}(u) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{2.59}$$

Note that, unlike $\breve{F}_{\Delta\mu}(\mu)$, the Fourier transform $\check{F}_{\Delta\mu}(\mu)$ of $\check{f}_{\Delta\mu}(u)$ does not agree with $F(\mu)$ on the interval $[-\Delta\mu/2, \Delta\mu/2]$ because of aliasing. Hence, unlike (2.26) and (2.27), we have

$$\langle F(\mu) - \check{F}_{\Delta\mu}(\mu),\, \check{F}_{\Delta\mu}(\mu) - \hat{F}_{\Delta u,\Delta\mu}(\mu)\rangle \neq 0 \tag{2.60}$$

$$\begin{aligned} \int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du \;\; &\neq \;\; \int |f(u) - \check{f}_{\Delta\mu}(u)|^2\, du \\ &\quad + \int |\check{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du \end{aligned} \tag{2.61}$$

Therefore, we need another approach to analyze the finite sample reconstruction error $\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du$. Here, we opt for the triangle inequality

$$\begin{aligned} \left(\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \;\; &\leq \;\; \left(\int |f(u) - \check{f}_{\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \\ &\quad + \left(\int |\check{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \end{aligned} \tag{2.62}$$

as the starting point of our error analysis.

Similar to (2.20), the equality

$$\int |\check{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du = \frac{1}{\Delta\mu} \sum_{|n|>\lfloor \Delta u \Delta\mu/2 \rfloor} \left| f\left(\frac{n}{\Delta\mu}\right)\right|^2 \tag{2.63}$$

is valid, and then (2.62) becomes

$$\begin{aligned} \left(\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \;\; &\leq \;\; \left(\int |f(u) - \check{f}_{\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \\ &\quad + \left(\frac{1}{\Delta\mu} \sum_{|n|>\lfloor \Delta u \Delta\mu/2 \rfloor} \left| f\left(\frac{n}{\Delta\mu}\right)\right|^2\right)^{\frac{1}{2}} \end{aligned} \tag{2.64}$$

By using Parseval's equality, (2.64) can be rewritten as

$$\begin{aligned} \left(\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \;\; &\leq \;\; \left(\int |F(\mu) - \check{F}_{\Delta\mu}(\mu)|^2\, d\mu\right)^{\frac{1}{2}} \\ &\quad + \left(\frac{1}{\Delta\mu} \sum_{|n|>\lfloor \Delta u \Delta\mu/2 \rfloor} \left| f\left(\frac{n}{\Delta\mu}\right)\right|^2\right)^{\frac{1}{2}} \end{aligned} \tag{2.65}$$

In order to analyze the term $\int |F(\mu) - \check{F}_{\Delta\mu}(\mu)|^2 \, d\mu$ apperaring in (2.65), we make use of Nyquist's sampling theorem to express $\check{F}_{\Delta\mu}(\mu)$ as

$$\check{F}_{\Delta\mu}(\mu) = \mathrm{rect}\left(\frac{\mu}{\Delta\mu}\right) \sum_{n=-\infty}^{\infty} F(\mu - \Delta\mu\, n) \tag{2.66}$$

Then, we get

$$\int |F(\mu) - \check{F}_{\Delta\mu}(\mu)|^2 \, d\mu = \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, d\mu + \int_{-\Delta\mu/2}^{\Delta\mu/2} \left| \sum_{n\neq 0} F(\mu - \Delta\mu\, n) \right|^2 d\mu \tag{2.67}$$

The term $\int_{-\Delta\mu/2}^{\Delta\mu/2} |\sum_{n\neq 0} F(\mu - \Delta\mu\, n)|^2 \, d\mu$ can be upperbounded as

$$\begin{aligned} \int_{-\Delta\mu/2}^{\Delta\mu/2} \left| \sum_{n\neq 0} F(\mu - \Delta\mu\, n) \right|^2 d\mu &= \sum_{m\neq 0} \sum_{n\neq 0} \int_{-\Delta\mu/2}^{\Delta\mu/2} F(\mu - \Delta\mu\, n)\, F^*(\mu - \Delta\mu\, m)\, d\mu \\ &\leq \sum_{m\neq 0} \sum_{n\neq 0} \left| \int_{-\Delta\mu/2}^{\Delta\mu/2} F(\mu - \Delta\mu\, n)\, F^*(\mu - \Delta\mu\, m)\, d\mu \right| \end{aligned} \tag{2.68}$$

From the Cauchy-Schwarz inequality for function spaces, we have

$$\begin{aligned} &\left| \int_{-\Delta\mu/2}^{\Delta\mu/2} F(\mu - \Delta\mu\, n)\, F^*(\mu - \Delta\mu\, m)\, d\mu \right|^2 \leq \\ &\int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu - \Delta\mu\, n)|^2 \, d\mu \int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu - \Delta\mu\, m)|^2 \, d\mu \end{aligned} \tag{2.69}$$

Then, combining this result with (2.68), we get

$$\int_{-\Delta\mu/2}^{\Delta\mu/2} \left| \sum_{n\neq 0} F(\mu - \Delta\mu\, n) \right|^2 d\mu \leq \left(\sum_{n\neq 0} \left(\int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu - \Delta\mu\, n)|^2 \, d\mu \right)^{\frac{1}{2}} \right)^2 \tag{2.70}$$

Thus, from (2.67), we obtain

$$\begin{aligned} \int |F(\mu) - \check{F}_{\Delta\mu}(\mu)|^2 \, d\mu \;\leq\; & \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, d\mu \\ & + \left(\sum_{n\neq 0} \left(\int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu - \Delta\mu\, n)|^2 \, d\mu \right)^{\frac{1}{2}} \right)^2 \end{aligned} \tag{2.71}$$

At this point, we can loose the upperbound here, and write

$$\begin{aligned}\left(\int |F(\mu)-\check{F}_{\Delta\mu}(\mu)|^2\,d\mu\right)^{\frac{1}{2}} &\leq \left(\int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \\ &\quad + \sum_{n\neq 0}\left(\int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu-\Delta\mu\, n)|^2\,d\mu\right)^{\frac{1}{2}} \quad (2.72)\\ &= \left(\int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \\ &\quad + \sum_{n\neq 0}\left(\int_{(n-\frac{1}{2})\Delta\mu}^{(n+\frac{1}{2})\Delta\mu} |F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \quad (2.73)\end{aligned}$$

Then, we use (2.65) to obtain

$$\begin{aligned}\left(\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right)^{\frac{1}{2}} &\leq \left(\int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \\ &\quad + \sum_{n\neq 0}\left(\int_{(n-\frac{1}{2})\Delta\mu}^{(n+\frac{1}{2})\Delta\mu} |F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \\ &\quad + \left(\frac{1}{\Delta\mu}\sum_{|n|>\lfloor\Delta u\Delta\mu/2\rfloor}\left|f\left(\frac{n}{\Delta\mu}\right)\right|^2\right)^{\frac{1}{2}} \quad (2.74)\end{aligned}$$

as the upperbound for the square root of the finite sample reconstruction error.

Similar to (2.44) and (2.48), provided that the function $|f(u)|$ is decreasing in the region $|u|>\frac{\lfloor\Delta u\Delta\mu/2\rfloor}{\Delta\mu}$ and $\Delta u\Delta\mu\gg 1$, we have

$$\frac{1}{\Delta\mu}\sum_{|n|>\lfloor\Delta u\Delta\mu/2\rfloor}\left|f\left(\frac{n}{\Delta\mu}\right)\right|^2 \approx \int_{|u|>\Delta u/2}|f(u)|^2\,du \quad (2.75)$$

After this approximation, we can rewrite (2.74) as

$$\begin{aligned}\left(\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right)^{\frac{1}{2}} &\leq \left(\int_{|u|>\Delta u/2}|f(u)|^2\,du\right)^{\frac{1}{2}} + \left(\int_{|\mu|>\Delta\mu/2}|F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \\ &\quad + \sum_{n\neq 0}\left(\int_{(n-\frac{1}{2})\Delta\mu}^{(n+\frac{1}{2})\Delta\mu}|F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \quad (2.76)\end{aligned}$$

Since (2.52) is equal to the sum of squares of the first and second terms of the summation in the RHS of (2.76), we conclude that the upperbound we have obtained here for the finite sample reconstruction error $\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du$

is larger than (2.52), as we stated in the beginning of this section. For a random process $f(u)$, since this argument works for all realizations, the upperbound we obtain here for the average finite sample reconstruction error $E[\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du]$ is larger than (2.53).

Now, we want to say a few words on the third term contributing to the RHS of (2.76). For any $a, b \in \mathbb{R}$, from Cauchy-Schwarz inequality, we have

$$\begin{aligned}(b-a)\int_a^b |F(\mu)|^2\, d\mu &= \int_a^b 1^2\, d\mu \int_a^b |F(\mu)|^2\, d\mu \\ &\geq \left(\int_a^b |F(\mu)|\, d\mu\right)^2 \end{aligned} \tag{2.77}$$

Thus, inserting $a = (n-1/2)\Delta\mu$ and $b = (n+1/2)\Delta\mu$ in (2.77), we conclude

$$\left(\int_{(n-\frac{1}{2})\Delta\mu}^{(n+\frac{1}{2})\Delta\mu} |F(\mu)|^2\, d\mu\right)^{\frac{1}{2}} \geq \frac{1}{\sqrt{\Delta\mu}} \int_{(n-\frac{1}{2})\Delta\mu}^{(n+\frac{1}{2})\Delta\mu} |F(\mu)|\, d\mu \tag{2.78}$$

$$\sum_{n\neq 0} \left(\int_{(n-\frac{1}{2})\Delta\mu}^{(n+\frac{1}{2})\Delta\mu} |F(\mu)|^2\, d\mu\right)^{\frac{1}{2}} \geq \frac{1}{\sqrt{\Delta\mu}} \int_{|\mu|>\Delta\mu/2} |F(\mu)|\, d\mu \tag{2.79}$$

Therefore, the third term of RHS of (2.76) is larger than the $\Delta\mu$ truncation error of the 1-norm of $F(\mu)$. Thus, in order to make our error upperbound (2.76) as small as desired, we first have to take $\Delta\mu$ truncation error of the 1-norm of $F(\mu)$ under control.

For a random process $f(u)$, taking the expectation of both sides in (2.76), and using the inequalities

$$E\left(\int_{|u|>\Delta u/2} |f(u)|^2\, du\right)^{\frac{1}{2}} \leq \left(E\left[\int_{|u|>\Delta u/2} |f(u)|^2\, du\right]\right)^{\frac{1}{2}} = \left(\int_{|u|>\Delta u/2} R(u,u)\, du\right)^{\frac{1}{2}} \tag{2.80}$$

$$E\left(\int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\, d\mu\right)^{\frac{1}{2}} \leq \left(E\left[\int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\, d\mu\right]\right)^{\frac{1}{2}} = \left(\int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\, d\mu\right)^{\frac{1}{2}} \tag{2.81}$$

stemming from the inequality $(E[X])^2 \leq E[X^2]$ where X is a real random variable, which can be rewritten as $E[X] \leq \sqrt{E[X^2]}$ when X does not take negative

values, we have

$$E\left(\int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du\right)^{\frac{1}{2}} \leq \left(\int_{|u|>\Delta u/2} R(u,u)\,du\right)^{\frac{1}{2}} + \left(\int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\,d\mu\right)^{\frac{1}{2}} + \sum_{n\neq 0} E\left(\int_{(n-\frac{1}{2})\Delta\mu}^{(n+\frac{1}{2})\Delta\mu} |F(\mu)|^2\,d\mu\right)^{\frac{1}{2}} \tag{2.82}$$

Similarly, from (2.79), we see that the average $\Delta\mu$ truncation error of the 1-norm of $F(\mu)$ should be made small enough first to make the error upperbound (2.82) sufficiently small.

2.5 Optimal Δu, $\Delta\mu$ and the Corresponding Best Achievable Finite Sample Reconstruction Error

Naturally, we want to use the smallest number of samples to achieve a specified finite sample reconstruction error and we desire to obtain the smallest possible finite sample reconstruction error for a given number of samples. This section is devoted to the application of the method of Lagrange multipliers to solve these two optimization problems. The parameters we need to optimize are Δu and $\Delta\mu$.

In Section 2.2, we have shown that the reconstruction error of FSR can be written as in (2.28) and (2.35) for the first and second options, respectively. In Section 2.3, we demonstrated that, under reasonable conditions, both (2.28) and (2.35) can be approximated as simply the sum of spatial and spectral truncation errors, namely (2.52). Thus, (2.52) is the ultimate form that the finite sample reconstruction error takes for both of the FSR options after some approximations. On the other hand, as given in (2.16), the number of samples, namely the number of DOF for the reconstruction signal, can be taken as $\Delta u\Delta\mu$. Based on these remarks, we can formulate these two optimization problems as

- Minimizing $n(\Delta u, \Delta\mu)$ subject to the constraint $e(\Delta u, \Delta\mu)$ is a specified constant.
- Minimizing $e(\Delta u, \Delta\mu)$ subject to the constraint $n(\Delta u, \Delta\mu)$ is a specified constant.

where

$$n(\Delta u, \Delta\mu) = \Delta u \Delta\mu \tag{2.83}$$

$$e(\Delta u, \Delta\mu) = \int_{|u|>\Delta u/2} |f(u)|^2 \, du + \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, d\mu \tag{2.84}$$

In order to be more precise, one can alternatively define $e(\Delta u, \Delta\mu)$ as the RHS of (2.47) and the RHS of (2.49) for the first and second FSR options, respectively. In this case, the details of the derivation would be quite similar. Thus, we continue our development by taking $e(\Delta u, \Delta\mu)$ as in (2.84).

For both of the two problems we have explained, the method of Lagrange multipliers indicates that $\exists\lambda \in \mathbb{R}$, the optimal $(\Delta u, \Delta\mu)$ point should satisfy

$$\frac{\partial\, e(\Delta u, \Delta\mu)}{\partial \Delta u} + \lambda \Delta\mu = 0 \tag{2.85}$$

$$\frac{\partial\, e(\Delta u, \Delta\mu)}{\partial \Delta\mu} + \lambda \Delta u = 0 \tag{2.86}$$

Note that $e(\Delta u, \Delta\mu)$ can be expressed as

$$e(\Delta u, \Delta\mu) = e_1(\Delta u) + e_2(\Delta\mu) \tag{2.87}$$

where

$$e_1(x) = E_0 - \int_{-x/2}^{x/2} |f(x')|^2 \, dx' \tag{2.88}$$

$$e_2(y) = E_0 - \int_{-y/2}^{y/2} |F(y')|^2 \, dy' \tag{2.89}$$

$$E_0 = \int |f(x')|^2 \, dx' = \int |F(y')|^2 \, dy' \tag{2.90}$$

Now, we can rewrite (2.85) and (2.86) as

$$e_1'(\Delta u) + \lambda \Delta \mu = 0 \tag{2.91}$$

$$e_2'(\Delta \mu) + \lambda \Delta u = 0 \tag{2.92}$$

resulting in the equality

$$e_1'(\Delta u)\Delta u = e_2'(\Delta \mu)\Delta \mu \tag{2.93}$$

The derivative of (2.88) can be calculated as

$$e_1'(x) = -\frac{1}{2}\left(\left|f\left(\frac{x}{2}\right)\right|^2 + \left|f\left(-\frac{x}{2}\right)\right|^2\right) \tag{2.94}$$

Similarly we have

$$e_2'(y) = -\frac{1}{2}\left(\left|F\left(\frac{y}{2}\right)\right|^2 + \left|F\left(-\frac{y}{2}\right)\right|^2\right) \tag{2.95}$$

Then, (2.93) can be rewritten as

$$\frac{\Delta \mu}{\Delta u} = \frac{\left|f\left(\frac{\Delta u}{2}\right)\right|^2 + \left|f\left(-\frac{\Delta u}{2}\right)\right|^2}{\left|F\left(\frac{\Delta \mu}{2}\right)\right|^2 + \left|F\left(-\frac{\Delta \mu}{2}\right)\right|^2} \tag{2.96}$$

In order to find the optimal $(\Delta u, \Delta \mu)$ pair, (2.96) and the constraint equation need to be solved together. In this way, we can find the smallest possible $e(\Delta u, \Delta \mu)$ for the constraint $n(\Delta u, \Delta \mu)$ is a given constant, and vice versa. Therefore, we can plot number of samples vs finite sample reconstruction error curve consisting of the best achievable points.

For a random process $f(u)$, we define $e(\Delta u, \Delta \mu)$ similarly as

$$e(\Delta u, \Delta \mu) = \int_{|u|>\Delta u/2} R(u,u)\, du + \int_{|\mu|>\Delta \mu/2} S(\mu,\mu)\, d\mu \tag{2.97}$$

based on the approximation (2.56). Then, for both of the optimization problems we defined, using the method of Lagrange multipliers, we obtain

$$\frac{\Delta \mu}{\Delta u} = \frac{R(\frac{\Delta u}{2}, \frac{\Delta u}{2}) + R(-\frac{\Delta u}{2}, -\frac{\Delta u}{2})}{S(\frac{\Delta \mu}{2}, \frac{\Delta \mu}{2}) + S(-\frac{\Delta \mu}{2}, -\frac{\Delta \mu}{2})} \tag{2.98}$$

similar to (2.96). From both (2.96) and (2.98), we see that on the curve $\Delta u \Delta \mu = N$, the optimum $(\Delta u, \Delta \mu)$ point is the one moving from which in upward or downward direction does not decrease $e(\Delta u, \Delta \mu)$. On the other hand, although it turns out that $\int_{|u|>\Delta u/2} R(u,u)\,du$ and $\int_{|\mu|>\Delta \mu/2} S(\mu,\mu)\,d\mu$ are equal to each other for optimal Δu and $\Delta \mu$ in the examples we consider in our work, we do not think that (2.98) necessarily imply $\int_{|u|>\Delta u/2} R(u,u)\,du = \int_{|\mu|>\Delta \mu/2} S(\mu,\mu)\,d\mu$.

In order to find the optimal $(\Delta u, \Delta \mu)$ pair, (2.98) and the constraint equation need to be solved together. Then, for a random process $f(u)$, we can plot number of samples vs the expectation of finite sample reconstruction error curve consisting of the best achievable points, i.e., we can plot number of samples vs the average finite sample reconstruction error Pareto optimal curve.

We will now provide a numerical example for the special case when the random process $f(u)$ of interest has an autocorrelation function of the form

$$R(u_1, u_2) = \psi_n(u_1)\psi_n(u_2) \tag{2.99}$$

where $\psi_n(u)$ is the n^{th} order Hermite-Gaussian function. Since Hermite-Gaussian functions are the eigenfunctions of the Fourier transform having eigenvalues of unit magnitude [113], autocorrelation and autocorrelation of the Fourier transform are exactly the same in this case. Therefore (2.98) simply reduces to $\Delta u = \Delta \mu \times 1\,s^2$. Then, under the constraint that the number of samples to be used is a constant N, (2.97) can be simplified as

$$2\int_{|u|>\sqrt{N}/2} \psi_n^2(u)\,du \tag{2.100}$$

From (2.100), $n(\Delta u, \Delta \mu)$ vs $e(\Delta u, \Delta \mu)$ Pareto optimal curves are obtained for several n values, as given in Figure 2.1.

As the order of the Hermite polynomial increases, both the spatial and the spectral width of the corresponding Hermite-Gaussian function increases as well. Therefore, in Figure 2.1, it is natural to observe that larger n results in usage of more samples to achieve the same error performance.

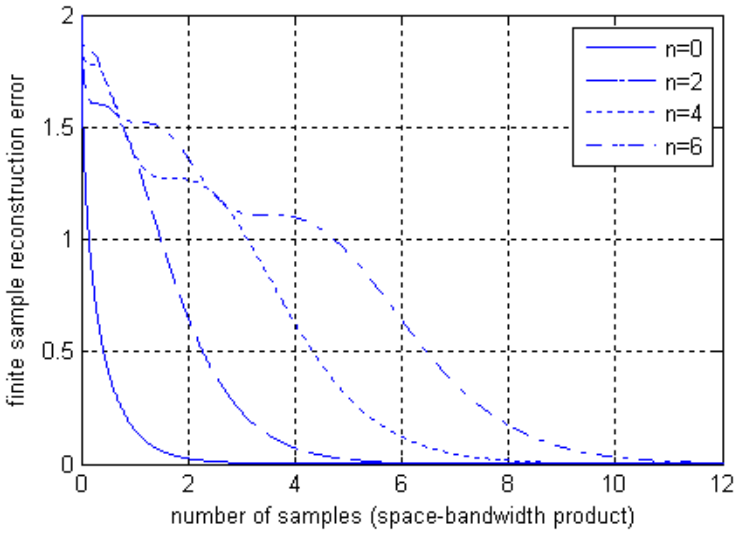

Figure 2.1: Number of samples vs finite sample reconstruction error Pareto optimal curves for the random processes having autocorrelation function $R(u_1, u_2) = \psi_n(u_1)\psi_n(u_2)$, where $\psi_n(u)$ refers to the n$^{\text{th}}$ order Hermite-Gaussian function.

As another example, we consider a random process $f(u)$ having a Gaussian Schell-model(GSM) type autocorrelation function

$$R(u_1, u_2) = A\, e^{-(u_1^2+u_2^2)/4\sigma_I^2}\, e^{-(u_1-u_2)^2/2\sigma_\mu^2} \tag{2.101}$$

In [114], it is proven that (2.101) can be decomposed as

$$R(u_1, u_2) = \sum_{n=-\infty}^{\infty} \lambda_n \sqrt{\frac{c}{\pi}}\, \psi_n\left(\sqrt{\frac{c}{\pi}}\, u_1\right) \psi_n\left(\sqrt{\frac{c}{\pi}}\, u_2\right) \tag{2.102}$$

where $\psi_n(u)$ is the the n$^{\text{th}}$ order Hermite-Gaussian function, λ_n is a positive number depending on σ_I, σ_μ and n, which is explicitly given in [114], and

$$c = \left(\left(\frac{1}{4\sigma_I^2}\right)^2 + \frac{1}{4\sigma_I^2\sigma_\mu^2}\right)^{1/2} \tag{2.103}$$

Then, using the fact that the functions $\psi_n(u)$ are the eigenfunctions of Fourier transform all having unit magnitude eigenvalues, and using the scaling property of Fourier transform, we get

$$S(\mu_1, \mu_2) = \sum_{n=-\infty}^{\infty} \lambda_n \sqrt{\frac{\pi}{c}}\, \psi_n\left(\sqrt{\frac{\pi}{c}}\mu_1\right) \psi_n\left(\sqrt{\frac{\pi}{c}}\mu_2\right) \tag{2.104}$$

$$= \frac{\pi}{c}\, R\left(\frac{\pi}{c}\mu_1, \frac{\pi}{c}\mu_2\right) \tag{2.105}$$

Then, from (2.105), we can write

$$\frac{R\left(\frac{\Delta u}{2},\frac{\Delta u}{2}\right)+R\left(-\frac{\Delta u}{2},-\frac{\Delta u}{2}\right)}{S\left(\frac{c\Delta u}{2\pi},\frac{c\Delta u}{2\pi}\right)+S\left(-\frac{c\Delta u}{2\pi},-\frac{c\Delta u}{2\pi}\right)}=\frac{\Delta u\, c/\pi}{\Delta u} \tag{2.106}$$

(2.106) implies that (2.98) simply reduces to $\Delta\mu = \Delta u\, c/\pi$ for a GSM type autocorrelation function. In this case, under the constraint $\Delta u \Delta\mu = N$, we obtain the optimal Δu and $\Delta\mu$ as $\sqrt{N\pi/c}$ and $\sqrt{Nc/\pi}$, respectively. Then, using (2.105), (2.97) can be rewritten as

$$\begin{aligned} e(\Delta u,\Delta\mu) &= \int_{|u|>\Delta u/2} R(u,u)\,du + \int_{|\mu|>\Delta u\, c/2\pi} \frac{\pi}{c}\, R\left(\frac{\pi}{c}\mu,\frac{\pi}{c}\mu\right)\,d\mu && (2.107)\\ &= 2\int_{|u|>\Delta u/2} R(u,u)\,du && (2.108)\\ &= 2A^2\int_{|u|>\sqrt{N\pi/4c}} e^{-u^2/2\sigma_I^2}\,du && (2.109)\\ &= 4A^2\sqrt{2\pi}\sigma_I\, Q\left(\sqrt{\frac{N\pi}{4\,c\,\sigma_I^2}}\right) && (2.110) \end{aligned}$$

Setting the insignificant amplitude factor A aside, the two parameters that determine a GSM type $R(u_1,u_2)$ are σ_I and σ_μ. If both of these two parameters are increased κ times, then c decreases κ^2 times. Therefore, $c\,\sigma_I^2$ does not change, and thus the ratio of the minimum achievable average finite sample reconstruction error to the average energy of $f(u)$, namely

$$\begin{aligned} \frac{e(\Delta u,\Delta\mu)}{\int R(u,u)\,du} &= \frac{e(\Delta u,\Delta\mu)}{\int A^2\, e^{-u^2/2\sigma_I^2}\,du} && (2.111)\\ &= \frac{e(\Delta u,\Delta\mu)}{A^2\sqrt{2\pi}\,\sigma_I} && (2.112)\\ &= 4\,Q\left(\sqrt{\frac{N\pi}{4\,c\,\sigma_I^2}}\right) && (2.113) \end{aligned}$$

does not change, either. Hence, we conclude that the normalized best achievable finite sample reconstruction error depends only on the ratio of σ_I to σ_μ.

Figure 2.2 illustrates $n(\Delta u,\Delta\mu)$ vs percentage $e(\Delta u,\Delta\mu)$ (100 times (2.113)) Pareto optimal curves for a couple of σ_I/σ_μ values. As the intensity width σ_I increases and the correlation width σ_μ decreases, the number of independent

samples having nonnegligible variance increases. Therefore, it is natural to observe that higher σ_I/σ_μ ratios result in the usage of more samples to achieve the same error.

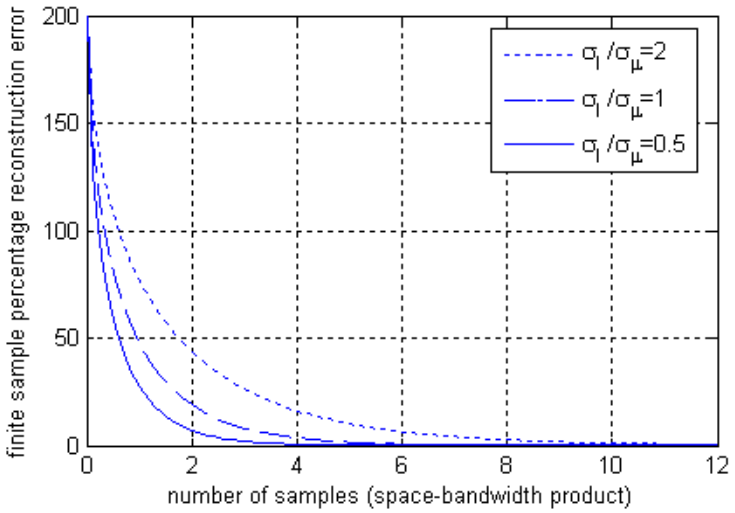

Figure 2.2: Number of samples vs finite sample reconstruction error Pareto optimal curves for random processes having GSM type autocorrelation function.

The variations of optimum $\Delta u = \sqrt{N\pi/c}$ and optimum $\Delta\mu = \sqrt{Nc/\pi}$ with respect to the number of samples N are shown in Figure 2.3 and 2.4, respectively. From these figures, we conclude that optimum Δu increases as σ_I or σ_μ increases. Whereas, optimum $\Delta\mu$ is inversely proportional to σ_I and σ_μ. Since the number of samples is equal to the product of Δu and $\Delta\mu$, comparing Figure 2.3 with Figure 2.4, we see that the (σ_I, σ_μ) pair having the largest optimal Δu has the smallest optimal $\Delta\mu$, and vice versa. In other words, the ordering of the curves in Figure 2.3 is reversed in Figure 2.4.

Moreover, comparing the curves of the (σ_I, σ_μ) pair $(1s, 0.5s)$ with $(2s, 1s)$, or comparing the curves corresponding to $(0.5s, 1s)$ with the curves corresponding to $(1s, 2s)$, we verify the fact that if both σ_I and σ_μ are increased κ times, then c decreases κ^2 times, resulting in a κ times increase in optimum Δu and a κ times decrease in optimum $\Delta\mu$.

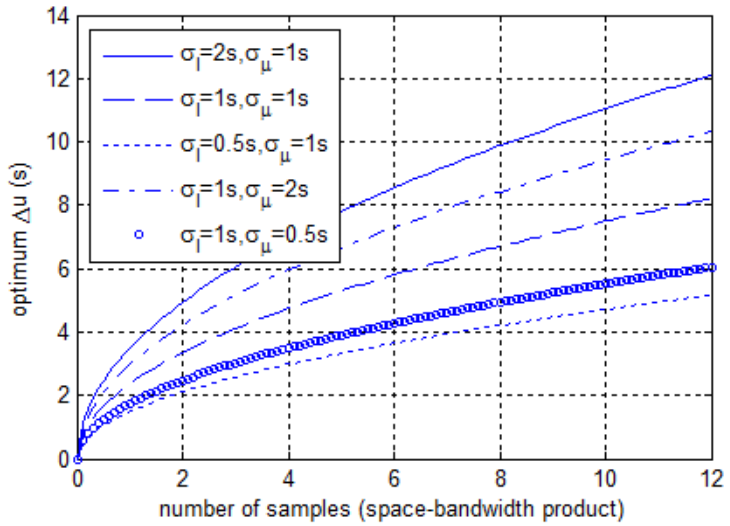

Figure 2.3: Number of samples vs optimum Δu curves for random processes having GSM type autocorrelation function.

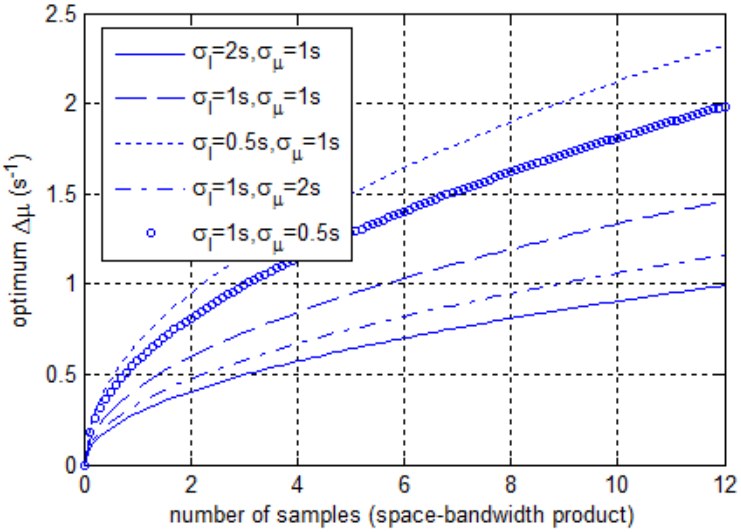

Figure 2.4: Number of samples vs optimum $\Delta\mu$ curves for random processes having GSM type autocorrelation function.

2.6 The Consequences of Prolate Spheroidal Functions on Our Work

In this section, we will discuss how the works on prolate spheroidal functions are related to our development. Prolate spheroidal functions are described in Slepian's well known paper [115] first, and some important properties of these functions are covered in Landau and Pollak papers [116, 117]. Here, we will first consider the results found in [116] with their consequences on the approximation of finite sample reconstruction error made in (2.52). Then, we will proceed to the results of [117] which are about the performance of the family of sincs (2.18) we used in reconstruction and prolate spheroidal functions in approximating bandlimited functions.

Except for Theorem 3, all the theorems given in this section are taken from [118], which includes the results of both [116] and [117]. However, all the remaining parts are our original work unless otherwise stated. For convenience, throughout this section, the signals considered have unit energy. Extending the results to the generic case when there is no restriction on the energy of signals is straightforward, as we did in the statement of Theorem 3.

Now, before starting our discussion, we give the following definitions which will be used throughout this section.

Definition 1. *The norm* $||f||$ *of a function* $f(u)$ *is defined as*

$$||f|| = \left(\int |f(u)|^2 \, du \right)^{\frac{1}{2}} \tag{2.114}$$

Definition 2. *The projection operator* A *confines the function to the interval* $[-\Delta u/2, \Delta u/2]$.

$$Af(u) = \begin{cases} f(u) & \text{if } |u| \leq \Delta u/2, \\ 0 & \text{else .} \end{cases} \tag{2.115}$$

Definition 3. *The projection operator B confines the Fourier transform of the function to the interval $[-\Delta\mu/2, \Delta\mu/2]$.*

$$Bf(u) = \int_{-\Delta\mu/2}^{\Delta\mu/2} F(\mu) e^{j2\pi\mu u}\, d\mu \tag{2.116}$$

Then, the operator BA can be expressed as

$$BAf(u) = \int_{-\Delta u/2}^{\Delta u/2} \Delta\mu\, \mathrm{sinc}[\Delta\mu(u-u')]\, f(u')\, du' \tag{2.117}$$

The eigenfunctions of BA operator are named as prolate spheroidal functions [115–117, 119]. Some of the properties of these functions and their eigenvalues are given in Theorem 4.

After giving the required definitions, we begin our discussion. Recall that, in Section 2.3, we have concluded that the reconstruction error for FSR of both the first and second options can be approximated as

$$\int |f(u) - \hat{f}_{\Delta u, \Delta\mu}(u)|^2 \approx \int_{|u|>\Delta u/2} |f(u)|^2\, du + \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\, d\mu \tag{2.118}$$

as written in (2.52). Since no signal $f(u)$ can be fully concentrated in both space and frequency domains, for fixed Δu and $\Delta\mu$, we cannot make both $\int_{|u|>\Delta u/2} |f(u)|^2\, du$ and $\int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\, d\mu$ as small as we desire by choosing $f(u)$ conveniently. In other words, we cannot make both

$$\alpha^2 = \int_{-\Delta u/2}^{\Delta u/2} |f(u)|^2\, du \tag{2.119}$$

and

$$\beta^2 = \int_{-\Delta\mu/2}^{\Delta\mu/2} |F(\mu)|^2\, d\mu \tag{2.120}$$

as close to $\int |f(u)|^2\, du = \int |F(\mu)|^2\, d\mu$ as we like, and consequently we cannot make (2.118) arbitrarily small. Therefore, once Δu and $\Delta\mu$ is fixed, irrespective of the function $f(u)$ to be represented by finite number of samples, we have to consent to a certain nonzero finite sample reconstruction error. Here, we aim to find this minimum finite sample reconstruction error in terms of Δu and $\Delta\mu$.

As an extension of Uncertainty Principle, there are some works in the literature about the spatial truncation error (2.119) and the spectral truncation error (2.120) which are concerned with the problem of finding the tightest bound on the (α, β) pairs achievable by a function $f(u)$. This problem is firstly considered and solved in [116]. Then, it is covered in [118, 119]. The solution of this problem will be useful in finding the minimum value that finite sample reconstruction error takes.

We begin stating our theorems with a simple and brief one.

Theorem 1. *A bandlimited signal cannot be identically* 0 *on any interval. Similarly, the Fourier transform of a spacelimited signal cannot be identically* 0 *on any interval.*

From this theorem, we easily conclude that the (α, β) pairs $(0, 1)$, $(1, 0)$ and $(1, 1)$ are not achievable. The next question is that whether there are any other (α, β) pairs which cannot be achieved by any unit energy function $f(u)$. The following theorem answers this question.

Theorem 2. *Inside the unit square* $[0, 1] \times [0, 1]$*, the set of achievable* (α, β) *pairs are the region defined by*

$$\cos^{-1}\alpha + \cos^{-1}\beta \geq \cos^{-1}\sqrt{\gamma} \tag{2.121}$$

excluding the points $(0, 1)$ *and* $(1, 0)$*, where* $0 \leq \gamma \leq 1$ *is the largest eigenvalue of the operator* BA*, and a concave and increasing function of the product* $\Delta u \Delta \mu$*. Moreover,* $\gamma\big|_{\Delta u \Delta \mu = 0} = 0$ *and* $\lim_{\Delta u \Delta \mu \to \infty} \gamma = 1$*. For* $\alpha > \sqrt{\gamma}$*, the functions achieving the bound of the region described by (2.121) are*

$$f(u) = \frac{\alpha}{\sqrt{\gamma}} A e_1(u) + \left(\frac{1-\alpha^2}{1-\gamma}\right)^{\frac{1}{2}} (e_1(u) - A e_1(u)) \tag{2.122}$$

where $e_1(u)$ *is the prolate spheroidal function having the largest eigenvalue* γ*.*

Actually, in none of the works [116, 118, 119], the function $\gamma(\Delta u \Delta \mu)$ is explicitly given. In these works, $\Delta u \Delta \mu$ vs γ plot similar to Figure 2.5 is provided instead.

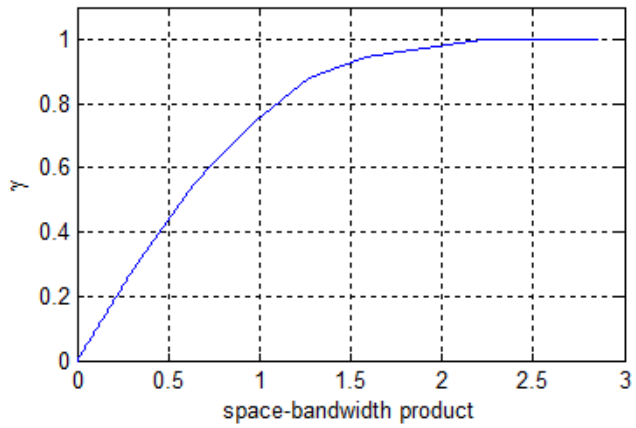

Figure 2.5: $\Delta u \Delta \mu$ vs γ curve obtained by reading off from Figure 2 of [116].

It is interesting that, for a given Δu and $\Delta \mu$, the set of the achievable points depends only on the product $\Delta u \Delta \mu$, as Theorem 2 implies.

Note that, if $\alpha^2 + \beta^2 \leq 1$, we have

$$\alpha \leq \sqrt{1-\beta^2} = \sin(\cos^{-1}\beta) = \cos(\pi/2 - \cos^{-1}\beta) \tag{2.123}$$

Since $\cos^{-1}$ is a decreasing function, taking $\cos^{-1}$ of each side, we get

$$\cos^{-1}\alpha + \cos^{-1}\beta \geq \pi/2 = \cos^{-1} 0 \geq \cos^{-1}\sqrt{\gamma} \tag{2.124}$$

Hence, from Theorem 2, we conclude that, inside the unit square $[0,1] \times [0,1]$, all the (α, β) pairs lying inside the unit circle centered at the origin is achievable, irrespective of $\Delta u > 0$ and $\Delta \mu > 0$.

Another implication of Theorem 2 is that if $\alpha \leq \sqrt{\gamma}$, then there is no restriction on β, namely $\forall \beta \in [0,1]$ is achievable. (Naturally, we also equivalently have if $\beta \leq \sqrt{\gamma}$, then $\forall \alpha \in [0,1]$ is achievable.) Note that since $\cos^{-1}$ is a decreasing function, if $\alpha \leq \sqrt{\gamma}$, then we have

$$\cos^{-1}\alpha \geq \cos^{-1}\sqrt{\gamma} \tag{2.125}$$

Then, since $\cos^{-1}\beta$ is always nonnegative, we immediately conclude

$$\cos^{-1}\alpha + \cos^{-1}\beta \geq \cos^{-1}\sqrt{\gamma},\ \forall\beta \in (0,1] \tag{2.126}$$

(2.121) also implies that for the class of unit energy functions bandlimited to $[-\Delta\mu/2, \Delta\mu/2]$, α^2 cannot exceed γ. (Equivalently, for the class of unit energy functions space limited to the interval $[-\Delta u/2, \Delta u/2]$, β^2 cannot exceed γ.) Actually, before proving Theorem 2, in [118], γ is defined as supremum of (2.119) taken over the class of bandlimited functions.

On the other hand, if $\alpha \leq \sqrt{\gamma}$ does not hold, we first rewrite (2.121) as

$$\cos^{-1}\beta \geq \cos^{-1}\sqrt{\gamma} - \cos^{-1}\alpha \tag{2.127}$$

Since we consider the case $\alpha > \sqrt{\gamma}$ here, we have

$$\cos^{-1}\sqrt{\gamma} - \cos^{-1}\alpha > 0 \tag{2.128}$$

Then, using the fact that the cosine function is decreasing on the interval $[0, \pi/2]$, (2.127) can be expressed as

$$\beta \leq \cos(\cos^{-1}\sqrt{\gamma} - \cos^{-1}\alpha) \tag{2.129}$$

$$\beta \leq \alpha\sqrt{\gamma} + \sin(\cos^{-1}\alpha)\sin(\cos^{-1}\sqrt{\gamma}) \tag{2.130}$$

$$\beta \leq \alpha\sqrt{\gamma} + \sqrt{1-\alpha^2}\sqrt{1-\gamma} \tag{2.131}$$

Therefore, if $\alpha > \sqrt{\gamma}$, (2.131) and (2.121) can be used interchangeably to express the region of achievable (α, β) pairs.

In (2.131), taking the square of both sides, we get

$$\beta^2 \leq \alpha^2(2\gamma - 1) + 2\alpha\sqrt{1-\alpha^2}\sqrt{\gamma-\gamma^2} + 1 - \gamma \tag{2.132}$$

Then, from (2.132), we obtain the inequality

$$2 - \alpha^2 - \beta^2 \geq (1-\gamma) + 2\gamma(1-\alpha^2) - 2\alpha\sqrt{1-\alpha^2}\sqrt{\gamma-\gamma^2} \tag{2.133}$$

the left hand side of which is nothing but

$$\begin{aligned} 2-\alpha^2-\beta^2 &= (1-\alpha^2)+(1-\beta^2) \\ &= \int_{|u|>\Delta u/2} |f(u)|^2\,du + \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2\,d\mu \\ &\approx \int |f(u)-\hat{f}_{\Delta u,\Delta\mu}(u)|^2\,du \end{aligned} \tag{2.134}$$

That is why we are interested in lowerbounding $2-\alpha^2-\beta^2$. As explained at the beginning of this section, for fixed Δu and $\Delta\mu$, there is an inevitable finite sample reconstruction error and our aim is to find this error which we cannot avoid independent of the function $f(u)$ to be reconstructed.

(2.133) implies that for the unit energy functions satisfying (2.119) for a certain α greater than $\sqrt{\gamma}$, the minimum value that $2-\alpha^2-\beta^2$ can take is

$$(1-\gamma)+2\gamma(1-\alpha^2)-2\alpha\sqrt{1-\alpha^2}\sqrt{\gamma-\gamma^2} \tag{2.135}$$

However, note that (2.133) is valid when $\alpha > \sqrt{\gamma}$. On the other hand, if $\alpha \leq \sqrt{\gamma}$, then

$$2-\alpha^2-\beta^2 \geq 2-\gamma-\beta^2 \geq 2-\gamma-1 = 1-\gamma \tag{2.136}$$

where the inequality is achieved by equality for $(\alpha,\beta) = (\sqrt{\gamma},1)$. But, when $\alpha > \sqrt{\gamma}$, we will also achieve $2-\alpha^2-\beta^2 = 1-\gamma$ by the point $(\alpha,\beta) = (1,\sqrt{\gamma})$. Thus, denoting the indispensible finite sample reconstruction error we aim to find as e_{min}, we have

$$\begin{aligned} e_{min} &= \min\left\{ \min_{\alpha>\sqrt{\gamma}} \{(1-\gamma)+2\gamma(1-\alpha^2)-2\alpha\sqrt{1-\alpha^2}\sqrt{\gamma-\gamma^2}\}, 1-\gamma \right\} \\ &= \min_{\alpha>\sqrt{\gamma}} \{(1-\gamma)+2\gamma(1-\alpha^2)-2\alpha\sqrt{1-\alpha^2}\sqrt{\gamma-\gamma^2}\} \end{aligned} \tag{2.137}$$

Since $\alpha\sqrt{1-\alpha^2}$ is increasing when $\alpha \leq 1/\sqrt{2}$, (2.135) is decreasing for the case $\alpha \leq 1/\sqrt{2}$. Indeed, we have

$$\begin{aligned} &\frac{d}{d\alpha}\left[(1-\gamma)+2\gamma(1-\alpha^2)-2\alpha\sqrt{1-\alpha^2}\sqrt{\gamma-\gamma^2}\right] \\ &= -2\left(2\alpha\gamma+\sqrt{\gamma-\gamma^2}\,\frac{1-2\alpha^2}{\sqrt{1-\alpha^2}}\right) \\ &\leq 0 \end{aligned} \tag{2.138}$$

for $\alpha \in [0, 1/\sqrt{2}]$. Now, in order to compute (2.137), we want to see whether there exists a number α_0 greater than both $1/\sqrt{2}$ and $\sqrt{\gamma}$ until which (2.135) continues to decrease, or equivalently

$$2\alpha\gamma + \sqrt{\gamma - \gamma^2}\,\frac{1-2\alpha^2}{\sqrt{1-\alpha^2}} \geq 0 \tag{2.139}$$

continues to be true. (2.139) can be rewritten as

$$2\alpha\gamma \geq \sqrt{\gamma - \gamma^2}\,\frac{2\alpha^2-1}{\sqrt{1-\alpha^2}} \tag{2.140}$$

Since we consider the case $\alpha^2 > 1/2$, both sides of (2.140) are positive. Thus, taking the square of both sides, (2.140) can also be expressed as

$$4\alpha^2\gamma^2 \geq (\gamma - \gamma^2)\,\frac{4\alpha^4 - 4\alpha^2 + 1}{1-\alpha^2} \tag{2.141}$$

After arranging the terms accordingly, from (2.141), we get

$$4\gamma\alpha^4 - 4\gamma\alpha^2 + \gamma - \gamma^2 \leq 0 \tag{2.142}$$

$$4\gamma\left(\alpha^2 - \frac{1-\sqrt{\gamma}}{2}\right)\left(\alpha^2 - \frac{1+\sqrt{\gamma}}{2}\right) \leq 0 \tag{2.143}$$

From (2.143), we conclude that (2.135) is decreasing when $1/2 \leq \alpha^2 \leq (1+\sqrt{\gamma})/2$ as well as the case $\alpha^2 \leq 1/2$. Moreover, (2.143) implies that (2.135) no longer becomes a decreasing function of α after α^2 exceeds the threshold $(1+\sqrt{\gamma})/2$. Therefore, noting that

$$\alpha_0 = \sqrt{\frac{1+\sqrt{\gamma}}{2}} \geq \sqrt{\frac{\gamma+\gamma}{2}} = \sqrt{\gamma} \tag{2.144}$$

we find e_{min} as

$$\begin{aligned}
e_{min} &= \left[(1-\gamma) + 2\gamma(1-\alpha^2) - 2\alpha\sqrt{1-\alpha^2}\sqrt{\gamma-\gamma^2}\right]\Big|_{\alpha^2=(1+\sqrt{\gamma})/2} && (2.145)\\
&= 1 - \gamma + 2\gamma\,\frac{1-\sqrt{\gamma}}{2} - 2\sqrt{\frac{1+\sqrt{\gamma}}{2}}\sqrt{\frac{1-\sqrt{\gamma}}{2}}\sqrt{\gamma-\gamma^2} && (2.146)\\
&= 1 - \gamma\sqrt{\gamma} - \sqrt{\gamma}(1-\gamma) && (2.147)\\
&= 1 - \sqrt{\gamma} && (2.148)
\end{aligned}$$

which is achieved only when $\alpha^2 = (1+\sqrt{\gamma})/2$ and

$$\beta^2 = 2 - \frac{1+\sqrt{\gamma}}{2} - e_{min} = \frac{1+\sqrt{\gamma}}{2} = \alpha^2 \tag{2.149}$$

Moreover, from Theorem 2, we see that the minimum finite sample reconstruction error e_{min} is achieved by the function

$$f(u) = \left[\frac{\alpha}{\sqrt{\gamma}} Ae_1(u) + \left(\frac{1-\alpha^2}{1-\gamma}\right)^{\frac{1}{2}} (e_1(u) - Ae_1(u))\right]\Bigg|_{\alpha^2=(1+\sqrt{\gamma})/2} \tag{2.150}$$

$$= \left(\frac{1+\sqrt{\gamma}}{2\gamma}\right)^{\frac{1}{2}} \left[Ae_1(u) + \frac{\sqrt{\gamma}}{1+\sqrt{\gamma}}(e_1(u) - Ae_1(u))\right] \tag{2.151}$$

We summarize these results in the following theorem.

Theorem 3. *For any signal $f(u)$, the finite sample reconstruction error expressed in (2.118) is at least $1-\sqrt{\gamma}$ fraction of its energy. The minimum finite sample reconstruction error*

$$(1-\sqrt{\gamma}) \int |f(u)|^2 \, du \tag{2.152}$$

is achieved by the function

$$f(u) = C\left[Ae_1(u) + \frac{\sqrt{\gamma}}{1+\sqrt{\gamma}}(e_1(u) - Ae_1(u))\right] \tag{2.153}$$

where C is any nonzero number. Moreover, the minimum finite sample reconstruction error is achieved only when the spatial truncation error $\int_{|u|>\Delta u/2} |f(u)|^2 \, du$ and the spectral truncation error $\int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, du$ are the same and equal to $\frac{1-\sqrt{\gamma}}{2} \int |f(u)|^2 \, du$.

Theorem 3 implies that for the extreme cases $\Delta u = 0$ and $\Delta\mu = 0$, namely for the case $\Delta u \Delta\mu = 0$, the finite sample reconstruction error will be as large as the whole energy of the signal to be reconstructed, which is a trivial result. Moreover, according to Theorem 3, for the other extreme case $\Delta u \Delta\mu = \infty$, there exists signals for which the finite sample reconstruction error is zero. To verify this, we can simply consider the signals space limited to $[-\Delta u/2, \Delta u/2]$ and the signals bandlimited to $[-\Delta\mu/2, \Delta\mu/2]$ for the cases when $\Delta\mu = \infty$ and $\Delta u = \infty$, respectively. Therefore, this is an expected result as well.

By plotting $\Delta u \Delta\mu$ vs $1-\sqrt{\gamma}$ graph, we can demonstrate how the minimum finite sample reconstruction error we have to accept changes depending on the

number of samples. On the other hand, the problem of minimizing finite sample reconstruction error for a specific signal $f(u)$ under the constraint $\Delta u \Delta \mu$ is constant is solved in Section 2.5, where we adjusted Δu and $\Delta \mu$ accordingly so that the error is minimized. Whereas, as we see in this section, changing Δu and $\Delta \mu$ do not have any effect on the minimum achievable finite sample reconstruction error as long as $\Delta u \Delta \mu$ is kept constant.

Figure 2.6 illustrates the comparison of the $\Delta u \Delta \mu$ vs $1 - \sqrt{\gamma}$ curve with the Pareto optimal $\Delta u \Delta \mu$ vs finite sample reconstruction error curve given in Figure 2.1 for $n = 0$.

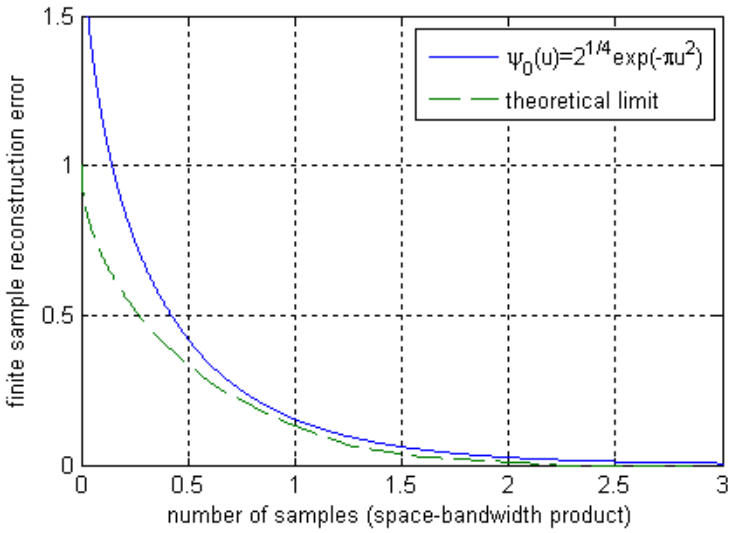

Figure 2.6: Comparison of the theoretical $1 - \sqrt{\gamma}$ limit and space-bandwidth product vs finite sample reconstruction error Pareto optimal curve for $f(u) = \psi_0(u) = 2^{1/4} e^{-\pi u^2}$.

From the point of view of Uncertainty Principle [113,118], $e^{-\pi u^2}$ is the function which is most concentrated in both space and frequency domain. However, if the measure of being concentrated in both domains is taken as the sum of spatial and spectral truncation errors, from Theorem 3, we know that the function most concentrated in both domains is the one given in (2.153). Nevertheless, we conclude from Figure 2.6 that the difference between theoretical limit achieved

by (2.153) and $e^{-\pi u^2}$ becomes negligible when $\Delta u \Delta \mu > 1$, consistent with the result of Uncertainty Principle.

Now, we give some of the properties of prolate spheroidal functions and their eigenvalues in the following theorem.

Theorem 4. *The operator BA has countably many eigenvalues*

$$1 > \gamma = \gamma_1 \geq \gamma_2 \geq \gamma_3 \cdots \to 0$$

The eigenvalue sequence γ_n satisfies

- $\sum_{n=1}^{\infty} \gamma_n^2 \leq \Delta u \Delta \mu$
- $\sum_{n=1}^{\infty} \gamma_n = \Delta u \Delta \mu$

Moreover the associated eigenfunctions e_n, namely prolate spheroidal functions, have the following properties:

- $\{e_n | n \geq 1\}$ *is an orthonormal basis of the class of functions bandlimited to* $[-\Delta\mu/2, \Delta\mu/2]$.
- $\left\{\gamma_n^{-1/2} A e_n | n \geq 1\right\}$ *is an orthonormal basis of the class of functions space limited to* $[-\Delta u/2, \Delta u/2]$.
- *The functions e_n, suitably truncated and scaled, equal their Fourier transforms [119].*

At this point, we are ready to present our theorem on approximating a unit energy function f bandlimited to $[-\Delta\mu/2, \Delta\mu/2]$ with an orthonormal set $\{f_k | k = 1, 2, \ldots, n\}$ and discuss its consequences on our work.

Theorem 5. *Define $\Delta[f_1, ...f_n]$ as the least upper bound of*

$$\left|\left| f - \sum_{k=1}^{n} \langle f, f_k \rangle f_k \right|\right| \tag{2.154}$$

over the unit energy functions bandlimited to $[-\Delta\mu/2, \Delta\mu/2]$ *satisfying* $||Af|| = \alpha$ *for a constant* α*.*

a) $\Delta[f_1, ..., f_n]$ *is least for* $f_1 = e_1, ..., f_n = e_n$*, and this is the case* $\forall n \geq 1$*.*

b) $\Delta^2[e_1, ..., e_n] \leq 12(1 - \alpha^2), \forall n > \Delta u \Delta\mu$*.*

c) $\Delta^2[e_1, ..., e_{[\Delta u \Delta\mu + 1] + n}] \geq (0.916)^{-1}(1 - \alpha^2 - 2\sqrt{2}e^{-\pi \Delta u \Delta\mu/4})$*, if* $1 - \alpha^2 < 0.916$*, n is fixed and* $\Delta u \Delta\mu$ *is sufficiently large.*

d) $\Delta^2[e_1, ..., e_n] \leq (1 + \delta)(1 - \alpha^2)$*, for* $n = \Delta u \Delta\mu + C(\delta)\log(\Delta u \Delta\mu + 1)$*, where* δ *is any positive number and* $C(\delta)$ *is a constant which depends only on* δ*.*

Although Theorem 5 is taken from [118], except for d), this theorem is firstly stated and proven in [117]. Theorem 5-d) is due to Shannon. To be more precise, Theorem 5-a),b),c) and d) is nothing but Theorem 1, Theorem 3, Theorem 8, and Theorem 4 in [117], respectively.

As given in (2.15), recall that the reconstruction signal for the first option is

$$\hat{f}_{\Delta u, \Delta\mu}(u) = \sum_{n=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{2.155}$$

Actually, this equation can be rewritten as

$$\hat{f}_{\Delta u, \Delta\mu}(u) = \sum_{k=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \langle \breve{f}_{\Delta\mu}, f_k \rangle\, f_k(u) \tag{2.156}$$

where

$$f_k(u) = \sqrt{\Delta\mu}\operatorname{sinc}(\Delta\mu\, u - k) \tag{2.157}$$

From the Theorem 5-a), we see that, in terms of the worst case value of

$$\int |\breve{f}_{\Delta\mu}(u) - \hat{f}_{\Delta u, \Delta\mu}(u)|^2\, du = \left|\left| \breve{f}_{\Delta\mu} - \sum_{k=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \langle \breve{f}_{\Delta\mu}, f_k \rangle\, f_k \right|\right|^2 \tag{2.158}$$

choosing the family of sincs as the orthonormal set $\{f_k\}$, as we actually did in our work, is suboptimal. Actually, according to Theorem 10 and 11 of [117], the

contrary of Theorem 5-b) and d) are valid for the family of sincs. But this does not mean that for every bandlimited function $\breve{f}_{\Delta\mu}$ satisfying $||A\breve{f}_{\Delta\mu}|| = \alpha$ for a constant α, the reconstruction performance of the orthonormal set

$$\left\{ e_k \left| k = 1, 2, ..., 2\left\lfloor \frac{\Delta u \Delta\mu}{2} \right\rfloor + 1 \right. \right\} \tag{2.159}$$

is better than that of

$$\left\{ \sqrt{\Delta\mu}\,\mathrm{sinc}(\Delta\mu\, u - k) \right| - \left\lfloor \frac{\Delta u \Delta\mu}{2} \right\rfloor \leq k \leq \left\lfloor \frac{\Delta u \Delta\mu}{2} \right\rfloor \Bigg\} \tag{2.160}$$

On the other hand, there is another result given in [118] which makes us optimistic about the reconstruction performance of our set given in (2.160). Defining $e(\delta)$ as the square of the error in approximating $\breve{f}_{\Delta\mu}(u+\delta)$ by the function set (2.160), namely expressing $e(\delta)$ as

$$e(\delta) = \left|\left| \breve{f}_{\Delta\mu}(u+\delta) - \sum_{n=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu} + \delta \right) \mathrm{sinc}\left(\Delta\mu\, u - n\right) \right|\right|^2 \tag{2.161}$$

we have

$$\int_0^{1/\Delta\mu} e(\delta)\, d\delta \leq \frac{1}{\Delta\mu} \int_{|u|>\Delta u/2} |\breve{f}_{\Delta\mu}(u)|^2\, du = \frac{1-\alpha^2}{\Delta\mu} \tag{2.162}$$

as calculated in [118]. Thus, there exists a lag $0 \leq \delta' \leq 1/\Delta\mu$ such that

$$e(\delta') \leq 1 - \alpha^2 \tag{2.163}$$

On the other hand, provided that $1-\alpha^2 < 0.916$ and $\Delta u \Delta\mu$ is sufficiently large, from Theorem 5-c), we get

$$\Delta^2[e_1, ..., e_{2\lfloor \Delta u \Delta\mu/2 \rfloor + 1}] \geq (0.916)^{-1}(1 - \alpha^2 - 2\sqrt{2}e^{-\pi \Delta u \Delta\mu/4}) \tag{2.164}$$

For large $\Delta u \Delta\mu$, it is also the case that RHS of (2.164) is larger than $1-\alpha^2$. Therefore, comparing this fact with (2.163), we conclude that, for large $\Delta u \Delta\mu$, there exists some functions for which the error in approximating them with the set (2.159) of prolate spheroidal functions is larger than the error in approximating a delayed version of them with the set (2.160) of family of sincs. However,

in [118], it is stated that the relation between the optimal lag δ and $\breve{f}_{\Delta\mu}$ is very complicated and nonlinear.

Lastly, we remind that the reconstruction signal for the second option is the inverse Fourier transform of

$$\hat{F}_{\Delta u,\Delta\mu}(\mu) = \sum_{n=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \tilde{F}_{\Delta u}\left(\frac{n}{\Delta u}\right) \operatorname{sinc}(\Delta u\, \mu - n) \tag{2.165}$$

as given in (2.31). Comparing this equation with (2.15), we conclude that all the arguments and results we gave after Theorem 5 is valid for the second option as well, if we simply replace $\hat{f}_{\Delta u,\Delta\mu}(u)$ by $\hat{F}_{\Delta u,\Delta\mu}(\mu)$, $\breve{f}_{\Delta\mu}(u)$ by $\tilde{F}_{\Delta u}(\mu)$, Δu by $\Delta\mu$ and $\Delta\mu$ by Δu.

Chapter 3

ENCODING OF THE SAMPLES

In Chapter 2, for a random or deterministic finite energy signal $f(u)$, we proposed

$$\hat{f}_{\Delta u,\Delta\mu}(u) = \sum_{n=-\lfloor \Delta u\Delta\mu/2 \rfloor}^{\lfloor \Delta u\Delta\mu/2 \rfloor} \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{3.1}$$

as the reconstruction signal, and

$$\mathbf{f} = \left(\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \middle| -\left\lfloor \frac{\Delta u\Delta\mu}{2} \right\rfloor \leq n \leq \left\lfloor \frac{\Delta u\Delta\mu}{2} \right\rfloor \right) \tag{3.2}$$

as the FSR. A dual approach, namely the second option for FSR, is also discussed.

In this chapter, we will consider the quantization of the samples forming $\mathbf{f}$ to encode $f(u)$ by finitely many bits at the expense of the associated finite bit reconstruction error. Here, we analyze and compare the performances of scalar uniform quantization, vector quantization of uniformly quantized samples, spatial non-uniform quantization depending on the sample variances, and the optimal quantization induced by rate distortion theory. Moreover, for the vector quantization covered, the parameters (Δu, $\Delta\mu$ and number of levels K) that number of bits and overall reconstruction error depend on are optimized, and

consequently number of bits vs overall reconstruction error Pareto optimal curve is obtained.

3.1 Uniform Quantization of Samples

In this section, we uniformly quantize the samples $\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})$ as $\breve{f}^q_{\Delta\mu}(\frac{n}{\Delta\mu})$ and obtain the reconstruction signal

$$f^q_{\Delta u,\Delta\mu}(u) = \sum_{n=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{3.3}$$

Finite number of bits are sufficient to determine $f^q_{\Delta u,\Delta\mu}(u)$. Therefore, we name $f^q_{\Delta u,\Delta\mu}(u)$ as finite bit reconstruction signal.

As written in (2.90), let the energy of the signal $f(u)$ be denoted by E_0. Then, since the energy of $\breve{f}_{\Delta\mu}(u)$ cannot exceed that of $f(u)$, from (2.21), we conclude $|\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})| \leq \sqrt{E_0\Delta\mu}$, $\forall n \in \mathcal{Z}$. Therefore, both real and imaginary parts of the samples $\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})$ are confined to the interval $[-\sqrt{E_0\Delta\mu}, \sqrt{E_0\Delta\mu}]$. Thus, the uniform quantization is to be done in this interval. If both real and imaginary parts of the samples are to be quantized by K number of levels, then the amplitude step between consecutive levels is

$$\frac{\sqrt{E_0\Delta\mu} - (-\sqrt{E_0\Delta\mu})}{K} = \frac{2\sqrt{E_0\Delta\mu}}{K} \tag{3.4}$$

and the maximum quantization error that can be made for a real or imaginary part of a sample $\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})$ is one half of (3.4), namely $\frac{\sqrt{E_0\Delta\mu}}{K}$. Hence, we have

$$\begin{aligned}
\left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) - \breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2 &= \left(\operatorname{Re}\left\{\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) - \breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right\}\right)^2 \\
&\quad + \left(\operatorname{Im}\left\{\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) - \breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right\}\right)^2 \\
&\leq \left(\frac{\sqrt{E_0\Delta\mu}}{K}\right)^2 + \left(\frac{\sqrt{E_0\Delta\mu}}{K}\right)^2 \qquad (3.5) \\
&= \frac{2E_0\Delta\mu}{K^2} \qquad (3.6)
\end{aligned}$$

Then, defining the quantization error as

$$e_q(\Delta u, \Delta\mu) = \int |\hat{f}_{\Delta u,\Delta\mu}(u) - f^q_{\Delta u,\Delta\mu}(u)|^2 \, du \tag{3.7}$$

and using the orthogonality of sincs, we get

$$\begin{aligned} e_q(\Delta u, \Delta\mu) &= \frac{1}{\Delta\mu} \sum_{n=-\lfloor \Delta u\Delta\mu/2 \rfloor}^{\lfloor \Delta u\Delta\mu/2 \rfloor} \left| \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) - \breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \right|^2 && (3.8) \\ &\leq \frac{1}{\Delta\mu}\left(2\left\lfloor \frac{\Delta u\Delta\mu}{2} \right\rfloor + 1\right)\frac{2E_0\Delta\mu}{K^2} && (3.9) \\ &\approx \frac{2E_0\Delta u\Delta\mu}{K^2} && (3.10) \end{aligned}$$

where $\frac{2E_0\Delta\mu}{K^2}$ in (3.9) comes from (3.6).

From (3.10), we conclude that, for any given $\epsilon_q > 0$, if the number of levels K is selected as $\sqrt{\frac{2E_0\Delta u\Delta\mu}{\epsilon_q}}$, then the quantization error $e_q(\Delta u, \Delta\mu)$ becomes less than or equal to ϵ_q. Since each sample consists of real and imaginary parts, there are two real variables to be quantized for each sample, resulting in a total of

$$2\left(2\left\lfloor \frac{\Delta u\Delta\mu}{2} \right\rfloor + 1\right) \approx 2\Delta u\Delta\mu \tag{3.11}$$

scalar quantizations, each requiring

$$\log_2 K = \frac{1}{2}\log_2\left(\frac{2E_0\Delta u\Delta\mu}{\epsilon_q}\right) \tag{3.12}$$

bits. Therefore, in this way, which is named as scalar uniform quantization,

$$2\Delta u\Delta\mu \times \frac{1}{2}\log_2\left(\frac{2E_0\Delta u\Delta\mu}{\epsilon_q}\right) = \Delta u\Delta\mu\log_2\left(\frac{2E_0\Delta u\Delta\mu}{\epsilon_q}\right) \tag{3.13}$$

bits are sufficient to ensure $e_q(\Delta u, \Delta\mu) \leq \epsilon_q$.

Now, consider a class of signals $\mathcal{F}$ such that the energy of none of the signals belonging to it exceeds E_0. Since all the arguments we presented so far are valid for any signal having energy less than or equal to E_0, we conclude that as many as (3.13) bits are sufficient to make $e_q(\Delta u, \Delta\mu) \leq \epsilon_q$, for all $f \in \mathcal{F}$. Thus, worst case quantization error for $\mathcal{F}$ cannot exceed ϵ_q.

However, such a quantization is quite inefficient. Because, actually there are not

$$K^{2\Delta u\Delta\mu} = \left(\frac{2E_0\Delta u\Delta\mu}{\epsilon_q}\right)^{\Delta u\Delta\mu} \tag{3.14}$$

different possible quantization points

$$\hat{\mathbf{f}} = \left(\breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\middle| -\left\lfloor\frac{\Delta u\Delta\mu}{2}\right\rfloor \le n \le \left\lfloor\frac{\Delta u\Delta\mu}{2}\right\rfloor\right) \tag{3.15}$$

due to the limitation coming from

$$\frac{1}{\Delta\mu}\sum_{n=-\lfloor\Delta u\Delta\mu/2\rfloor}^{\lfloor\Delta u\Delta\mu/2\rfloor}\left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2 \le \frac{1}{\Delta\mu}\sum_{n=-\infty}^{\infty}\left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2 = \int|\breve{f}_{\Delta\mu}(u)|^2\,du \le E_0 \tag{3.16}$$

(3.16) implies that the quantization points $\hat{\mathbf{f}}$ outide the hypersphere of radius $\sqrt{E_0\Delta\mu}$ are useless. Actually, the number of quantization points staying inside the hypersphere of radius $\sqrt{E_0\Delta\mu}$ is much more smaller than (3.14), as we will show.

Thinking the real and imaginary parts of the samples $\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})$ seperately, we can regard the quantization points $\hat{\mathbf{f}}$ as vectors in $\mathbb{R}^{2\Delta u\Delta\mu}$, by taking the approximation in (3.11) into account. Inside the hypersphere we mentioned, each vector

$$\mathbf{f} = \left(\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\middle| -\left\lfloor\frac{\Delta u\Delta\mu}{2}\right\rfloor \le n \le \left\lfloor\frac{\Delta u\Delta\mu}{2}\right\rfloor\right) \tag{3.17}$$

will be represented as $\hat{\mathbf{f}}$ after uniform quantization if none of the $2\Delta u\Delta\mu$ components of $\hat{\mathbf{f}}$ is far away from the corresponding component of $\mathbf{f}$ more than one half of (3.4). Therefore, for all $\hat{\mathbf{f}}$, the locus of the vectors $\mathbf{f}$ represented by $\hat{\mathbf{f}}$ is a hypercube of edge length

$$2\times\frac{2\sqrt{E_0\Delta\mu}/K}{2} = \frac{2\sqrt{E_0\Delta\mu}}{K} \tag{3.18}$$

and dimension $2\Delta u\Delta\mu$, having a volume of

$$\left(\frac{2\sqrt{E_0\Delta\mu}}{K}\right)^{2\Delta u\Delta\mu} = \left(\frac{2\sqrt{E_0\Delta\mu}}{\sqrt{2E_0\Delta u\Delta\mu/\epsilon_q}}\right)^{2\Delta u\Delta\mu} = \left(\frac{2\epsilon_q}{\Delta u}\right)^{\Delta u\Delta\mu} \tag{3.19}$$

On the other hand, our hypersphere of radius $\sqrt{E_0 \Delta\mu}$ and dimension $2\Delta u \Delta\mu$ has a volume of

$$\frac{\pi^{\Delta u \Delta\mu}}{(\Delta u \Delta\mu)!} (E_0 \Delta\mu)^{\Delta u \Delta\mu} \tag{3.20}$$

Then, dividing (3.20) by (3.19), we find the number of quantization points $\hat{\mathbf{f}}$ inside the hypersphere as

$$\frac{1}{(\Delta u \Delta\mu)!} \left(\frac{\pi}{2}\right)^{\Delta u \Delta\mu} \left(\frac{E_0 \Delta u \Delta\mu}{\epsilon_q}\right)^{\Delta u \Delta\mu} \tag{3.21}$$

which is only

$$\frac{1}{(\Delta u \Delta\mu)!} \left(\frac{\pi}{4}\right)^{\Delta u \Delta\mu} \tag{3.22}$$

fraction of (3.14). Instead of scalar quantization, after observing the vector (3.17), one can detect which one of the different quantization points as many as (3.21) the vector is mapped to. Thus, by using only

$$\begin{aligned} &\log_2 \left(\frac{1}{(\Delta u \Delta\mu)!} \left(\frac{\pi}{2}\right)^{\Delta u \Delta\mu} \left(\frac{E_0 \Delta u \Delta\mu}{\epsilon_q}\right)^{\Delta u \Delta\mu} \right) \\ &= \Delta u \Delta\mu \log_2 \left(\frac{\pi E_0 \Delta u \Delta\mu}{2\epsilon_q} \right) - \log_2 (\Delta u \Delta\mu)! \end{aligned} \tag{3.23}$$

bits, $e_q(\Delta u, \Delta\mu) \leq \epsilon_q$ can be achieved. Such kind of quantization is an example of vector quantization, because all the samples $\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})$ are encoded together as a vector instead of applying uniform quantization to them independently. Since the positions of the quantization points $\hat{\mathbf{f}}$ are inherited from the usual uniform scalar quantization, we can name this quantization method as uniform vector quantization. Comparing (3.23) with (3.13), we see that uniform vector quantization makes it possible to have the same quantization performance by using

$$\log_2 (\Delta u \Delta\mu)! + \Delta u \Delta\mu \log_2 \left(\frac{4}{\pi}\right) \tag{3.24}$$

bits less. Needless to repeat, as well as scalar uniform quantization case, the results we presented here for vector quantization is valid not only for a single function $f(u)$ having a certain energy E_0, but also for any signal class $\mathcal{F}$ the signals in which have energy E_0 at most.

Moreover, note that using Stirling's approximation

$$\ln N! \approx N \ln N - N + \frac{1}{2} \ln(2\pi N) \tag{3.25}$$

we can approximate (3.23) as

$$\Delta u \Delta \mu \log_2 \left(\frac{\pi e E_0}{2\epsilon_q} \right) - \frac{1}{2} \log_2(2\pi \Delta u \Delta \mu) \tag{3.26}$$

On the other hand, we note that the overall (finite bit) reconstruction error can be upperbounded as

$$\begin{aligned} \left(\int |f(u) - f^q_{\Delta u, \Delta \mu}(u)|^2 \, du \right)^{\frac{1}{2}} \leq & \left(\int |f(u) - \hat{f}_{\Delta u, \Delta \mu}(u)|^2 \, du \right)^{\frac{1}{2}} \\ & + \left(\int |\hat{f}_{\Delta u, \Delta \mu}(u) - f^q_{\Delta u, \Delta \mu}(u)|^2 \, du \right)^{\frac{1}{2}} \end{aligned} \tag{3.27}$$

when $f(u)$ is a deterministic signal. For a class of signals $\mathcal{F}$, or equivalently a random process $f(u)$, taking the expectation of both sides in (3.27) and changing the order of expectation and square root as done in (2.80) and (2.81), we get

$$\begin{aligned} E \left(\int |f(u) - f^q_{\Delta u, \Delta \mu}(u)|^2 \, du \right)^{\frac{1}{2}} \leq & \left(E \left[\int |f(u) - \hat{f}_{\Delta u, \Delta \mu}(u)|^2 \, du \right] \right)^{\frac{1}{2}} \\ & + \left(E \left[\int |\hat{f}_{\Delta u, \Delta \mu}(u) - f^q_{\Delta u, \Delta \mu}(u)|^2 \, du \right] \right)^{\frac{1}{2}} \end{aligned} \tag{3.28}$$

as the overall error upperbound for stochastic case. If the two terms on the RHS of (3.28) can be made arbitrarily small by appropriately choosing Δu, $\Delta \mu$ and K, then the overall reconstruction error can also be made arbitrarily small, as will be the case for many processes of physical interest. Nevertheless, we suspect the existence of certain random processes for which this may not be true.

Lastly, we remark that the reconstruction signal for the second option is the inverse Fourier transform of

$$\hat{F}_{\Delta u, \Delta \mu}(\mu) = \sum_{n=-\lfloor \Delta u \Delta \mu / 2 \rfloor}^{\lfloor \Delta u \Delta \mu / 2 \rfloor} \tilde{F}_{\Delta u} \left(\frac{n}{\Delta u} \right) \operatorname{sinc}(\Delta u \, \mu - n) \tag{3.29}$$

as given in (2.31). After uniformly quantizing the samples $\tilde{F}_{\Delta u} \left(\frac{n}{\Delta u} \right)$ as $\tilde{F}^q_{\Delta u} \left(\frac{n}{\Delta u} \right)$, we obtain the finite bit reconstruction signal, having the Fourier transform

$$F^q_{\Delta u, \Delta \mu}(\mu) = \sum_{n=-\lfloor \Delta u \Delta \mu / 2 \rfloor}^{\lfloor \Delta u \Delta \mu / 2 \rfloor} \tilde{F}^q_{\Delta u} \left(\frac{n}{\Delta u} \right) \operatorname{sinc}(\Delta u \, \mu - n) \tag{3.30}$$

Now, comparing (2.31) and (3.30) with (2.15) and (3.3) respectively, we conclude that, after replacing $f(u)$ by $F(\mu)$, $\hat{f}_{\Delta u,\Delta\mu}(u)$ by $\hat{F}_{\Delta u,\Delta\mu}(\mu)$, $\breve{f}_{\Delta\mu}$ by $\tilde{F}_{\Delta u}$, $\breve{f}^{q}_{\Delta\mu}$ by $\tilde{F}^{q}_{\Delta u}$, $f^{q}_{\Delta u,\Delta\mu}(u)$ by $F^{q}_{\Delta u,\Delta\mu}(\mu)$, $\Delta\mu$ by Δu, and Δu by $\Delta\mu$, all the work done in this section is valid for the second FSR option as well.

3.2 Number of Bits vs Error Pareto Optimal Curve: The Method of Lagrange Multipliers Revisited

In Section 3.1, after covering scalar uniform quantization, we considered a vector quantization technique based on the fact that the quantization points are enclosed by a hypersphere. For vector quantization, we have found the sufficient number of bits in (3.23) in terms of Δu, $\Delta\mu$ and K to have a quantization error less than a specified threshold ϵ_q. In this section, we will optimize Δu, $\Delta\mu$ and K by using the method of Lagrange multipliers to solve the problem of finding the smallest number of bits to achieve a specified reconstruction error and finding the smallest possible reconstruction error for a given number of bits. Here, we first consider a single function $f(u)$ having energy E_0, then proceed to the case when $f(u)$ is a random process the realizations of which do not have an energy larger than E_0 (Or equivalently, we will proceed to the case when $\mathcal{F}$ is a signal class such that energy of the signals in it does not exceed E_0).

Before proceeding, we first express the number of bits used for the vector quantization we proposed in terms of K, rather than ϵ_q. Without inserting $\sqrt{\frac{2E_0\Delta u\Delta\mu}{\epsilon_q}}$ to K, if we divide (3.20) by (3.19), we get

$$\frac{(E_0\pi\Delta\mu)^{\Delta u\Delta\mu}/(\Delta u\Delta\mu)!}{\left(2\sqrt{E_0\Delta\mu}/K\right)^{2\Delta u\Delta\mu}} = \frac{1}{(\Delta u\Delta\mu)!}\left(\frac{\pi K^2}{4}\right)^{\Delta u\Delta\mu} \tag{3.31}$$

Thus, in terms of $\Delta u, \Delta\mu$ and K, the number of bits can be written as

$$\Delta u \Delta\mu \log_2 \left(\frac{\pi K^2}{4} \right) - \log_2(\Delta u \Delta\mu)! \tag{3.32}$$

Since $\Delta u \Delta\mu \gg 1$ in practice, we can drop the term $\frac{1}{2}\ln(2\pi N)$ in Stirling's approximation we stated in (3.25), and write

$$\ln(\Delta u \Delta\mu)! \approx \Delta u \Delta\mu \ln(\Delta u \Delta\mu) - \Delta u \Delta\mu \tag{3.33}$$

Thus, we approximete (3.32) as

$$b(\Delta u, \Delta\mu, K) = \Delta u \Delta\mu \log_2 \left(\frac{\pi e K^2}{4 \Delta u \Delta\mu} \right) \tag{3.34}$$

Now, although we are unable to express the overall reconstruction error $\int |f(u) - f^q_{\Delta u, \Delta\mu}(u)|^2 \, du$ in terms of Δu, $\Delta\mu$ and K directly, we can find an upperbound for the square root of it which can be written as the function of Δu, $\Delta\mu$ and K. In order to find such an upperbound, we first combine (3.27) with (2.52) and get

$$\begin{aligned} \left(\int |f(u) - f^q_{\Delta u, \Delta\mu}(u)|^2 \, du \right)^{\frac{1}{2}} \;\leq\; & \left(\int_{|u|>\Delta u/2} |f(u)|^2 \, du + \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, d\mu \right)^{\frac{1}{2}} \\ & + \left(\int |\hat{f}_{\Delta u, \Delta\mu}(u) - f^q_{\Delta u, \Delta\mu}(u)|^2 \, du \right)^{\frac{1}{2}} \end{aligned} \tag{3.35}$$

Then, we use (3.10) to simplify (3.35) as

$$\begin{aligned} \left(\int |f(u) - f^q_{\Delta u, \Delta\mu}(u)|^2 \, du \right)^{\frac{1}{2}} \;\leq\; & \left(\int_{|u|>\Delta u/2} |f(u)|^2 \, du + \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, d\mu \right)^{\frac{1}{2}} \\ & + \sqrt{2E_0} \, \frac{\sqrt{\Delta u \Delta\mu}}{K} \end{aligned} \tag{3.36}$$

RHS of (3.36) is the upperbound we are looking for. Thus, we define $e(\Delta u, \Delta\mu, K)$ as

$$e(\Delta u, \Delta\mu, K) = \left(\int_{|u|>\Delta u/2} |f(u)|^2 \, du + \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, d\mu \right)^{\frac{1}{2}} + \sqrt{2E_0} \, \frac{\sqrt{\Delta u \Delta\mu}}{K} \tag{3.37}$$

Here, note that we do not deviate too much from the original reconstruction error by defining $e(\Delta u, \Delta\mu, K)$ based on the upperbound coming from (3.27).

Because, the overall error due to quantization and sampling is typically greater than the error coming from sampling and the error coming from quantization. Thus, typically we have

$$\left(\int |f(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \geq$$
$$\frac{1}{2}\left[\left(\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} + \left(\int |\hat{f}_{\Delta u,\Delta\mu}(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}}\right] \quad (3.38)$$

Therefore, (3.27) is typically tight enough and setting the usage of (3.10) aside, (3.37) is accurate within a factor of 2 as an approximation of the square root of $\int |f(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du$. Moreover, assuming that the amplitude step between consecutive quantization levels is so small that the $2\Delta u\Delta\mu$ samples are evenly distributed to the quantization interval $\left[-\frac{\sqrt{E_0\Delta\mu}}{K}, \frac{\sqrt{E_0\Delta\mu}}{K}\right]$, we can rewrite (3.9) as

$$e_q(\Delta u, \Delta\mu) \approx \frac{1}{\Delta\mu} \times 2\Delta u\Delta\mu \times \frac{1}{3}\left(\frac{\sqrt{E_0\Delta\mu}}{K}\right)^2 = \frac{2E_0\Delta u\Delta\mu}{3K^2} \quad (3.39)$$

where $\frac{1}{3}$ in (3.39) comes from the fact that variance of a random variable uniformly distributed in the interval $[-L, L]$ is $\frac{1}{3}L^2$. Thus, comparing (3.10) with (3.39), we see that (3.10) is accurate within a factor of 3.

Hence, we conclude that the inequalities resulting in (3.37) are reasonably tight and (3.37) is accurate enough to be used instead of the square root of $\int |f(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du$.

Now, similar to Section 2.5, we use Lagrange multipliers method to solve the problems of minimizing $b(\Delta u, \Delta\mu, K)$ subject to the constraint $e(\Delta u, \Delta\mu, K)$ is a given constant and minimizing $e(\Delta u, \Delta\mu, K)$ subject to the constraint $b(\Delta u, \Delta\mu, K)$ is a given constant. For both of these optimization problems, the method of Lagrange multipliers indicates that $\exists\lambda \in \mathbb{R}$, the optimal $(\Delta u, \Delta\mu, K)$

triple should satisfy

$$\frac{\partial\, e(\Delta u, \Delta\mu, K)}{\partial \Delta u} + \lambda \frac{\partial\, b(\Delta u, \Delta\mu, K)}{\partial \Delta u} = 0 \tag{3.40}$$

$$\frac{\partial\, e(\Delta u, \Delta\mu, K)}{\partial \Delta\mu} + \lambda \frac{\partial\, b(\Delta u, \Delta\mu, K)}{\partial \Delta\mu} = 0 \tag{3.41}$$

$$\frac{\partial\, e(\Delta u, \Delta\mu, K)}{\partial K} + \lambda \frac{\partial\, b(\Delta u, \Delta\mu, K)}{\partial K} = 0 \tag{3.42}$$

From (3.42), we obtain

$$-\sqrt{2E_0}\ \frac{\sqrt{\Delta u \Delta\mu}}{K^2} + 2\log_2 e\, \lambda \frac{\Delta u \Delta\mu}{K} = 0 \tag{3.43}$$

$$\lambda = \frac{\ln 2}{K}\sqrt{\frac{E_0}{2\Delta u \Delta\mu}} \tag{3.44}$$

Now, after some algebraic manipulations, (3.40) and (3.41) can be rewritten as

$$-\frac{\left|f\left(\frac{\Delta u}{2}\right)\right|^2 + \left|f\left(-\frac{\Delta u}{2}\right)\right|^2}{4\sqrt{e(\Delta u, \Delta\mu)}} + \frac{1}{K}\sqrt{\frac{E_0 \Delta\mu}{2\Delta u}}\ln\left(\frac{\pi e K^2}{4\Delta u \Delta\mu}\right) = 0 \tag{3.45}$$

$$-\frac{\left|F\left(\frac{\Delta \mu}{2}\right)\right|^2 + \left|F\left(-\frac{\Delta \mu}{2}\right)\right|^2}{4\sqrt{e(\Delta u, \Delta\mu)}} + \frac{1}{K}\sqrt{\frac{E_0 \Delta u}{2\Delta\mu}}\ln\left(\frac{\pi e K^2}{4\Delta u \Delta\mu}\right) = 0 \tag{3.46}$$

where $e(\Delta u, \Delta\mu)$ is equal to (2.52), as defined in (2.84). Multiplying both sides of (3.45) by Δu and both sides of (3.46) by $\Delta\mu$, we obtain

$$\frac{\Delta\mu}{\Delta u} = \frac{\left|f\left(\frac{\Delta u}{2}\right)\right|^2 + \left|f\left(-\frac{\Delta u}{2}\right)\right|^2}{\left|F\left(\frac{\Delta \mu}{2}\right)\right|^2 + \left|F\left(-\frac{\Delta \mu}{2}\right)\right|^2} \tag{3.47}$$

This equation is nothing but (2.96) in Section 2.5! It is nice to observe that the equation that Δu and $\Delta\mu$ should satisfy for the optimum performance does not change when quantization is taken into account.

In order to find the optimal $(\Delta u, \Delta\mu, K)$ point, one needs to solve (3.45), (3.46) and the constraint equation together. In this way, we can find the smallest possible $e(\Delta u, \Delta\mu, K)$ for the constraint $b(\Delta u, \Delta\mu, K)$ is a given constant, and vice versa. Therefore, we can plot number of bits vs reconstruction error curve consisting of the best achievable points. In other words, we can obtain number of bits vs reconstruction error Pareto optimal curve.

Now, as we declared in the beginning of this section, we examine the case when $f(u)$ is a random process the energy of the realizations of which is upperbounded by a certain number E_0. In order to define $e(\Delta u, \Delta\mu, K)$, we first use (2.56) in (3.28) to obtain

$$E\left(\int |f(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \leq \left(\int_{|u|>\Delta u/2} R(u,u)\, du + \int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\, d\mu\right)^{\frac{1}{2}} + \left(E\left[\int |\hat{f}_{\Delta u,\Delta\mu}(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du\right]\right)^{\frac{1}{2}} \tag{3.48}$$

Now, since the inequality (3.10) is valid for all realizations of $f(u)$, it should be valid for the expectation as well. Therefore, we have

$$E\left[\int |\hat{f}_{\Delta u,\Delta\mu}(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du\right] \leq \frac{2E_0\Delta u\Delta\mu}{K^2} \tag{3.49}$$

Using (3.49) in (3.48), we get

$$E\left(\int |f(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du\right)^{\frac{1}{2}} \leq \left(\int_{|u|>\Delta u/2} R(u,u)\, du + \int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\, d\mu\right)^{\frac{1}{2}} + \sqrt{2E_0}\frac{\sqrt{\Delta u\Delta\mu}}{K} \tag{3.50}$$

Then, we define $e(\Delta u, \Delta\mu, K)$ for the random process case as the RHS of (3.50), namely

$$e(\Delta u, \Delta\mu, K) = \left(\int_{|u|>\Delta u/2} R(u,u)\, du + \int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\, d\mu\right)^{\frac{1}{2}} + \sqrt{2E_0}\frac{\sqrt{\Delta u\Delta\mu}}{K} \tag{3.51}$$

Because of the same reasons explained before, the inequality (3.50) is considerably tight, as well. Thus, defining $e(\Delta u, \Delta\mu, K)$ as in (3.51) is plausible.

Now, comparing (3.51) with (3.37), we see that the only difference is usage of $R(u,u)$ and $S(\mu,\mu)$ instead of $|f(u)|^2$ and $|F(\mu)|^2$, respectively. Thus, after using the method of Lagrange multipliers, the equations we obtain are

$$-\frac{R(\frac{\Delta u}{2},\frac{\Delta u}{2}) + R(-\frac{\Delta u}{2},-\frac{\Delta u}{2})}{4\sqrt{e(\Delta u,\Delta\mu)}} + \frac{1}{K}\sqrt{\frac{E_0\Delta\mu}{2\Delta u}}\ln\left(\frac{\pi eK^2}{4\Delta u\Delta\mu}\right) = 0 \tag{3.52}$$

$$-\frac{S(\frac{\Delta\mu}{2},\frac{\Delta\mu}{2}) + S(-\frac{\Delta\mu}{2},-\frac{\Delta\mu}{2})}{4\sqrt{e(\Delta u,\Delta\mu)}} + \frac{1}{K}\sqrt{\frac{E_0\Delta u}{2\Delta\mu}}\ln\left(\frac{\pi eK^2}{4\Delta u\Delta\mu}\right) = 0 \tag{3.53}$$

similar to (3.45) and (3.46), where $e(\Delta u, \Delta\mu)$ is as defined in (2.97). From (3.52) and (3.53), we similarly derive

$$\frac{\Delta\mu}{\Delta u} = \frac{R(\frac{\Delta u}{2}, \frac{\Delta u}{2}) + R(-\frac{\Delta u}{2}, -\frac{\Delta u}{2})}{S(\frac{\Delta\mu}{2}, \frac{\Delta\mu}{2}) + S(-\frac{\Delta\mu}{2}, -\frac{\Delta\mu}{2})} \tag{3.54}$$

which is exactly the same as (2.98). Therefore, the equation that the optimal Δu and $\Delta\mu$ satisfy does not change when quantization is taken into account.

In order to find the least possible $e(\Delta u, \Delta\mu, K)$ for the constraint $b(\Delta u, \Delta\mu, K)$ is a given constant and vice versa, we solve (3.52), (3.53) and the constraint together to find the three unknowns Δu, $\Delta\mu$ and K. Equivalently, one can also solve (3.54), the constraint and either (3.52) or (3.53) together. Then, we can obtain number of bits vs the average reconstruction error curve consisting of the best achievable points, which is reminiscent of the rate-distortion curve in information theory.

As an example, similar to Section 2.5, we consider the special case when the random process $f(u)$ has an autocorrelation function $R(u_1, u_2)$ of the form

$$R(u_1, u_2) = \psi_n(u_1)\psi_n(u_2) \tag{3.55}$$

where $\psi_n(u)$ is the n$^{\text{th}}$ order Hermite-Gaussian function. As explained in Section 2.5, (3.54) is equivalent to $\Delta u = \Delta\mu \times 1\, s^2$ in this case. Then, under the constraint that the number of bits to be used is a constant R, (3.52) can be simplified as

$$\frac{\psi_n^2\left(\frac{\Delta u}{2}\right)}{\left(\int_{|u|>\Delta u/2} \psi_n^2(u)\, du\right)^{\frac{1}{2}}} = \ln 2 \sqrt{\pi e}\, \frac{2^{-R/2(\Delta u)^2} R}{(\Delta u)^3} \tag{3.56}$$

Solving (3.56) numerically, we find the optimal Δu and $\Delta\mu$ for a fixed R. Then, from (3.51), we obtain the least possible, or equivalently the best achievable, $e(\Delta u, \Delta\mu, K)$ for the constraint $b(\Delta u, \Delta\mu, K) = R$. The rate distortion curves of our development, namely $b(\Delta u, \Delta\mu, K)$ vs square of $e(\Delta u, \Delta\mu, K)$ Pareto optimal curves, are given in Figure 3.1, for several n values.

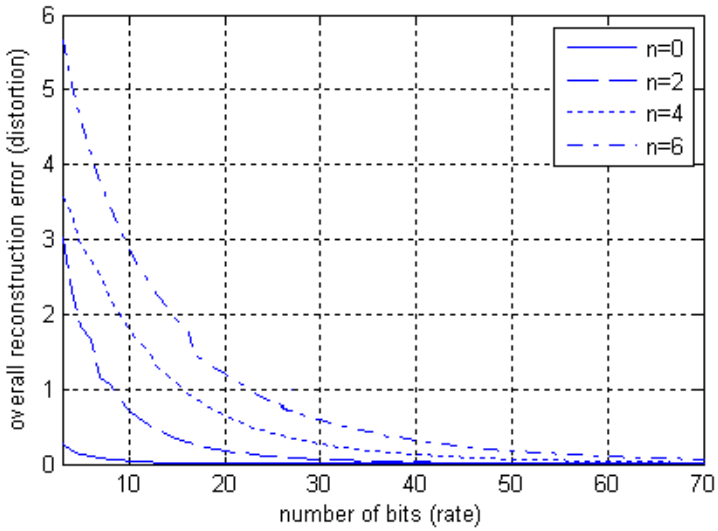

Figure 3.1: Rate distortion curves for the random processes having autocorrelation function $R(u_1, u_2) = \psi_n(u_1)\psi_n(u_2)$, where $\psi_n(u)$ refers to the n$^{\text{th}}$ order Hermite-Gaussian function.

As the order of the Hermite polynomial increases, both the spatial and the spectral width of the corresponding Hermite-Gaussian function increases as well. Therefore, similar to Section 2.5, in Figure 3.1, it is natural to observe that larger n results in usage of more bits to achieve the same error performance.

As another example, similar to Section 2.5, we consider a random process having an autocorrelation function of the form

$$R(u_1, u_2) = A\, e^{-(u_1^2+u_2^2)/4\sigma_I^2}\, e^{-(u_1-u_2)^2/2\sigma_\mu^2} \tag{3.57}$$

In Section 2.5, it was shown that the solution of (3.54) is $\Delta\mu = \Delta u\, c/\pi$ for a GSM type autocorrelation function, i.e., for an autocorrelation in the form (3.57). Then, under the constraint that the number of bits is equal to R, from (3.52), after some algebraic manipulations, we get

$$\frac{e^{-(\Delta u)^2/8\sigma_I^2}}{\sqrt{Q\left(\frac{\Delta u}{2\sigma_I}\right)}} = 2\pi\sqrt{e}\,\sigma_I \ln 2\, \frac{\pi R}{c(\Delta u)^3} 2^{-\pi R/2c(\Delta u)^2} \tag{3.58}$$

Solving (3.58) numerically, we compute optimal Δu and $\Delta\mu$ corresponding to R. Then, inserting the optimal Δu, $\Delta\mu$, and K in (3.51), we obtain the best achievable $e(\Delta u, \Delta\mu, K)$ under the condition $b(\Delta u, \Delta\mu, K) = R$.

As mentioned in Section 2.5, a κ times increase in σ_I and σ_μ results in a κ^2 times decrease in c. After rewriting (3.58) as

$$\frac{e^{-(\kappa\Delta u)^2/8(\kappa\sigma_I)^2}}{\sqrt{Q\left(\frac{\kappa\Delta u}{2\kappa\sigma_I}\right)}} = 2\pi\sqrt{e}\,\kappa\sigma_I \ln 2\,\frac{\pi R}{(c/\kappa^2)(\kappa\Delta u)^3} 2^{-\pi R/2(c/\kappa^2)(\kappa\Delta u)^2} \tag{3.59}$$

we see that if both σ_I and σ_μ are increased κ times, the optimal Δu increases κ times as well. Thus, the optimal $\Delta\mu = \Delta u\, c/\pi$ decreases κ times and $\Delta u\Delta\mu$ does not change. Since $b(\Delta u, \Delta\mu, K)$ depends only on $\Delta u\Delta\mu$ except for K, we conclude that optimal number of levels K does not change, either.

As shown in the equations (2.107)-(2.108) of Section 2.5, (3.54) implies $\int_{|u|>\Delta u/2} R(u,u)\,du = \int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\,d\mu$ for optimal Δu and $\Delta\mu$. Moreover, since $E_0 \propto \sigma_I$ as found in (2.112), after the κ times increase in σ_I and σ_μ, the new minimum achievable error $e_{\text{new}}(\Delta u, \Delta\mu, K)$ becomes

$$\begin{aligned} e_{\text{new}}(\Delta u, \Delta\mu, K) &= \left(2\int_{|u|>\kappa\Delta u/2} R\left(\frac{u}{\kappa},\frac{u}{\kappa}\right)\,du\right)^{\frac{1}{2}} + \sqrt{2\kappa E_0}\frac{\sqrt{\Delta u\Delta\mu}}{K} && (3.60)\\ &= \left(2\kappa\int_{|u|>\Delta u/2} R(u,u)\,du\right)^{\frac{1}{2}} + \sqrt{\kappa}\sqrt{2E_0}\frac{\sqrt{\Delta u\Delta\mu}}{K} && (3.61)\\ &= \sqrt{\kappa}e_{\text{old}}(\Delta u, \Delta\mu, K) && (3.62) \end{aligned}$$

From (3.62), we conclude that if both σ_I and σ_μ are increased κ times, $e^2(\Delta u, \Delta\mu, K)$ increases κ times as well. However, the ratio of the least achievable $e^2(\Delta u, \Delta\mu, K)$ to the average energy of $f(u)$, that is, $e^2(\Delta u, \Delta\mu, K)/\int R(u,u)\,du$, remains constant since $\int R(u,u)\,du \propto \sigma_I$. Therefore, the normalized best achievable overall reconstruction error $e^2(\Delta u, \Delta\mu, K)$ depends only on the ratio of σ_I to σ_μ. Recall that a similar fact was proven in Section 2.5 for finite sample reconstruction error.

$b(\Delta u, \Delta\mu, K)$ vs percentage $e^2(\Delta u, \Delta\mu, K)$ Pareto optimal curves, namely our rate distortion curves, are given for a couple of σ_I/σ_μ values in Figure 3.2.

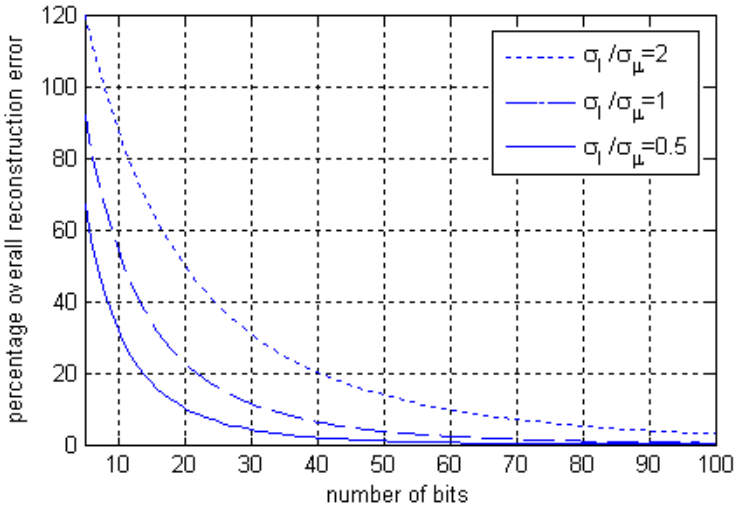

Figure 3.2: Rate distortion curves for random processes having GSM type autocorrelation function.

From Figure 3.2, we see that if the ratio σ_I/σ_μ increases, the required number of bits to achieve the same percentage error increases as well. This is an expected result, since the increase in the intensity width σ_I and the decrease in the correlation width σ_μ increases the information content of the random process $f(u)$, as explained in Section 2.5.

Number of bits vs optimum Δu and optimum $\Delta\mu$ plots are provided in Figure 3.3 and 3.4, respectively. In accordance with the corresponding figures of Section 2.5, these plots indicate that the increase in σ_I and σ_μ results in an increase in optimum Δu and a decrease in optimum $\Delta\mu$.

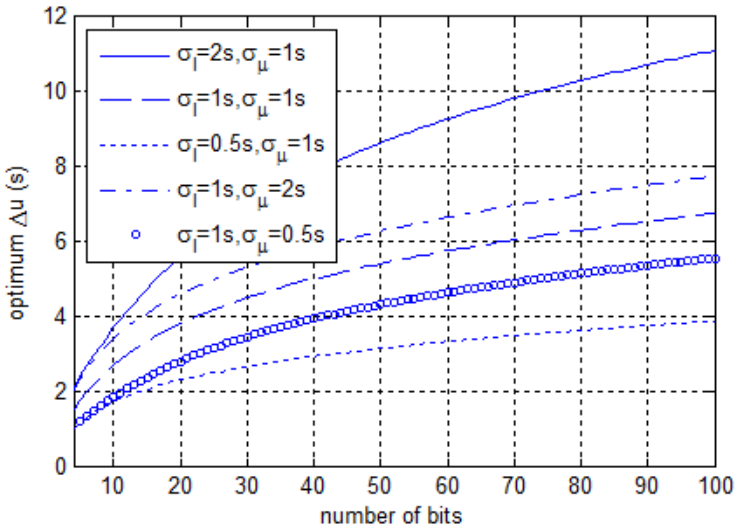

Figure 3.3: Number of bits vs optimum Δu curves for random processes having GSM type autocorrelation function.

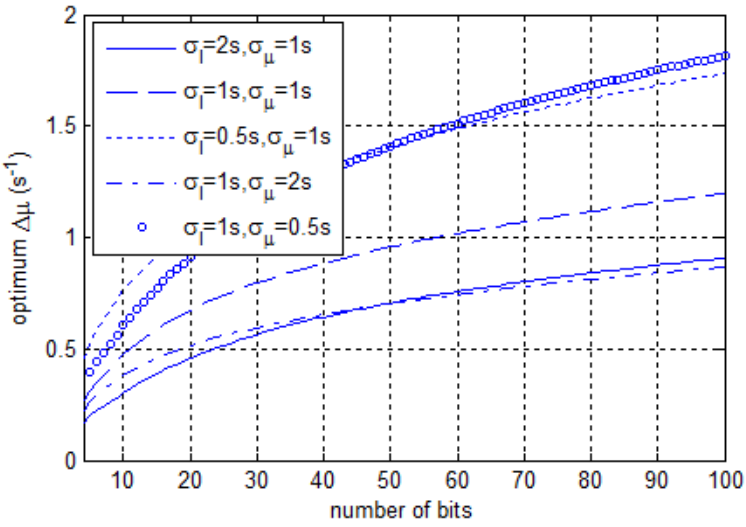

Figure 3.4: Number of bits vs optimum $\Delta\mu$ curves for random processes having GSM type autocorrelation function.

Moreover, as we did in Section 2.5, comparing the graphs corresponding to the (σ_I, σ_μ) pair $(1s, 0.5s)$ with those of $(2s, 1s)$ or comparing the graphs corresponding to $(0.5s, 1s)$ with those of $(1s, 2s)$, we see that optimal Δu increases κ times and optimal $\Delta\mu$ decreases κ times if both σ_I and σ_μ are increased κ times, the reason of which is explained after (3.59). In the same lines following (3.59), we have also explained that the optimal space-bandwidth product $\Delta u \Delta\mu$ and

the optimal number of levels K depends only on the ratio σ_I/σ_μ. The optimal $\Delta u \Delta\mu$ and K graphs are provided in Figure 3.5 and 3.6, respectively.

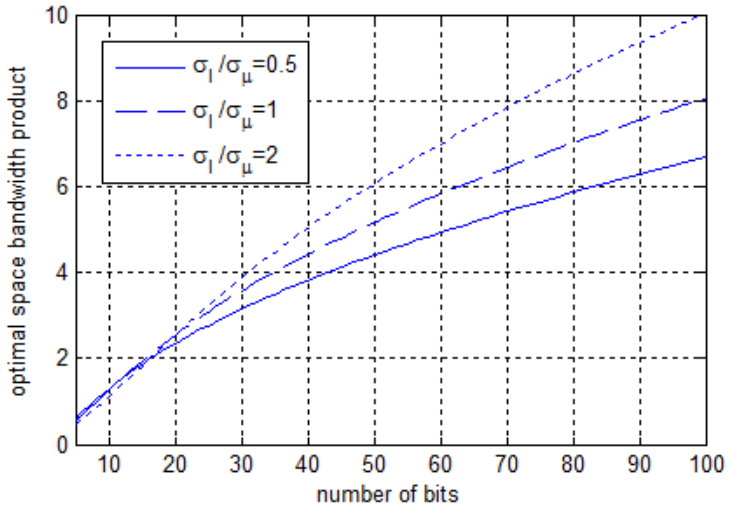

Figure 3.5: Number of bits vs optimum space-bandwidth product curves for random processes having GSM type autocorrelation function.

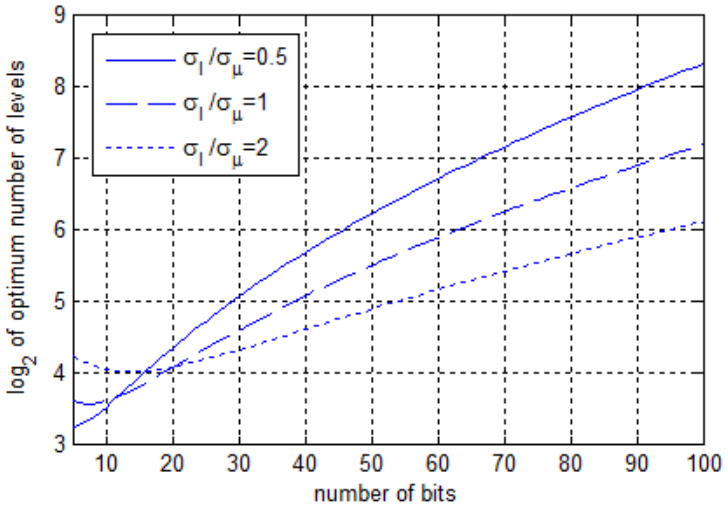

Figure 3.6: Number of bits vs optimum number of levels curves for random processes having GSM type autocorrelation function.

From Figure 3.5 and 3.6, we see that larger σ_I/σ_μ ratio results in larger optimal space-bandwidth product and smaller optimum number of levels, after number of bits exceeds a certain threshold.

3.3 Performance Comparison of Spatially Uniform and Non-Uniform Quantization

The improvement in the quantization performance when the samples having different variances are quantized differently is illustrated in this section.

From (3.34) and (3.39) in Section 3.2, we conclude that approximately $\Delta u \Delta \mu \log_2 \left(\frac{\pi e K^2}{4 \Delta u \Delta \mu} \right)$ number of bits are sufficient to obtain an average quantization error $\frac{2E_0 \Delta u \Delta \mu}{3K^2}$ for a signal class $\mathcal{F}$ the energy of the members of which is upperbounded by E_0. In other words, approximately

$$C(\epsilon_q) = \Delta u \Delta \mu \log_2 \left[\frac{\pi e (2E_0 \Delta u \Delta \mu / 3\epsilon_q)}{4 \Delta u \Delta \mu} \right] \tag{3.63}$$

$$= \Delta u \Delta \mu \log_2 \left(\frac{\pi e E_0}{6 \epsilon_q} \right) \tag{3.64}$$

number of bits are sufficient to make average quantization error ϵ_q.

Here (3.64) can be interpreted as the cost of making average quantization error ϵ_q. Conversely, if it is not allowed to exceed a specified cost C, then from (3.64), the minimum achievable average quantization error can be found as

$$\epsilon_q(C) = \frac{\pi e E_0}{6} \, 2^{-C/\Delta u \Delta \mu} \tag{3.65}$$

However, in Section 3.1, we have taken the quantization interval the same for all the samples and this results in the inefficiency of allocating redundant bits for the samples having small variances. Now, we will discuss the improvement in $\epsilon_q(C)$ if the quantization interval of the samples are chosen differently depending on the variance they have. To demonstrate this improvement, we will consider the quantization model formulated in [120]. As we mentioned in the beginning of this section, here we consider a signal class $\mathcal{F}$ (or equivalently, a random process $f(u)$) the maximum energy of the members of which is E_0.

Imitating the notation of Section 3.1, for each sample $\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})$, we denote the result of the new quantization we described as $\breve{f}^q_{\Delta\mu}(\frac{n}{\Delta\mu})$. Similarly, we use the

same notation, namely $f^q_{\Delta u,\Delta\mu}(u)$, for the reconstruction signal. Then, we repeat (3.8) here, and write

$$\int |\hat{f}_{\Delta u,\Delta\mu}(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du = \frac{1}{\Delta\mu} \sum_{n=-\lfloor \Delta u\Delta\mu/2\rfloor}^{\lfloor \Delta u\Delta\mu/2\rfloor} \left| \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) - \breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2 \tag{3.66}$$

Now, taking the expectation of both sides in (3.66), and defining $\mathbf{f}$ and $\hat{\mathbf{f}}$ as

$$\mathbf{f} = \left(\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \middle| -\left\lfloor \frac{\Delta u\Delta\mu}{2}\right\rfloor \le n \le \left\lfloor \frac{\Delta u\Delta\mu}{2}\right\rfloor \right) \tag{3.67}$$

$$\hat{\mathbf{f}} = \left(\breve{f}^q_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \middle| -\left\lfloor \frac{\Delta u\Delta\mu}{2}\right\rfloor \le n \le \left\lfloor \frac{\Delta u\Delta\mu}{2}\right\rfloor \right) \tag{3.68}$$

as done in (3.17) and (3.15) respectively, we write

$$\epsilon_q = E\left[\int |\hat{f}_{\Delta u,\Delta\mu}(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du\right] = \frac{E[||\mathbf{f}-\hat{\mathbf{f}}||_2^2]}{\Delta\mu} \tag{3.69}$$

Note that we can consider the real and imaginary parts of the samples seperately and regard $\mathbf{f}$ and $\hat{\mathbf{f}}$ as vectors in $\mathbb{R}^{2\Delta u\Delta\mu}$ rather than $\mathbb{C}^{\Delta u\Delta\mu}$. Here, as done in [120], we model the quantization as additive zero mean measurement noise $\mathbf{m} \in \mathbb{R}^{2\Delta u\Delta\mu}$ independent of $\mathbf{f}$, having independent components, each having variance $\sigma^2_{m_i}, i = 1, 2, \ldots, 2\Delta u\Delta\mu$. Then, we assume that $\mathbf{f}$ is recovered as $\hat{\mathbf{f}}$ by using a matrix $B \in \mathbb{R}^{2\Delta u\Delta\mu\times 2\Delta u\Delta\mu}$, for example a possible recovery can be $\hat{\mathbf{f}} = \mathbf{B}(\mathbf{f}+\mathbf{m})$. Note that $\mathbf{f}$, $\hat{\mathbf{f}}$, and $\mathbf{m}$ are taken as column matrices in $\mathbb{R}^{2\Delta u\Delta\mu\times 1}$ in this section. We also assume that $\mathbf{f}$ is zero mean. If there are some samples which are not zero mean, their mean can be found and subtracted, and can be added back to $\hat{\mathbf{f}}$. Therefore, there is no loss of generality in zero mean assumption.

Moreover, in this quantization model, we define the number of bits used as

$$C = \sum_{i=1}^{2\Delta u\Delta\mu} \frac{1}{2}\log_2\left(1 + \frac{\sigma^2_{f_i}}{\sigma^2_{m_i}}\right) \tag{3.70}$$

where $\sigma^2_{f_i}$ is the variance of the i$^{\text{th}}$ component of $\mathbf{f}$. This cost function is discussed in detail in [120].

The diagonal of the autocorrelation matrix $\mathbf{K_f}$ of $\mathbf{f}$ is $\sigma^2_{f_1}, \sigma^2_{f_2}, \ldots, \sigma^2_{f_{2\Delta u \Delta \mu}}$. Although the offdiagonal entries of $\mathbf{K_f}$ are not necessarily zero, $\mathbf{K_f}$ can be diagonalized as $\mathbf{K_f} = \mathbf{Q}^{\mathrm{T}}\mathbf{D}\mathbf{Q}$, where $\mathbf{Q}$ is a real $2\Delta u \Delta \mu \times 2\Delta u \Delta \mu$ unitary matrix and $\mathbf{D}$ is a diagonal matrix having the eigenvalues of $\mathbf{K_f}$, which are nonnegative, on its diagonal. Note that, in this case, the autocorrelation matrix $\mathbf{K_g}$ of the random vector $\mathbf{g} = \mathbf{Q}\mathbf{f}$ is $\mathbf{Q}\mathbf{K_f}\mathbf{Q}^{\mathrm{T}} = \mathbf{D}$. Therefore $\mathbf{g}$ has uncorrelated components.

Now, we propose to measure $\mathbf{g}$ instead of $\mathbf{f}$. In this case, we will recover $\mathbf{g}$ as $\hat{\mathbf{g}} = \mathbf{B}(\mathbf{g} + \mathbf{m})$. Then, $\mathbf{f}$ will be recovered as $\hat{\mathbf{f}} = \mathbf{Q}^{\mathrm{T}}\hat{\mathbf{g}}$. Figure 3.7 is the block diagram of the approach considered here.

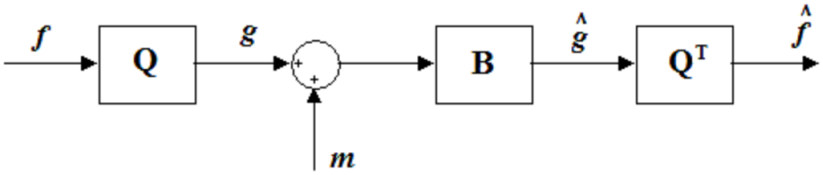

Figure 3.7: Block diagram of measurement system

Here, we also remark that error in approximating $\mathbf{f}$ by $\hat{\mathbf{f}}$ is equal to the error in approximating $\mathbf{g}$ by $\hat{\mathbf{g}}$, since

$$
\begin{aligned}
||\mathbf{f} - \hat{\mathbf{f}}||_2^2 &= \mathrm{tr}\{(\mathbf{f} - \hat{\mathbf{f}})(\mathbf{f} - \hat{\mathbf{f}})^{\mathrm{T}}\} && (3.71)\\
&= \mathrm{tr}\{\mathbf{Q}(\mathbf{f} - \hat{\mathbf{f}})(\mathbf{f} - \hat{\mathbf{f}})^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\} && (3.72)\\
&= \mathrm{tr}\{(\mathbf{Q}\mathbf{f} - \mathbf{Q}\hat{\mathbf{f}})(\mathbf{Q}\mathbf{f} - \mathbf{Q}\hat{\mathbf{f}})^{\mathrm{T}}\} && (3.73)\\
&= \mathrm{tr}\{(\mathbf{g} - \hat{\mathbf{g}})(\mathbf{g} - \hat{\mathbf{g}})^{\mathrm{T}}\} && (3.74)\\
&= ||\mathbf{g} - \hat{\mathbf{g}}||_2^2 && (3.75)
\end{aligned}
$$

Therefore, we have reduced the problem of quantizing $\mathbf{f}$ to the problem of quantizing $\mathbf{g}$, which have a diagonal autocorrelation matrix $\mathbf{K_g}$. $\mathrm{tr}(\mathbf{K_g})$ can be expressed as

$$\begin{aligned}
\mathrm{tr}(\mathbf{K_g}) &= \mathrm{tr}(\mathbf{Q}\mathbf{K_f}\mathbf{Q}^{\mathrm{T}}) && (3.76)\\
&= \mathrm{tr}(\mathbf{K_f}) && (3.77)\\
&= \sum_{i=1}^{2\Delta u\Delta\mu} \sigma_{f_i}^2 && (3.78)\\
&= \sum_{n=-\lfloor \Delta u\Delta\mu/2\rfloor}^{\lfloor \Delta u\Delta\mu/2\rfloor} E\left[\left|\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right)\right|^2\right] && (3.79)
\end{aligned}$$

Then, using (2.37) with (3.79), we get

$$\begin{aligned}
\mathrm{tr}(\mathbf{K_g}) &\leq \Delta\mu\, E\left[\int |\breve{f}_{\Delta\mu}(u)|^2\, du\right] && (3.80)\\
&\leq \Delta\mu\, E\left[\int |f(u)|^2\, du\right] && (3.81)\\
&\leq \Delta\mu E_0 && (3.82)
\end{aligned}$$

From (3.82), we see that for a number ρ between 0 and 1, we have

$$\mathrm{tr}(\mathbf{K_g}) = \rho\Delta\mu E_0 \tag{3.83}$$

Now, after expressing $\mathrm{tr}(\mathbf{K_g})$ in a convenient form, we turn our attention to finding $\mathbf{B}$ for which (3.69) is minimum. From (3.75), we see that minimizing (3.69) is fully equivalent to minimizing

$$\begin{aligned}
E[||\mathbf{g}-\hat{\mathbf{g}}||_2^2] &= E[(\mathbf{g}-\hat{\mathbf{g}})^{\mathrm{T}}(\mathbf{g}-\hat{\mathbf{g}})]\\
&= E[(\mathbf{g}-(\mathbf{B}(\mathbf{g}+\mathbf{m})))^{\mathrm{T}}(\mathbf{g}-(\mathbf{B}(\mathbf{g}+\mathbf{m})))] && (3.84)
\end{aligned}$$

for given measurement variances $\sigma_{m_1}^2, \sigma_{m_2}^2, \ldots, \sigma_{m_{2\Delta u\Delta\mu}}^2$ and the autocorrelation matrix $\mathbf{K_g} = \mathrm{diag}\{\sigma_{g_1}^2, \sigma_{g_2}^2, \ldots, \sigma_{g_{2\Delta u\Delta\mu}}^2\}$. From orthogonality condition, we have $E\{(\mathbf{g}-\mathbf{B}(\mathbf{g}+\mathbf{m}))(\mathbf{g}+\mathbf{m})^{\mathrm{T}}\} = \mathbf{0} \in \mathbb{R}^{2\Delta u\Delta\mu\times 2\Delta u\Delta\mu}$, which can be rewritten as

$$E\{((\mathbf{I}-\mathbf{B})\mathbf{g}-\mathbf{B}\mathbf{m})(\mathbf{g}^{\mathrm{T}}+\mathbf{m}^{\mathrm{T}})\} = (\mathbf{I}-\mathbf{B})\mathbf{K_g}-\mathbf{B}\mathbf{K_m} = 0 \tag{3.85}$$

since measurement noise and the input $\mathbf{f}$ are statistically independent and zero mean. Then, we find $\mathbf{B}$ as $\mathbf{B} = \mathbf{K_g}(\mathbf{K_g}+\mathbf{K_m})^{-1}$, namely

$$B_{ik} = \frac{\sigma_{g_i}^2}{\sigma_{g_i}^2+\sigma_{m_i}^2}\delta_{ik} \tag{3.86}$$

Note that (3.84) can also be expressed as

$$\sum_{i=1}^{2\Delta u\Delta\mu} E\left[\left(g_i - \sum_{k=1}^{2\Delta u\Delta\mu} B_{ik}(g_k + m_k)\right)^2\right] \tag{3.87}$$

where g_k and m_k correspond to the k$^{\text{th}}$ entry of the random vectors $\mathbf{g}$ and $\mathbf{m}$, respectively. For $\mathbf{B} = \mathbf{K_g}(\mathbf{K_g} + \mathbf{K_m})^{-1}$, this expression reduces to

$$\sum_{i=1}^{2\Delta u\Delta\mu}\left[\left(\frac{\sigma_{m_i}^2}{\sigma_{g_i}^2+\sigma_{m_i}^2}\right)^2\sigma_{g_i}^2 + \left(\frac{\sigma_{g_i}^2}{\sigma_{g_i}^2+\sigma_{m_i}^2}\right)^2\sigma_{m_i}^2\right] = \sum_{i=1}^{2\Delta u\Delta\mu}\frac{\sigma_{g_i}^2\sigma_{m_i}^2}{\sigma_{g_i}^2+\sigma_{m_i}^2}$$

After finding the error for optimal $\mathbf{B}$, to calculate $\epsilon_q(C)$, we need to obtain the measurement variances $\sigma_{m_1}^2, \sigma_{m_2}^2, \ldots, \sigma_{m_{2\Delta u\Delta\mu}}^2$ which minimize

$$\epsilon_q = \frac{E[||\mathbf{f}-\hat{\mathbf{f}}||_2^2]}{\Delta\mu} = \frac{1}{\Delta\mu}\sum_{i=1}^{2\Delta u\Delta\mu}\frac{\sigma_{g_i}^2\sigma_{m_i}^2}{\sigma_{g_i}^2+\sigma_{m_i}^2} = \frac{1}{\Delta\mu}\sum_{i=1}^{2\Delta u\Delta\mu}\left(\frac{1}{\sigma_{g_i}^2}+\frac{1}{\sigma_{m_i}^2}\right)^{-1} \tag{3.88}$$

subject to the constraint

$$\sum_{i=1}^{2\Delta u\Delta\mu}\frac{1}{2}\log_2\left(1+\frac{\sigma_{g_i}^2}{\sigma_{m_i}^2}\right) = C \tag{3.89}$$

coming from (3.70).

The solution of this optimization problem is

$$\sigma_{m_i}^2 = \begin{cases} \frac{\nu\sigma_{g_i}^2}{\sigma_{g_i}^2-\nu}, & \text{if } \sigma_{g_i}^2 > \nu \\ \infty, & \text{if } \sigma_{g_i}^2 \le \nu \end{cases} \tag{3.90}$$

where ν is chosen so that (3.89) holds, i.e.,

$$\sum_{i:\sigma_{g_i}^2>\nu}\frac{1}{2}\log_2\left(\frac{\sigma_{g_i}^2}{\nu}\right) = C \tag{3.91}$$

Then, $\epsilon_q(C)$ can be written as

$$\epsilon_q(C) = \frac{\sum_{i:\sigma_{g_i}^2>\nu}\nu + \sum_{i:\sigma_{g_i}^2\le\nu}\sigma_{g_i}^2}{\Delta\mu} \tag{3.92}$$

For the samples at which measurement is performed, namely $\sigma_{m_i}^2$ is finite, we have

$$\frac{1}{\sigma_{g_i}^2}+\frac{1}{\sigma_{m_i}^2} = \frac{1}{\nu} = \text{constant} \tag{3.93}$$

This result is also consistent with the method of Lagrange multipliers. If we define two new class of variables as $m_i' = \frac{1}{\sigma_{m_i}^2}$ and $g_i' = \frac{1}{\sigma_{g_i}^2}$, then the optimization problem can be restated as minimizing

$$\frac{1}{\Delta\mu} \sum_{i=1}^{2\Delta u\Delta\mu} \frac{1}{g_i' + m_i'} \tag{3.94}$$

subject to the constraint

$$\sum_{i=1}^{2\Delta u\Delta\mu} \frac{1}{2} \log_2 \left(1 + \frac{m_i'}{g_i'}\right) = C \tag{3.95}$$

Then, from the equation that the optimal point should satisfy

$$\frac{\partial}{\partial m_i'} \left(\sum_{i=1}^{2\Delta u\Delta\mu} \frac{1}{2} \log_2 \left(1 + \frac{m_i'}{g_i'}\right) \right) + \lambda \frac{\partial}{\partial m_i'} \left(\frac{1}{\Delta\mu} \sum_{i=1}^{2\Delta u\Delta\mu} \frac{1}{g_i' + m_i'} \right) = 0 \tag{3.96}$$

we get

$$\frac{1}{2\ln 2} \frac{1}{g_i' + m_i'} = \frac{\lambda}{\Delta\mu} \frac{1}{(g_i' + m_i')^2} \tag{3.97}$$

$$g_i' + m_i' = \frac{2\lambda \ln 2}{\Delta\mu} = \text{constant} \tag{3.98}$$

consistent with (3.93).

Now, we will analyze how $\epsilon_q(C)$ changes depending on the related parameters and compare the results with uniform quantization. First we consider the extreme case $\rho = 1$, and

$$\sigma_{g_1}^2 = \sigma_{g_2}^2 = \cdots = \sigma_{g_{2\Delta u\Delta\mu}}^2 = \frac{\Delta\mu E_0}{2\Delta\mu\Delta u} = \frac{E_0}{2\Delta u} \tag{3.99}$$

Since all the input variances are equal in this case, $\nu < \sigma_{g_i}^2 \, \forall i$. Then, (3.91) reduces to

$$\Delta u\Delta\mu \log_2 \left(\frac{E_0}{2\Delta u\nu}\right) = C \tag{3.100}$$

$$\nu = \frac{E_0}{2\Delta u} 2^{-C/\Delta u\Delta\mu} \tag{3.101}$$

From (3.92), $\epsilon_q(C)$ is found as

$$\epsilon_q(C) = 2\Delta u\nu = E_0 \, 2^{-C/\Delta u\Delta\mu} \tag{3.102}$$

Comparing this result with (3.65), we see that uniform quantization and the spatially non-uniform quantization we described in this section have similar performances if the samples have equal variances. This is an expected result since the merit of spatially nonuniform quantization is to exploit the imbalance of variances, which does not exist in this case.

Now we consider the situation when the input variances $\sigma^2_{g_1} > \cdots > \sigma^2_{g_{2\Delta u \Delta \mu}}$ decay as the pdf of a zero mean Gaussian with standart deviation $2\Delta u \Delta \mu / \alpha$. Here, the parameter α is a measure of how the variances are close to each other. $\alpha = 0$ corresponds to the extreme case when all the variances are the same, therefore spatially nonuniform quantization has the worst performance at $\alpha = 0$. On the other hand, if α is sufficiently large, then there are only few samples having significant variance which worth measuring. In this case, the performance of spatially nonuniform quantization becomes much better than uniform quantization. Figure 3.8 shows how the performance of spatially nonuniform quantization improves as α increases. The case $\alpha = 0$ effectively corresponds to uniform quantization.

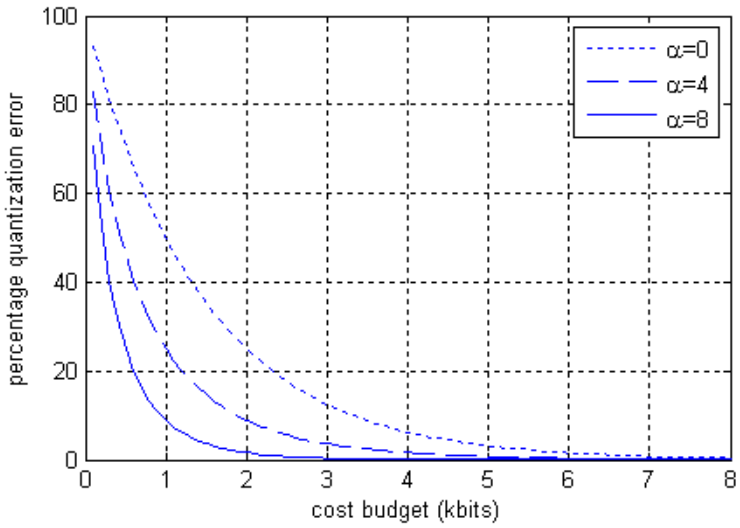

Figure 3.8: $\epsilon_q(C)$ curve for $\rho = 1$, $E_0 = 1000\,\Phi^2 s$, $\Delta u = 10\sqrt{10}\,s$, $\Delta \mu = 10\sqrt{10}\,s^{-1}$.

For fixed α and $\Delta u \Delta \mu$, the effects of other parameters such as $\rho, E_0, \Delta\mu$ on $\epsilon_q(C)$ curve are trivial. If one of the parameters $\rho, E_0, \Delta\mu$ are increased κ times, then the variances $\sigma^2_{g_1}, \sigma^2_{g_2}, ..., \sigma^2_{g_{2\Delta u \Delta \mu}}$ increase κ times. From (3.91), we see that the solution parameter ν also increases κ times. Lastly, from (3.92), we conclude that the percentage quantization error $\epsilon_q(C)/E_0 \times 100$ does not depend on E_0 and $\Delta\mu$, but increases κ times if a κ times increase is performed on ρ. However, if $\Delta\mu$ or Δu is increased independently, ρ and $\Delta\mu\Delta u$ automatically increase, resulting in the increase in the sum of variances $\mathrm{tr}(\mathbf{K_g})$ together with the increase in the samples having significant variance. Thus, $\epsilon_q(C)$ increases consequently.

Lastly, we remind that all the work we have done in this section is valid for the second FSR option as well, if $\breve{f}_{\Delta\mu}(\frac{n}{\Delta\mu})$, $\breve{f}^q_{\Delta\mu}(\frac{n}{\Delta\mu})$, $\Delta\mu$ and Δu are replaced by $\tilde{F}_{\Delta u}(\frac{n}{\Delta u})$, $\tilde{F}^q_{\Delta u}(\frac{n}{\Delta u})$, Δu and $\Delta\mu$ respectively.

3.4 The Application of Rate Distortion Theory

Firstly, we state Shannon's theorem on rate distortion theory. The notation and definitions are taken from [121].

3.4.1 Shannon's Rate Distortion Theorem

Let X be an i.i.d. (independent and identically distributed) source with distribution $p_X(x)$ and $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$ be a mapping, where

- $\mathcal{X}$ is the set of values that X can take, called set of *source alphabet*.
- $\hat{\mathcal{X}}$ is another set, called set of *reproduction alphabet*.

The function d is called as *distortion function*. Now, we extend d to the domain $\mathcal{X}^n \times \hat{\mathcal{X}}^n$ as

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) \tag{3.103}$$

Let f_n be a function with domain $\mathcal{X}^n$ and range $\{1, 2, ..., 2^{nR}\}$ and g_n be another function with domain $\{1, 2, ..., 2^{nR}\}$ and range $\hat{\mathcal{X}}^n$. Those two functions are called *encoding* and *decoding functions*, respectively. Let distortion of the pair (f_n, g_n) be defined as

$$D(f_n, g_n) = E\left[d(X^n, g_n(f_n(X^n)))\right] \tag{3.104}$$

A rate distortion pair (R, D) is called *achievable*, if there exists (f_n, g_n) pairs (having domain/range parameters n, 2^{nR}) satisfying

$$\lim_{n\to\infty} D(f_n, g_n) \leq D \tag{3.105}$$

Then, the *rate distortion function* $R(D)$ is defined as the infimum of rates R such that (R, D) is achievable for a given D.

Now, we can state Shannon's rate distortion theorem.

Theorem 6. *$R(D) = \min_S I(X; \hat{X})$, where S is the set of conditional distributions*

$$S = \{p_{\hat{X}|X}(\hat{x}|x) : E[d(X, \hat{X})] \leq D\} \tag{3.106}$$

and $I(X; \hat{X})$ is

$$I(X; \hat{X}) = E\left[\log_2 \left(\frac{p_{X,\hat{X}}(X, \hat{X})}{p_X(X) p_{\hat{X}}(\hat{X})}\right)\right] \tag{3.107}$$

the mutual information of the random variables X and $\hat{X}$, and X is the random variable having the distribution (discrete or continuous) $p(x)$ we want to decode and encode, as defined at the beginning.

3.4.2 Rate Distortion Theory and FSR

In this section, we consider a signal class $\mathcal{F}$ (or equivalently a random process $f(u)$) the average energy of which is finite. Now, after the FSR induced by the

signal

$$\hat{f}_{\Delta u,\Delta\mu}(u) = \sum_{n=-\lfloor \Delta u \Delta\mu/2 \rfloor}^{\lfloor \Delta u \Delta\mu/2 \rfloor} \breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \operatorname{sinc}(\Delta\mu\, u - n) \tag{3.108}$$

as given in (2.15), the random process $f(u)$ is reduced to the random vector

$$\mathbf{f} = \left(\breve{f}_{\Delta\mu}\left(\frac{n}{\Delta\mu}\right) \middle| -\left\lfloor \frac{\Delta u \Delta\mu}{2} \right\rfloor \le n \le \left\lfloor \frac{\Delta u \Delta\mu}{2} \right\rfloor \right) \tag{3.109}$$

as expressed in (3.67), at the expense of an approximate average error of

$$E\left[\int |f(u) - \hat{f}_{\Delta u,\Delta\mu}(u)|^2 \, du \right] \approx \int_{|u|>\Delta u/2} R(u,u)\, du + \int_{|\mu|>\Delta\mu/2} S(\mu,\mu)\, d\mu \tag{3.110}$$

as given in (2.56). In Section 3.1, we first considered scalar uniform quantization of $\mathbf{f}$ to represent $f(u)$ by finitely many bits, then used the fact that $\mathbf{f}$ is confined to a hypersphere to improve the quantization performance. In Section 3.3, we showed that the quantization performance can be improved more by quantizing the samples belonging to $\mathbf{f}$ depending on their variances. In this section, our aim is to apply rate distortion theory to see the best achievable performance for the quantization of $\mathbf{f}$.

However, since the samples constituting $\mathbf{f}$ are not i.i.d. in general, we cannot use rate distortion theory directly. To overcome this problem, we assume that i.i.d. generated realizations of the random process $f(u)$ are available. In other words, we assume the existence of a source which produces a realization of the random process $f(u)$ at each instant independent from the past and future realizations. In this case, the vectors $\mathbf{f}$ we obtain will be i.i.d., since the realizations from which these vectors are obtained are independently generated.

As an intermediate step, for a fixed n, we may consider joint encoding of i.i.d. vectors $\mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(n)}$ coming from n independent realizations $f^{(1)}(u), f^{(2)}(u), \ldots, f^{(n)}(u)$. But, rate distortion theory allows us to choose n as large as we desire to achieve an (R, D) pair.

After encoding and decoding let the vectors $\mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(n)}$ be recovered as $\hat{\mathbf{f}}^{(1)}, \hat{\mathbf{f}}^{(2)}, \ldots, \hat{\mathbf{f}}^{(n)}$. From (3.69), we see that the arithmetic mean of the expectations of the quantization error for $f^{(1)}(u), f^{(2)}(u), \ldots, f^{(n)}(u)$ is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Delta\mu}E[||\mathbf{f}^{(i)}-\hat{\mathbf{f}}^{(i)}||_2^2] = E[d(\mathbf{f}^n,\hat{\mathbf{f}}^n)] \tag{3.111}$$

as (3.103) implies, where

$$d(\mathbf{f},\hat{\mathbf{f}}) = \frac{||\mathbf{f}-\hat{\mathbf{f}}||_2^2}{\Delta\mu} \tag{3.112}$$

$$\mathbf{f}^n = (\mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(n)}) \tag{3.113}$$

$$\hat{\mathbf{f}}^n = (\hat{\mathbf{f}}^{(1)}, \hat{\mathbf{f}}^{(2)}, \ldots, \hat{\mathbf{f}}^{(n)}) \tag{3.114}$$

Therefore, we need to take the distortion function as (3.112) in order to ensure that the distortion of rate distortion theory corresponds to the expectation of our quantization error, namely $E[\int |\hat{f}_{\Delta u,\Delta\mu}(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du]$.

Now, from Shannon's rate distortion theorem given in Section 3.4.1, we conclude that to make the arithmetic mean of the expectations of the quantization error for i.i.d. generated realizations of a random process $f(u)$ equal to D,

$$R(D) = \min_{\{p(\hat{\mathbf{f}}|\mathbf{f}):E[d(\mathbf{f},\hat{\mathbf{f}})]\le D\}} I(\mathbf{f};\hat{\mathbf{f}}) \tag{3.115}$$

bits/realization are sufficient.

Needless to repeat, all the work done in this section is applicable to the case when the second option is used to obtain FSR. In this case, the vector $\mathbf{F}$ we want to quantize is

$$\mathbf{F} = \left(\tilde{F}_{\Delta u}\left(\frac{n}{\Delta u}\right)\middle| -\left\lfloor\frac{\Delta u\Delta\mu}{2}\right\rfloor \le n \le \left\lfloor\frac{\Delta u\Delta\mu}{2}\right\rfloor\right) \tag{3.116}$$

and the distortion function is

$$d(\mathbf{F},\hat{\mathbf{F}}) = \frac{||\mathbf{F}-\hat{\mathbf{F}}||_2^2}{\Delta u} \tag{3.117}$$

but the rest is the same. For the first FSR option, the joint distribution of the samples forming $\mathbf{f}$ determines the curve $R(D)$, whereas for the second FSR

option, it is the joint distribution of the samples forming $\mathbf{F}$ that determines $R(D)$.

In Section 3.1, we have considered scalar uniform quantization first. Then, based on the observation that $\mathbf{f}$ stays inside a hypersphere, we introduced vector quantization in which the samples constituting $\mathbf{f}$ are jointly encoded depending on the quantization point inside the hypersphere $\mathbf{f}$ mapped to. On the other hand, the encoding technique we discuss in this section differs from those described in previous sections, because it is based on joint encoding of the vectors $\mathbf{f}^{(i)}$ coming from consecutive independent realizations of $f(u)$. In other words, what we consider here is vector quantization of vectors, not vector quantization of individual samples. Therefore, the complexity of the encoding that we propose in this section is much more high compared to the ones considered in previous sections. But, as Shannon's rate distortion theorem implies, it is impossible to find any encoding/decoding technique having better performance than the encoder/decoder we consider in this section. Figure 3.9 and Figure 3.10 illustrate the overall finite bit reconstruction system we propose here for the first and second FSR options respectively, including the sampling part.

On the other hand, Shannon's rate distortion theorem does not tell us anything about how to reduce $f(u)$ to finitely many samples $\mathbf{f}$ consists of. In order to obtain the optimal finite bit reconstruction, we need to solve as well the problem of finding optimum Δu and $\Delta\mu$ to minimize the overall reconstruction error $E[\int |f(u) - f^q_{\Delta u,\Delta\mu}(u)|^2\, du]$. We have solved this problem in Section 3.2 for the vector quantization covered in Section 3.1 coming from hypersphere restriction, and thus optimized the sampling part as well. However, finding the optimal Δu and $\Delta\mu$ here is quite complicated and depends on the distribution of $\mathbf{f}$, as (3.115) implies.

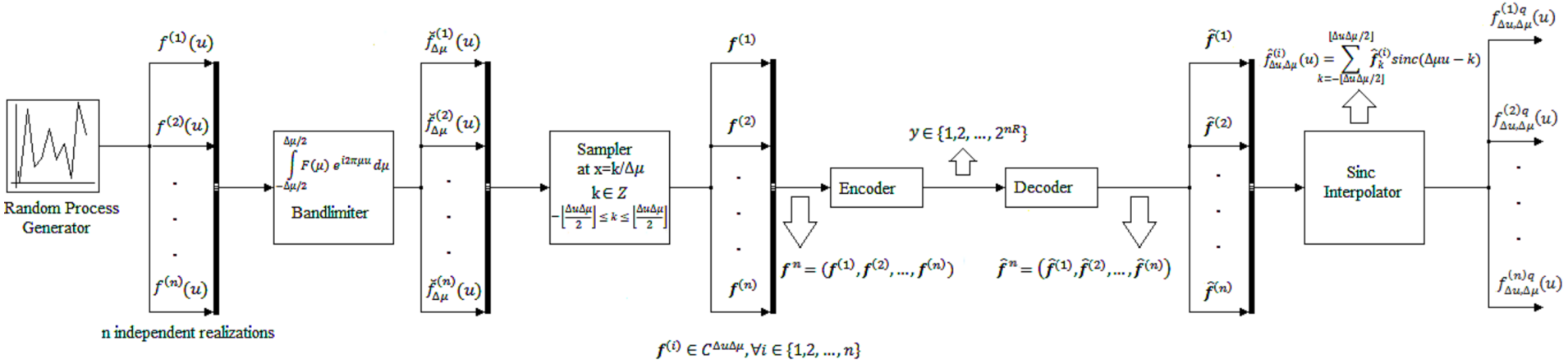

Figure 3.9: The overall finite bit reconstruction system for the first FSR option making use of the encoder/decoder of Shannon's rate distortion theorem. Each realization $f^{(i)}(u)$ is reconstructed as $f^{(i)q}_{\Delta u,\Delta\mu}(u)$.

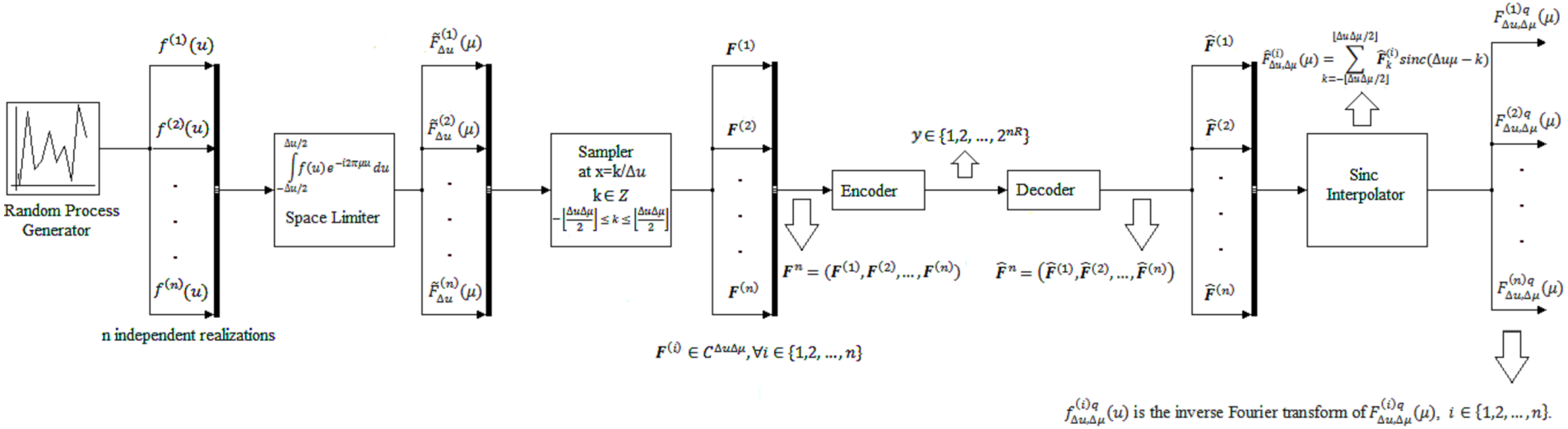

Figure 3.10: The overall finite bit reconstruction system for the second FSR option making use of the encoder/decoder of Shannon's rate distortion theorem. Each realization $f^{(i)}(u)$ is reconstructed as $f^{(i)q}_{\Delta u,\Delta\mu}(u)$.

Chapter 4

CONCLUSIONS

Any deterministic finite energy signal $f(u)$ and any random process $f(u)$ the average energy of which is finite can be reconstructed by using only finitely many samples of them with arbitrarily small error, by choosing the parameters of the reconstruction signal sufficiently large. Moreover, for the finite sample representation technique we propose, under some reasonable assumptions, the finite sample reconstruction error can be simplified as

$$\int_{|u|>\Delta u/2} |f(u)|^2 \, du + \int_{|\mu|>\Delta\mu/2} |F(\mu)|^2 \, d\mu \tag{4.1}$$

for a deterministic signal $f(u)$, and

$$\int_{|u|>\Delta u/2} R(u,u) \, du + \int_{|\mu|>\Delta\mu/2} S(\mu,\mu) \, d\mu \tag{4.2}$$

for a random process $f(u)$, where Δu and $\Delta\mu$ are the approximate spatial and spectral width of the finite sample reconstruction signal, respectively, $F(\mu)$ is the Fourier transform of the deterministic signal $f(u)$, $R(u_1, u_2)$ is the autocorrelation of the random process $f(u)$, and $S(\mu_1, \mu_2)$ is the autocorrelation of the Fourier transform of the random process $f(u)$. It is important to observe that the truncation made in space and frequency domain directly appear in the error expression without any cross terms or amplification. Here, the number

of samples used in reconstruction signal is $\Delta u \Delta \mu$, which is also equal to the space-bandwidth product of this signal.

From the method of Lagrange multipliers, we see that to minimize (4.1) for a given number of samples and to use minimum number of samples to ensure (4.1) is equal to a given constant, the optimal Δu and $\Delta \mu$ needs to satisfy the equality

$$\frac{\Delta \mu}{\Delta u} = \frac{\left|f\left(\frac{\Delta u}{2}\right)\right|^2 + \left|f\left(-\frac{\Delta u}{2}\right)\right|^2}{\left|F\left(\frac{\Delta \mu}{2}\right)\right|^2 + \left|F\left(-\frac{\Delta \mu}{2}\right)\right|^2} \tag{4.3}$$

For the random process case, from (4.2), we similarly obtain the equation of the optimal Δu and $\Delta \mu$ as

$$\frac{\Delta \mu}{\Delta u} = \frac{R(\frac{\Delta u}{2}, \frac{\Delta u}{2}) + R(-\frac{\Delta u}{2}, -\frac{\Delta u}{2})}{S(\frac{\Delta \mu}{2}, \frac{\Delta \mu}{2}) + S(-\frac{\Delta \mu}{2}, -\frac{\Delta \mu}{2})} \tag{4.4}$$

Then, using (4.4) with the constraint equation, we find optimal $(\Delta u, \Delta \mu)$ points and then we obtain the number of samples vs finite sample reconstruction error Pareto optimal curve.

If the antialiasing filter is not used before sampling, the corresponding finite sample reconstruction error is difficult to analyze. In this case, the error is upperbounded by a term greater than (4.1) for deterministic $f(u)$ and (4.2) for stochastic $f(u)$.

For any signal $f(u)$, (4.1) is greater than or equal to $1 - \sqrt{\gamma}$ fraction of its energy, where γ is the largest eigenvalue of the operator

$$Tf = \int_{-\Delta u/2}^{\Delta u/2} \Delta \mu \, \mathrm{sinc}[\Delta \mu (u - u')] \, f(u') \, du' \tag{4.5}$$

and the inequality is achieved by equality when $\int_{|u|>\Delta u/2} |f(u)|^2 \, du$ and $\int_{|\mu|>\Delta \mu/2} |F(\mu)|^2 \, d\mu$ are the same and equal to $(1 - \sqrt{\gamma})/2$ fraction of the energy of $f(u)$. As explained in [118], the eigenfunctions of the operator (4.5), namely the prolate spheroidal functions, form the optimal set for which the worst case finite sample reconstruction error of bandlimited signals is minimum. However, the family of sincs overcomes the suboptimality by a convenient shift in the sampling instants.

After representing the finite energy signal of interest by finitely many samples, the next step is quantization of these samples to reduce the signal of interest to finitely many bits. For a random process $f(u)$ none of the realizations of which have an energy larger than a certain number E_0, or equivalently for a class of signals the energy of none of the members of which exceeds E_0, scalar uniform quantization of samples makes it possible to have a quantization error less than ϵ_q for all the realizations, by using

$$\Delta u \Delta \mu \log_2 \left(\frac{2E_0 \Delta u \Delta \mu}{\epsilon_q} \right) \tag{4.6}$$

number of bits. But the vector quantization we propose achieves the same performance with

$$\Delta u \Delta \mu \log_2 \left(\frac{\pi E_0 \Delta u \Delta \mu}{2\epsilon_q} \right) - \log_2 (\Delta u \Delta \mu)! \tag{4.7}$$

number of bits. Moreover, the performance of vector quantization can be improved more by quantizing the samples differently depending on the variance they have.

For the vector quantization considered, (4.7) can be approximated as $\Delta u \Delta \mu \log_2 \left(\frac{\pi e E_0}{2\epsilon_q} \right)$. Then, using the method of Lagrange multipliers again, we see that to minimize the overall reconstruction error by using a specified number of bits and to achieve an overall reconstruction error by using minimum number of bits, the equation of the optimum Δu and $\Delta \mu$ becomes nothing but (4.4). Namely, the equation that optimal Δu and $\Delta \mu$ jointly satisfy does not change when the quantization is taken into account. After optimizing Δu, $\Delta \mu$ and the number of quantization levels, we obtain number of bits vs reconstruction error Pareto optimal curve consisting of the best achievable points, similar to the rate-distortion curve in information theory.

Rate distortion theory can be applied to our sample quantization problem if we assume that there is a source which produces a realization of the same random process $f(u)$ independent from past and future realizations. In this

case, we do not jointly encode the individual samples. What we jointly encode is the i.i.d. vectors consisting of the samples belonging to the same realization. The vector quantization of rate distortion theory cannot be outperformed by any other quantization technique as proven by Shannon in [73], therefore we know that the quantization method we consider based on rate distortion theory is the optimum one.

We considered uniform sampling with sinc interpolation in finite sample reconstruction of finite energy signals. Moreover, in quantization part, our starting point was uniform quantization. Therefore, our future work will consist of the usage of nonuniform sampling, different interpolation functions and nonuniform quantization to encode finite energy signals.

Bibliography

[1] A. Ozcelikkale, *Structural and metrical information in linear systems.* Master's thesis, Bilkent Univ., Ankara, Turkey, 2006.

[2] A. Ozcelikkale, *The Representation and Measurement of Signals in Physical Environments Progress Report.* Internal Report, Bilkent Univ., Ankara, Turkey, 2010.

[3] H. Nyquist, "Certain topics in telegraph transmission theory," *IEE. Trans.*, vol. 47, pp. 617–644, January 1928.

[4] C. E. Shannon, "Communications in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, January 1949.

[5] J. M. Whittaker, "The Fourier theory of the cardinal functions," *Proc. Mat. Soc. Edinburgh*, vol. 1, pp. 169–176, 1929.

[6] V. A. Kotel'nikov, "On the transmission capacity of "ether" and wire in electrocommunications," *Izd. Red. Upr. Svyazzi RKKA (Moscow)*, 1933.

[7] M. Zakai, "Band-limited functions and the sampling theorem," *Information and Computation/Information and Control-IANDC*, vol. 8, no. 2, pp. 143–158, 1965.

[8] A. J. Jerri, "The Shannon sampling theorem-Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, November 1977.

[9] P. L. Butzer and R. L. Stens, "Sampling theory for not necessarily band-limited functions:A historical overview," *SIAM Review*, vol. 34, pp. 40–53, March 1992.

[10] P. J. S. G. Ferreira, "On the approximation of nonbandlimited signals by nonuniform sampling series," *Proceedings of EUSIPCO-96, VIII European Signal Processing Conference*, pp. 1567–1570, September 1996.

[11] J. L. Brown, "Estimation of energy aliasing error for nonbandlimited signals," *Multdimens. Syst. Signal Process.*, vol. 15, pp. 51–56, 2004.

[12] M. Unser, "Sampling–50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, April 2000.

[13] P. P. Vaidyanathan, "Generalizations of the sampling theorem:Seven decades after Nyquist," *IEEE Trans. Circuits and Systems–I:Fundamental Theory and Appl.*, vol. 48, pp. 1094–1109, September 2001.

[14] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, 1967.

[15] K. Seip, "An irregular sampling theorem for functions bandlimited in a generalized sense," *SIAM J. Appl. Math.*, vol. 47, no. 5, pp. 1112–1116, 1987.

[16] J. J. Benedetto and W. Heller, "Irregular sampling and the theory of frames," *Math. Note*, vol. 10, pp. 103–125, 1990.

[17] J. J. Benedetto, "Irregular sampling and frames," in *Wavelets– A Tutorial in Theory and Applications* (C. K. Chui, ed.), pp. 445–507, CRC Press, 1992.

[18] H. G. Feichtinger and K. Grochenig, "Irregular sampling theorems and series expansions of band-limited functions," *J. Math. Anal. Appl.*, vol. 167, no. 2, pp. 530–556, 1992.

[19] H. G. Feichtinger and K. Grochenig, "Iterative reconstruction of multivariate band-limited functions from irregular sampling values," *SIAM J. Math. Anal.*, vol. 23, no. 1, pp. 244–261, 1992.

[20] K. Grochenig, "Reconstruction algorithms in irregular sampling," *Math. Comp.*, vol. 59, no. 199, pp. 181–194, 1992.

[21] H. G. Feichtinger, K. Grochenig, and T. Strohmer, "Efficient numerical methods in non-uniform sampling theory," *Numerische Mathematik*, vol. 69, no. 4, pp. 423–440, 1995.

[22] Y. M. Liu and G. G. Walter, "Irregular sampling in wavelet subspaces," *J. Fourier Anal. Appl.*, vol. 2, no. 2, pp. 181–189, 1995.

[23] G. S. Song and I. G. H. Ong, "Reconstruction of band-limited signals from irregular samples," *Signal Processing*, vol. 46, no. 3, pp. 315–329, 1995.

[24] P. P. Vaidyanathan and S. M. Phoong, "Reconstruction of sequences from nonuniform samples," *Proc. IEEE Int Symp. Circuits and Systems*, pp. 601–604, 1995.

[25] Y. M. Liu, "Irregular sampling for spline wavelet subspaces," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 623–627, 1996.

[26] A. Aldroubi and H. Feichtinger, "Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: The L_p theory," *Proc. Amer. Math. Soc.*, vol. 126, pp. 2677–2686, 1998.

[27] W. Chen, S. Itoh, and J. Shiki, "Irregular sampling theorems for wavelet subspaces," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1131–1142, 1998.

[28] A. Aldroubi and K. Grochenig, "Beurling–Landau-type theorems for non-uniform sampling in shift invariant spline spaces," *J. Fourier Anal. Appl.*, vol. 6, no. 1, pp. 91–101, 2000.

[29] Y. C. Eldar and A. V. Oppenheim, "Filter bank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans. Signal Processing*, vol. 48, pp. 2864–2875, October 2000.

[30] C. Zhao and P. Zhao, "Sampling theorem and irregular sampling theorem for multiwavelet subspaces," *IEEE Trans. Signal Processing*, vol. 53, pp. 707–713, February 2005.

[31] X. Zhu, A. T. S. Ho, and P. Marziliano, "A new semi-fragile image watermarking with robust tampering restoration using irregular sampling," *Signal Processing-Image Comm.*, vol. 22, pp. 515–528, June 2007.

[32] A. Nordio, C. F. Chiasserini, and E. Viterbo, "Reconstruction of multidimensional signals from irregular noisy samples," *IEEE Trans. Signal Processing*, vol. 56, pp. 4274–4285, September 2008.

[33] S. Mallat, "Multiresolution approximations and wavelet orthogonal bases of $L_2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, no. 1, pp. 69–87, 1989.

[34] S. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[35] A. Aldroubi and M. Unser, "Families of wavelet transforms in connection with Shannon sampling theory and the Gabor transform," in *Wavelets–A Tutorial in Theory and Applications* (C. K. Chui, ed.), pp. 509–528, Academic, 1992.

[36] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, 1992.

[37] G. G. Walter, "A sampling theorem for wavelet subspaces," *IEEE Trans. Inform. Theory*, pp. 881–884, March 1992.

[38] M. Vetterli and C. Herley, "Wavelets and filter banks-theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207–2232, September 1992.

[39] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Prentice-Hall, 1995.

[40] P. P. Vaidyanathan, "Sampling theorems from wavelet and filter bank theory," in *Microsystems Technology for Multimedia Applications:an Introduction* (B. Sheu, M. Ismail, E.Sanchez-Sinencio, and T. H. Wu, eds.), IEEE Press, 1995.

[41] P. P. Vaidyanathan and I. Djokovic, "Wavelet transforms," in *The Circuits and Filters Handbook* (W. K. Chen, ed.), pp. 134–219, CRC Press, 1995.

[42] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge, 1996.

[43] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic, 1998.

[44] T. Blu and M. Unser, "Quantitative Fourier analysis of approximation techniques–Part II: Wavelets," *IEEE Trans. Signal Processing*, vol. 47, pp. 2796–2806, October 1999.

[45] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Processing*, vol. 12, pp. 906–916, August 2003.

[46] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Mag.*, vol. 22, pp. 123–151, November 2005.

[47] I. J. Schoenberg, *Cardinal Spline Interpolation*. SIAM, 1973.

[48] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust.,Speech,Signal Processing*, vol. ASSP-26, no. 6, pp. 508–517, 1978.

[49] M. Unser, A. Aldroubi, and M. Eden, "Fast B–spline transforms for continuous image representation and interpolation," *IEEE Trans. Pattern Anal. Machine Intell*, vol. 13, pp. 277–285, March 1991.

[50] M. Unser, A. Aldroubi, and M. Eden, "Polynomial spline signal approximations: Filter design and asymptotic equivalence with Shannon's sampling theorem," *IEEE Trans. Inform. Theory*, vol. 38, pp. 95–103, Jan 1992.

[51] M. Unser, A. Aldroubi, and M. Eden, "On the asymptotic convergence of B–spline wavelets to Gabor functions," *IEEE Trans. Inform. Theory*, vol. 38, pp. 864–872, March 1992.

[52] M. Unser and A. Aldroubi, "Polynomial splines and wavelets," in *Wavelets–A Tutorial in Theory and Applications* (C. K. Chui, ed.), pp. 91–122, Academic, 1992.

[53] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part I–Theory," *IEEE Trans. Signal Processing*, vol. 41, pp. 821–833, February 1993.

[54] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part II–Efficient design and applications," *IEEE Trans. Signal Processing*, vol. 41, pp. 834–848, February 1993.

[55] M. Unser, "Ten good reasons for using spline wavelets," *Proc. SPIE Conf Wavelet Applications in Signal and Image Processing V*, pp. 422–431, August 1997.

[56] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Mag.*, vol. 16, no. 6, pp. 22–38, 1999.

[57] D. V. de Ville, T. Blu, M. Unser, W. Philips, I. Lemahieu, and R. V. de Welle, "Hex-splines:A novel spline family for hexagonal lattices," *IEEE Trans. Image Processing*, vol. 13, pp. 758–772, June 2004.

[58] F. Viola and W. F. Walker, "A spline-based algorithm for continuous time-delay estimation using sampled data," *IEEE Trans. Ultras.,Ferroe.,Freq. Control*, vol. 52, pp. 80–93, January 2005.

[59] L. Guo and H. Wang, "Fault detection and diagnosis for general stochastic systems using B-spline expansions and nonlinear filters," *IEEE Trans. Circuits and Systems I-Regular Papers*, vol. 52, pp. 1644–1652, August 2005.

[60] E. Masry, "On the truncation error of the sampling expansion for stationary bandlimited processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 2851–2853, October 1994.

[61] H. Choi and D. C. Munson, "Stochastic formulation of bandlimited signal interpolation," *IEEE Trans. Circuits and Systems–II: Analog and Digital Signal Processing*, vol. 47, pp. 82–85, January 2000.

[62] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.

[63] B. M. Oliver, J. Pierce, and C. E. Shannon, "The philosopy of PCM," *Proc. IRE*, vol. 36, pp. 1324–1331, November 1948.

[64] P. F. Panter and W. Dite, "Quantizing distortion in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, vol. 39, pp. 44–48, January 1951.

[65] B. Smith, "Instantaneous companding of quantized signals," *Bell Syst. Tech. J.*, vol. 36, pp. 653–709, 1957.

[66] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, March 1982.

[67] B. M. Oliver, J. Pierce, and C. E. Shannon, "Efficient coding," *Bell Syst. Tech. J.*, vol. 31, pp. 724–750, July 1952.

[68] C. W. Harrison, "Experiments with linear prediction in television," *Bell Syst. Tech. J.*, vol. 31, pp. 764–783, July 1952.

[69] P. Elias, "Predictive coding I and II," *IRE Trans. Inform. Theory*, vol. IT-1, pp. 16–33, March 1955.

[70] L. . H. Zetterberg, "A comparison between delta and pulse code modulation," *Ericsson Technics*, vol. 11, no. 1, pp. 95–154, 1955.

[71] H. V. de Weg, "Quantization noise of a single integration delta modulation system with an N-digit code," *Philips Res. Rep.*, vol. 8, pp. 568–569, August 1971.

[72] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423,623–656, 1948.

[73] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pp. 142–163, 1959.

[74] J. G. Dunn, "The performance of a class of n dimensional quantizers for a Gaussian source," *Proc. Columbia Symp. Signal Transmission Processing*, pp. 76–81, 1965.

[75] T. Berger, F. Jelinek, and J. K. Wolf, "Permutation codes for sources," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 160–169, January 1972.

[76] T. Berger, "Optimum quantizers and permutation codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 759–765, November 1972.

[77] D. L. Chaffee and J. K. Omura, "A very low rate voice compression system," *Abstracts of Papers IEEE Int. Symp. Information Theory*, October 1974.

[78] E. E. Hilbert, "Cluster compression algorithm:a joint clustering/data compression concept," *Jet Propulsion Lab., Pasadena, CA, Publication 77-43*, December 1977.

[79] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[80] J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, March 1982.

[81] J. H. Conway and N. J. A. Sloane, "Fast quantizing and decoding algorithms for lattice quantizers and codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 227–232, March 1982.

[82] J. H. Conway and N. J. A. Sloane, "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 820–824, November 1983.

[83] M. J. Sabin and R. M. Gray, "Product code vector quantizers for speech waveform coding," *Conf. Rec. GLOBECOM*, pp. 1087–1091, December 1982.

[84] M. J. Sabin and R. M. Gray, "Product code vector quantizers for waveform and voice coding," *IEEE Trans. Acoust.,Speech,Signal Processing*, vol. ASSP-32, pp. 474–488, June 1984.

[85] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 568–583, July 1986.

[86] A. Buzo, A. H. Gray, Jr., R. M. Gray, and J. D. Markel, "Speech coding based upon vector quantization," *IEEE Trans. Acoust.,Speech,Signal Processing*, vol. ASSP-28, pp. 562–574, October 1980.

[87] R. M. Gray and Y. Linde, "Vector quantizers and predictive quantizers for Gauss-Markov sources," *IEEE Trans. Commun.*, vol. COM-30, pp. 381–389, February 1982.

[88] B.-H. Juang and A. H. Gray, "Multiple stage vector quantization for speech coding," *Proc. IEEE Int. Conf. Acoustics,Speech and Signal Processing*, vol. 1, pp. 597–600, April 1982.

[89] Y.-S. Ho and A. Gersho, "Variable-rate multi-stage vector quantization for image coding," *Proc. IEEE Int. Conf. Acoustics,Speech and Signal Processing*, pp. 1156–1159, 1988.

[90] V. Cuperman and A. Gersho, "Vector predictive coding of speech at 16 Kb/s," *IEEE Trans. Commun.*, vol. COM-33, pp. 685–696, July 1985.

[91] H.-M. Hang and J. W. Woods, "Predictive vector quantization of images," *IEEE Trans. Commun.*, vol. COM-33, pp. 1208–1219, November 1985.

[92] P. C. Chang and R. M. Gray, "Gradient algorithms for designing predictive vector quantizers," *IEEE Trans. Acoust.,Speech,Signal Processing*, vol. ASSP-34, pp. 679–690, August 1986.

[93] J. S. Pan, Z. M. Lu, and S. H. Sun, "An efficient encoding algorithm for vector quantization based on subvector technique," *IEEE Trans. Image Processing*, vol. 12, pp. 265–270, March 2003.

[94] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1101–1112, March 2006.

[95] C. C. Chang, W. L. Tai, and C. C. Lin, "A reversible data hiding scheme based on side match vector quantization," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 16, pp. 1301–1308, October 2006.

[96] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 458–462, February 2007.

[97] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, October 1998.

[98] G. T. D. Francia, "Resolving power and information," *J. Opt. Soc. Amer.*, vol. 45, pp. 497–501, July 1955.

[99] D. Gabor, "Light and information," in *Progress in Optics* (E. Wolf, ed.), pp. 109–153, Elsevier, 1961.

[100] F. Gori and G. Guattari, "Shannon number and degrees of freedom of an image," *Opt. Commun.*, vol. 7, pp. 163–165, February 1973.

[101] A. Starikov, "Effective number of degrees of freedom of partially coherent sources," *J. Opt. Soc. Amer.*, vol. 72, pp. 1538–1544, 1982.

[102] O. Bucci and G. Franceschetti, "On the degrees of freedom of scattered fields," *IEEE Trans. Antennas Propag.*, vol. 37, pp. 918–926, July 1989.

[103] D. Mendlovic and A. W. Lohmann, "Space-bandwidth product adaptation and its application to superresolution: Fundamentals," *J. Opt. Soc. Amer. A*, vol. 14, pp. 558–562, March 1997.

[104] R. Piestun and D. A. B. Miller, "Electromagnetic degrees of freedom of an optical system," *J. Opt. Soc. Amer. A*, vol. 17, pp. 892–902, May 2000.

[105] A. Poon, R. Brodersen, and D. Tse, "Degrees of freedom in multiple-antenna channels:A signal space approach," *IEEE Trans. Inform. Theory*, vol. 51, pp. 523–526, February 2005.

[106] M. Migliore, "On the role of the number of degrees of freedom of the field in MIMO channels," *IEEE Trans. Antennas Propag.*, vol. 54, pp. 620–628, February 2006.

[107] J. Xu and R. Janaswamy, "Electromagnetic degrees of freedom in 2-D scattering environments," *IEEE Trans. Antennas Propag.*, vol. 54, pp. 3882–3894, December 2006.

[108] R. Kennedy, P. Sadeghi, T. Abhayapala, and H. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Signal Processing*, vol. 55, pp. 2542–2556, June 2007.

[109] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, 1971.

[110] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice-Hall, 1984.

[111] N. B. Haaser and J. A. Sullivan, *Real Analysis*. Dover, 1991.

[112] A. Koc, H. M. Ozaktas, C. Candan, and M. A. Kutay, "Digital computation of linear canonical transforms," *IEEE Trans. Signal Processing*, vol. 56, pp. 2383–2394, June 2008.

[113] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing*. Wiley, 2000.

[114] A. Starikov and E. Wolf, "Coherent-mode representation of gaussian schell-model sources and of their radiation fields," *J. Opt. Soc. Amer.*, vol. 72, July 1982.

[115] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-I," *Bell Syst. Tech. J.*, vol. 40, pp. 43–84, January 1961.

[116] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-II," *Bell Syst. Tech. J.*, vol. 40, pp. 65–84, January 1961.

[117] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-III:The dimension of the space of essentially time and bandlimited signals," *Bell Syst. Tech. J.*, vol. 41, pp. 1295–1336, July 1962.

[118] H. Dym and H. P. McKean, *Fourier Series and Integrals*. Academic Press, 1972.

[119] A. Papoulis, *Signal Analysis*. McGraw-Hill, 1977.

[120] A. Ozcelikkale, H. M. Ozaktas, and E. Arikan, "Signal recovery with cost-constrained measurements," *IEEE Trans. Signal Processing*, vol. 58, pp. 3607–3617, July 2010.

[121] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.