The Analysis and Design of Windowed Fourier Frame based Multiple Description Source Coding Schemes*

Radu Balan †
Ingrid Daubechies
Program in Applied and Computational Mathematics
Princeton University
Princeton , NJ 08544
e-mail address: rvbalan@math.princeton.edu , ingrid@math.princeton.edu
Vinay Vaishampayan
AT&T Laboratories
Florham Park , NJ 07932
e-mail address: vinay@research.att.com

August 6, 2001

Abstract

In this paper the windowed Fourier encoding-decoding scheme applied to the multiple description compression problem is analyzed. In the general case, four window functions are needed to define the encoder and decoder, although this number can be reduced to three or two by using time-shift or frequency-shift division schemes. The encoding coefficients are next divided into two groups according to the eveness of either modulation or translation index. The distortion on each channel is analyzed using the Zak transform. For the optimal windows, explicit representation formulas are obtained and non-localization results are proved. Asymptotic formulas of the total distortion and transmission rate are established and the redundancy is shown to trade-off between these two.

Key words: multiple description coding, windowed Fourier transform, redundant sets

Acknowledgments The authors wish to thank the anonymous referees for their comments and suggestions regarding the original manuscript.

*Published in: IEEE Trans on IT, vol. 46, no.7, November 2000, pp. 2491–2536

†current address: Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540

1 Introduction

The multiple description problem, a generalization of the problem of source coding subject to a fidelity criterion, is one of the fundamental problems of source coding theory. The objective of a multiple description coder is to construct several descriptions of the source sequence, with the property that the descriptions be good individually (in the rate distortion sense) and be better together. The simplest case (the one considered here) is of constructing two descriptions. Multiple description source codes are designed with the following scenario in mind. It is assumed that several (in this case two) channels connect the source to the destination, each with its own rate constraint. Each channel may fail; whether or not a channel has failed is known to the decoder but not to the encoder. The encoder wishes to send information about the source sequence over both channels, subject to the rate constraints, such that when both channels work a high fidelity replica of the source sequence is obtained, and if either channel fails, the degradation is graceful.

In addition to being an interesting and non-trivial problem in its own right, the multiple description problem is of significant practical interest because it results in compression systems that are better able to withstand frame erasures. Frame erasures are a significant problem in several communication systems of current interest, most notably wireless digital speech communications and packetized speech and video communications.

The formulation of the multiple description problem is attributed jointly to Gersho, Ozarow, Witsenhausen, Wolf, Wyner and Ziv. The main problem that information theorists consider is that of determining the rate distortion region for a given statistical model for the source and for a given fidelity criterion. Ozarow [Ozarow80], constructed the rate distortion region for the only case solved so far, namely the special case of a memoryless Gaussian source and the squared-error distortion criterion. An achievable rate region was given by El Gamal and Cover [ElGaCo82] for a memoryless source and a single-letter fidelity criterion. The binary symmetric memoryless source with an error frequency distortion criterion has been studied by Berger and Zhang [BerZha83], [ZhBe87], Ahlswede [Ahls85], Witsenhausen and Wyner [WiWy81], Wolf, Wyner and Ziv [WoWyZi80]. It was conjectured that the achievable rate region given in [ElGaCo82] coincided with the rate distortion region in cases other than the Gaussian memoryless source and the squared-error distortion criterion. However, this conjecture was disproved in [ZhBe87]. An important special case of the multiple description problem is the problem of successive refinement of information [EquCov91]. In [EquCov91], a necessary and sufficient condition for a rate distortion problem to be successively refinable is derived.

Published design techniques can loosely be divided into two categories, quantization-based approaches and subspace approaches. In quantization-based approaches, the starting point is a memoryless source and the most basic system is the multiple description quantizer [Vaish93, VaDo94]. Multiple description quantizers operate by sending information about each source sample over each channel. The simplest illustration is of two uniform step-size quantizers each with step size Δ, one offset from the other by half a step size. The first quantizer index is sent on the first channel and the second quantizer index is sent on the second channel. If both channels work, the decoder sees a quantizer with effective step size $\Delta/2$ whereas if only a single channel works, the effective step size is Δ. Thus two channels are better than one in the sense that a lower distortion is obtained when both channels work. However, when the step size is small, we are using $2R$ bits in order to obtain the performance of an $R+1$ bit quantizer, i.e., the rate overhead is almost 100%. More efficient constructions are presented in [Vaish93] and an asymptotic analysis is presented in [VaBa98].

For real-world sources such as speech and video, it is important to exploit the correlation in order to build efficient coders. Multiple description quantizers can be used efficiently for sources with memory by using standard decorrelating transforms [BaVa94, Vaish96].

Subspace methods begin by assuming that the source to be encoded is correlated. The objective is to construct two subspaces of the signal space and to send the projection of the signal on each space over a separate channel. If the spaces are well chosen, the two projections are correlated and it is possible to obtain acceptable quality when one channel is broken. Subspace methods are considered in [IngVai95], [VaSi95], [JayChr81], [OrWaVaRe97]. One common point in all the subspace methods cited above is that they do not use overcomplete expansions of the signal being coded. More recently several new approaches have been considered; some papers have considered overcomplete sets (see [ChMeWa99, GoKoVe99]); others have used vector quantization methods (see [FlEf99, SeVaSl99]), a forward error correction coding ([MoRiLa99]) or an iterative decoding approach ([Srin99]). Here we consider the design of a multiple description system based on overcomplete windowed Fourier expansions.

In this paper we use the discrete windowed Fourier transform to encode and decode the signal. A (discrete) *windowed Fourier transform* is defined by the following data: a function g called *window*, and two positive parameters $\alpha, \beta > 0$ called *modulation parameter*, respectively *translation parameter*. With these data one constructs the following set of functions called a *Weyl-Heisenberg set* (or a *WH set*):

$$\mathcal{WH}_{g;\alpha,\beta} = \{g_{mn;\alpha,\beta} \ ; \ m,n \in \mathbf{Z}\} \quad , \quad g_{mn;\alpha,\beta}(x) = e^{2\pi i m\alpha x} g(x - n\beta) \tag{1}$$

A *windowed Fourier transform-based encoder* converts a signal f into a sequence of coefficients $c_{mn} = < f, g_{mn} >$, where $< f_1, f_2 >= \int_{-\infty}^{\infty} f_1(x)\overline{f_2(x)}dx$ is the scalar product in $L^2(\mathbf{R})$ (in this paper we shall deal with continuous-time signals; in practice the signal is usually discrete and the scalar product becomes a discrete sum). The inverse operation is performed by the *windowed Fourier decoder.* This takes a double-indexed sequence of (complex) numbers $(d_{mn})_{m,n\in\mathbf{Z}}$ and returns a continuous-time signal of the form $\sum_{m,n} d_{mn} g_{mn}$. We use the notation g_{mn} signaling a double-indexed sequence of functions. Usually it is equivalent to $g_{m,n;\alpha,\beta}$ if not otherwise indicated.

Weyl-Heisenberg sets have been long studied in the literature (see [Daub90] or [HeWa89]). In Appendix A we briefly review the main known results. Two important definitions regarding these (and other sets) are the following:

DEFINITION 1 *A Weyl-Heisenberg set is called a* frame *if there are two positive constants $A, B > 0$ such that for every $f \in L^2(\mathbf{R})$:*

$$A||f||^2 \leq \sum_{m,n} | < f, g_{mn} > |^2 \leq B||f||^2 \tag{2}$$

The numbers A, B are called frame bounds. *If $A = B$, the frame is called* tight.

A WH set is called a Riesz basis for its span *(or a* s-Riesz basis*) if there are two positive constants $A, B > 0$ such that for every finite sequence $(c_{mn})_{m,n\in\mathbf{Z}}$ (i.e. only a finite number of elements are non-zero):*

$$A \sum_{m,n} |c_{mn}|^2 \leq ||\sum_{m,n} c_{mn} g_{mn}||^2 \leq B \sum_{m,n} |c_{mn}|^2 \tag{3}$$

The numbers A, B are called Riesz basis bounds. *If the WH set is simultaneously frame and s-Riesz basis, then it is simply called a* Riesz basis *(its closed span is, in this case, the entire space $L^2(\mathbf{R})$).*

An extension of a (single windowed) Weyl-Heisenberg set is given by a *Weyl-Heisenberg multiset* defined simply as a union of Weyl-Heisenberg sets. Thus, given $g^1, g^2 \in L^2(\mathbf{R})$ and $\alpha, \beta > 0$, we call $\mathcal{WH}_{(g^1,g^2);\alpha,\beta} = \mathcal{WH}_{g^1;\alpha,\beta} \cup \mathcal{WH}_{g^2;\alpha,\beta}$ a *Weyl-Heisenberg multiset.* Similarly we use the terms of *multiframe* and *multi s-Riesz basis* to suggest the multiset property as well as the frame, respectively the s-Riesz basis property of the WH multiset. Thus $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ is a *multiframe* if there are $A, B > 0$ such that for every $f \in L^2(\mathbf{R})$:

$$A\|f\|^2 \leq \sum_{m,n}(|<f, g^1_{mn}>|^2 + |<f, g^2_{mn}>|^2) \leq B\|f\|^2$$

The multiset $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ is called a *multi s-Riesz basis* if there are $A, B > 0$ such that for every $c^1, c^2 \in l^2(\mathbf{Z}^2)$:

$$A\sum_{m,n}(|c^1_{mn}|^2 + |c^2_{mn}|^2) \leq \|\sum_{m,n}(c^1_{mn}g^1_{mn} + c^2_{mn}g^2_{mn})\|^2 \leq B\sum_{m,n}(|c^1_{mn}|^2 + |c^2_{mn}|^2)$$

Weyl-Heisenberg multisets have been studied in [ZiZe97] and recently in [ChDeHe99]. In Appendix A we recall some of their results.

The block scheme of the multidescription transmission system we are proposing is given in Figure 1.

The *original signal* f is passed through the *analog encoders* defined for the first channel by a Weyl-Heisenberg set associated to the window g^1, respectively for the second channel by g^2. Their outputs represent the encoding coefficients $c^1_{mn} =< f, g^1_{mn} >$, $c^2_{mn} =< f, g^2_{mn} >$. These coefficients are passed through the *quantizers* Q_Δ and $d^1_{mn} = Q_\Delta(c^1_{mn})$, $d^2_{mn} = Q_\Delta(c^2_{mn})$ are their quantized values (we take the mid-point of the quantization inter-level). Next, the coefficients d^1_{mn}, d^2_{mn} are encoded, using for instance an *entropy encoder*, into the bit-sequences b^1_{mn}, b^2_{mn} and sent through the two channels. The receiver is made out of three decoders; the *side decoders* convert, in the first stage, the bit-sequences into the approximate coefficients $d^{1'}_{m,n}$, respectively $d^{2'}_{m,n}$ and then, in the second stage, decode these coefficients into approximating signals f^1, respectively f^2; the *central decoder* does the same thing, except for the fact that it uses both bit-streams $b^{1'}_{mn}$ and $b^{2'}_{mn}$.

Our problem is to analyze this scheme by computing the distortion, minimizing it under certain hypotheses, evaluating the transmission rate and determining the rate-distortion characteristics (or the side distortion - central distortion characteristics). As we shall see later, by varying the encoding redundancy we can *trade off between side distortions and the transmission rate.*

The paper is organized as follows: in section 2 we discuss the analog encoders and decoders, we analyze different signal models and state certain optimization problems; in section 3 we solve the optimization problems using the Zak transform; we also analyze the optimal, partial optimal and near-optimal cases by obtaining the distortion-redundancy characteristics; in section 4 we discuss the rate and obtain the rate-distortion and side distortion - central distortion characteristics; section 5 contains the conclusions and is followed by the bibliography.

2 Models and Optimization Problems

Let us now return to the block scheme in Figure 1. Suppose for the moment the quantizer does not introduce any error (for instance consider the asymptotic limit $\Delta \to 0$) and therefore the bit-stream (b^1_{mn}, b^2_{mn}) contains the same information as the coefficients (c^1_{mn}, c^2_{mn}). If both channels work, then the full information is known to the central decoder. Then we have to design the encoders/central decoder in such a way to losslessly reconstruct the original signal (recall that we make abstraction of the quantization error). This is possible only if the multiset $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ defining the encoder is a multiframe. The density result due to Christensen, Deng and Heil (see the Appendix A) shows the redundancy of this multiframe is $\frac{2}{\alpha\beta} \geq 1$. For the reconstruction (central decoder) we have many possibilities, all given by various dual frames. Even if we impose the dual frame to be given by a WH multiset, an infinite number of choices (assuming $\frac{2}{\alpha\beta} > 1$) remains (for instance see [Li95]). Among these we shall choose the standard dual frame (also known as *canonical*, or *minimal* dual frame) $(\tilde{g^1}, \tilde{g^2}; \alpha, \beta)$, whose construction will be indicated later, in the next section. [1] Let us now consider that only one channel works, say channel 1. Then the receiver knows the bit-stream $\{b^1_{m,n}\}$ solely. Unless we are prepared to spend a lot of rate, this bit-stream should contain only partial information on the original signal f (even neglecting the quantization error). Thus the coefficients $c^1_{m,n} =< f, g^1_{m,n} >$ typically represent an incomplete description of the signal. This means that the set $\mathcal{WH}_{g^1;\alpha,\beta}$ should be an incomplete set, which suggests $\frac{1}{\alpha\beta} \leq 1$ (see again Appendix A). Thus we obtain the natural condition:

$$1 \leq \frac{2}{\alpha\beta} \leq 2 \tag{4}$$

i.e. the redundancy of the original WH multiset $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ should be between 1 and 2.

On the other hand, the side decoders are assumed linear, of the form $\sum_{m,n} d^{1'}_{m,n} g^{1\#}_{m,n}$, respectively $\sum_{m,n} d^{2'}_{m,n} g^{2\#}_{m,n}$. Approximating again $d^{1'}_{m,n} = c^1_{m,n}$ we obtain: $f^1 = \sum_{m,n} c^1_{m,n} g^{1\#}_{m,n}$. In principle, the $\{g^{1\#}_{m,n}\}$ could be an arbitrary collection of functions not necessarily obtained via (1). If we impose two invariance conditions on this decoding scheme, we can show that the decoder necessarily has to be coherent, i.e. given by translations and modulations as in (1). The two (very natural) invariance conditions are the following:

1. If f is translated by β then the decoded signal f^1 translates by β as well, i.e. if $T_\beta : L^2(\mathbf{R}) \to L^2(\mathbf{R})$

[1] As reminded by one of the readers, this dual frame minimizes the reconstruction error variance in the case of white quantization error (see for instance [BolHla97], for single window case).

is the translation operator $T_\beta f(x) = f(x-\beta)$, then

$$\sum_{m,n} < T_\beta f, g^1_{m,n} > g^{1\#}_{m,n} = \sum_{m,n} < f, g^1_{m,n} > T_\beta g^{1\#}_{m,n} \quad , \quad \forall f \tag{5}$$

2. If f is modulated by $2\pi\alpha$, then the decoded signal f^1 modulates by $2\pi\alpha$ as well, i.e. if $M_{2\pi\alpha} : L^2(\mathbf{R}) \to L^2(\mathbf{R})$ is the modulation operator $M_{2\pi\alpha} f(x) = e^{2\pi\alpha i x} f(x)$ then:

$$\sum_{m,n} < M_{2\pi\alpha} f, g^1_{m,n} > g^{1\#}_{m,n} = \sum_{m,n} < f, g^1_{m,n} > M_{2\pi\alpha} g^{1\#}_{m,n} \quad , \quad \forall f \tag{6}$$

LEMMA 2 *If conditions (5) and (6) are satisfied and $\mathcal{WH}_{g^1;\alpha,\beta}$ is a s-Riesz basis, then there exists a function $g^{1\#}$ such that:*

$$g^{1\#}_{m,n}(x) = e^{2\pi i m\alpha x} g^{1\#}(x - n\beta) \tag{7}$$

Proof

Since the set $\{g^1_{m,n}\}$ is incomplete we can find f_0 such that $< f_0, g^1_{m,n} >= \delta_{m,0}\delta_{n,0}$ (where $\delta_{a,b} = 1$ if $a = b$ and 0 otherwise). From here, by setting $f_{M,N}(x) = e^{2\pi i M\alpha x} f_0(x - N\beta)$ we obtain $< f_{M,N}, g^1_{m,n} >= \delta_{m,M}\delta_{n,N}$. On the other hand one can easily check that $< T_\beta f, g^1_{m,n} >= e^{-2\pi i m\alpha\beta} < f, g^1_{m,n-1} >$ and $< M_{2\pi\alpha} f, g^1_{m,n} >=< f, g^1_{m-1,n} >$ for every f. In particular by plugging $f_{M,N}$ in (5) and (6) we get:

$$e^{-2\pi i M\alpha\beta} g^{1\#}_{M,N+1} = T_\beta g^{1\#}_{M,N} \quad , \quad g^{1\#}_{M+1,N} = M_{2\pi\alpha} g^{1\#}_{2M,N} \tag{8}$$

for every integer M, N. Iterating now these relations we obtain (7). $\square$

Assuming the same invariance hypotheses for the second channel we obtain similarly the following relations:

$$g^{2\#}_{m,n}(x) = e^{2\pi i M\alpha x} g^{2\#}(x - n\beta) \tag{9}$$

which shows that both $\{g^{1\#}_{m,n}\}$ and $\{g^{2\#}_{m,n}\}$ should be coherent, i.e. obtained as WH sets.

This discussion justifies our choice for the side decoders as given by the WH sets $\mathcal{WH}_{g^{1\#};\alpha,\beta}$, respectively $\mathcal{WH}_{g^{2\#};\alpha,\beta}$.

There are two particular choices for the encoder or decoder that we would like to single out. Both choices correspond to using a single windowed WH frame $\mathcal{WH}_{g;\alpha_0\beta_0}$ followed by a downsampling (division) of the coefficients. One possibility is to split with respect to the time shifts as follows:

$$g^1_{mn} = g_{m,2n;\alpha_0,\beta_0} \qquad g^2_{mn} = g_{m,2n+1;\alpha_0,\beta_0} \tag{10}$$

We call this encoder a *time-shift division encoder* (TSDE). Then it is obvious that c^1_{mn}, c^2_{mn} can be equivalently obtained via:

$$c^1_{mn} =< f, g^1_{mn;\alpha_0,2\beta_0} > \quad , \quad c^2_{mn} =< f, g^2_{mn;\alpha_0,2\beta_0} >$$

where $g^1 = g$ and $g^2 = T_{\beta_0} g$. Similar translation-modulation invariance conditions on the decoder ask for the following relations: $g^{1\#}_{mn} = g^{1\#}_{m,2n;\alpha_0,\beta_0}$, respectively $g^{2\#}_{mn} = g^{2\#}_{m,2n+1;\alpha_0\beta_0}$. The encoding-decoding scheme is represented in Figure 2.

The other possibility is to split the coefficients with respect to the frequency shifts as follows:

$$g^1_{mn} = g_{2m,n;\alpha_0\beta_0} \qquad g^2_{mn} = g_{2m+1,n;\alpha_0\beta_0} \tag{11}$$

We call this encoder a *frequency-shift division encoder* (FSDE). We can still obtain the encoding scheme via the general scheme presented in Figure 1, except for a constant phase factor in the second channel, which is canceled out by a similar choice of the second side-decoder. We have $c^1_{mn} =< f, g^1_{mn;2\alpha_0,\beta_0} >$ and $c^2_{mn} = e^{2\pi i n\alpha_0\beta_0} < f, g^2_{mn;2\alpha_0,\beta_0} >$ where $g^1 = g$ and $g^2 = M_{2\pi\alpha_0} g$. The invariance conditions on the decoder are satisfied if we use the following windows: $g^{1\#}_{mn} = g^{1\#}_{2m,n;\alpha_0,\beta_0}$ and $g^{2\#}_{mn} = g^{2\#}_{2m+1,n;\alpha_0,\beta_0}$. Note that $g^{2\#}_{2m+1,n;\alpha_0\beta_0} = e^{2\pi i n\alpha_0\beta_0}(M_{2\pi\alpha_0} g^{2\#})_{m,n;2\alpha_0,\beta_0}$ and this explains how the constant phase factor is removed. The encoding-decoding scheme is similar to the one in Figure 2 where the indexes $m, 2n$ and $m, 2n+1$ are replaced by $2m, n$, respectively $2m+1, n$.

Similarly, we can construct the side-decoders either by time-shift division or by frequency-shift division, of one given frame. Thus if we choose $g^{1\#}_{mn} = g^{\#}_{m,2n;\alpha_0,\beta_0}$ and $g^{2\#}_{mn} = g^{\#}_{m,2n+1;\alpha_0,\beta_0}$ we obtain the *time-shift division decoder* (TSDD) and the encoding-decoding scheme is shown in Figure 3. If we choose $g^{1\#}_{mn} = g^{\#}_{2m,n;\alpha_0,\beta_0}$ and $g^{2\#}_{mn} = g^{\#}_{2m+1,n;\alpha_0,\beta_0}$ we have the *frequency-shift division decoder* (FSDD) and the encoding-decoding scheme is analogous to the one drawn in Figure 3, where we replace the indexes $m, 2n$ and $m, 2n+1$ by $2m, n$ and $2m+1, n$, respectively.

Finally, we consider also the case when both the encoders and side-decoders are obtained by shift division. Then, the *time-shift division encoder-decoder* (TSDED) is obtained by $g^1_{mn} = g_{m,2n;\alpha_0,\beta_0}$, $g^2_{mn} = g_{m,2n+1;\alpha_0,\beta_0}$, $g^{1\#}_{mn} = g^{\#}_{m,2n;\alpha_0,\beta_0}$ and $g^{2\#}_{mn} = g^{\#}_{m,2n+1;\alpha_0,\beta_0}$. The *frequency-shift division encoder-decoder* (FSDED) is defined analogously by $g^1_{mn} = g_{2m,n;\alpha_0\beta_0}$, $g^2_{mn} = g_{2m+1,n;\alpha_0,\beta_0}$, $g^{1\#}_{mn} = g^{\#}_{2m,n;\alpha_0,\beta_0}$ and $g^{2\#} = g^{\#}_{2m+1,n;\alpha_0,\beta_0}$.

We analyze now certain signal models and we compute the one-channel approximation error.

Given a signal f, the approximation error furnished by the first side decoder is given by:

$$e_1(f) = \|\sum_{m,n} d^{1'}_{m,n} g^{1\#}_{m,n} - f\|^2 \tag{12}$$

Suppose again that $d' = c$. When no a priori information is known about the signal, a logical choice for the error measure would be to take the supremum of $e_1(f)$ over all f with $\|f\| = 1$. We obtain the following norm:

$$e_{wc} = \sup_{\|f\|=1} \|\sum_{m,n} < f, g^1_{m,n} > g^{1\#}_{m,n} - f\|^2 \tag{13}$$

The index wc stands for the *worst-case*. Indeed, e_{wc} measures the worst-case error when the encoder is fixed by g^1 and the decoder by $g^{1\#}$. Thus the designing issue seeks to solve the following optimization problem:

$$J^*_{wc} = \inf_{g^1, g^{1\#}} \sup_{\|f\|=1} \|\sum_{m,n} < f, g^1_{m,n} > g^{1\#}_{m,n} - f\|^2 \tag{14}$$

Since for $\alpha\beta > 1$ $\{g^1_{m,n}; m,n \in \mathbf{Z}\}$ is always incomplete in $L^2(\mathbf{R})$, then obviously $e_{wc} \geq 1$ (just take f to be orthogonal to all $g^1_{m,n}$). In fact it is easy to see that the optimal value in (14) is 1 for every $\alpha\beta$ with $\alpha\beta > 1$ and it is 0 for $\alpha\beta \leq 1$. The value 1 is the threshold for $\alpha\beta$ when $\{g^1_{mn;\alpha,\beta}\}_{m,n}$ may turn from an incomplete set when $\alpha\beta > 1$ into a complete one when $\alpha\beta \leq 1$ (see the Appendix A). To achieve the optimal value 1, we can choose g^1 and $g^{1\#}$ such that $\sum_{m,n} < \cdot, g^1_{m,n} > g^{1\#}_{m,n}$ represents, for instance, an orthogonal projection. This happens whenever $g^{1\#}$ is the generator of a WH set that is biorthogonal to $\mathcal{WH}_{g^1;\alpha,\beta}$, i.e. $g^{1\#} \in \overline{Span\,\mathcal{WH}_{g;\alpha,\beta}}$ and $< g^{1\#}, g_{m,n} >= \delta_{m,0}\delta_{n,0}$ for every $m,n \in \mathbf{Z}$. We shall return in the next section to the problem of finding $g^{1\#}$, the biorthogonal generator, given g^1. We note here only the discontinuity of J^*_{wc} as a function of $\frac{1}{\alpha\beta}$, at the threshold value 1 (see Figure 5 top plot). The stochastic model presented below will yield a continuous transition from 1 to 0 (see Figure 5 bottom plot).

We introduce now the stochastic model that is going to be the main topic of this paper. The abstract (mathematic) results needed to justify the formal computations are presented in Appendix B. More results and extensions are presented in [Balan98].

Our stochastic model is of a stationary signal with zero-mean and known autocovariance function:

$$\begin{aligned} \mathbf{Ef}(t) &= 0 \\ \mathbf{Ef}(t)\overline{\mathbf{f}(s)} &= R(t-s) \end{aligned} \tag{15}$$

The natural representation space can no longer be $L^2(\mathbf{R})$, the space of finite energy signals, since $\mathbf{E}|\mathbf{f}(t)^2|$ is not integrable; instead one can use the Wiener amalgam space $W(L^2, l^\infty)$, a space of finite power signals (hence the mathematical "complications" presented in Appendix B), or (less intrinsically) weighted L^2-spaces.

The approximation error is measured as an expected value of the weighted L^2-norm given by a non-negative weight function w as follows:

$$e^1_{st} = \mathbf{E}||\mathbf{f} - \sum_{m,n} < \mathbf{f}, g^1_{m,n} > g^{1\#}_{m,n}||^2_w \tag{16}$$

where $||h||^2_w = \int_{-\infty}^{\infty} |h(x)|^2 w(x)dx$. The lower index st stands for *stationary*, the upper index indicates the channel for which the approximation error is measured. Hence e^2_{st} means the approximation error of the second channel:

$$e^2_{st} = \mathbf{E}||\mathbf{f} - \sum_{m,n} < \mathbf{f}, g^2_{m,n} > g^{2\#}_{m,n}||^2_w. \tag{17}$$

Consequently, the two-channel error (which is not the reconstruction error of the central decoder) is:

$$e^{1+2}_{st} = e^1_{st} + e^2_{st} = \mathbf{E}||\mathbf{f} - \sum_{m,n} < \mathbf{f}, g^1_{m,n} > g^{1\#}_{m,n}||^2_w + \mathbf{E}||\mathbf{f} - \sum_{m,n} < \mathbf{f}, g^2_{m,n} > g^{2\#}_{m,n}||^2_w \tag{18}$$

The designing task is the following: given the stochastic model (15) and the weight function w, find the windows $g^1, g^2, g^{1\#}, g^{2\#}$ that minimize the two-channel approximation error, allow a perfect reconstruction when both channels work and the quantizer is ignored, and are well-localized in the time-frequency domain. To deal with the time-frequency localization we can append to the optimization criterion certain terms measuring the time-frequency spread, but this turned out to be very expensive computationally.

Let us state now the possible optimization problems related to the approximation errors (16)–(18). We denote by $H_1(g; \alpha, \beta)$, $H_2(g; \alpha, \beta)$ and $H(g^1, g^2; \alpha, \beta)$ the following hypotheses:

$$H_1(g; \alpha, \beta): \quad \text{The set } \mathcal{WH}_{g;\alpha,\beta} \text{ is a s-Riesz basis.} \tag{19}$$

$$H_2(g; \alpha, \beta): \quad \text{The set } \mathcal{WH}_{g;\alpha,\beta} \text{ is a frame.} \tag{20}$$

$$H(g^1, g^2; \alpha, \beta): \quad \text{The multiset } \mathcal{WH}_{(g^1,g^2);\alpha,\beta} \text{ is a multiframe.} \tag{21}$$

Ocasionally we shall use $H_1(g), H_2(g), H(g^1, g^2)$ when there is no danger of confusion. For $\alpha\beta = \frac{p}{q}$ we can also use $H_1(g; p, q), H_2(g; p, q)$ or $H(g^1, g^2; p, q)$ instead of (19), (20), or (21), respectively.

A. The Optimal Problems

A1. The One-Channel Optimal Problem:

$$e^1_{opt} = \inf_{\substack{g^1, g^{1\#} \\ H_1(g^1; \alpha, \beta)}} e^1_{st} \tag{22}$$

A2. The Two-Channel Optimal Problem

$$e^{1+2}_{opt} = \inf_{\substack{g^1, g^2, g^{1\#}, g^{2\#} \\ H(g^1, g^2; \alpha_0, 2\beta_0)}} e^{1+2}_{st} \tag{23}$$

A3. The optimal TSDE:

$$e^{TSDE}_{opt} = \inf_{\substack{g, g^{1\#}, g^{2\#} \\ H_2(g; \alpha_0, \beta_0)}} e^{1+2}_{st}(g^1 = g, g^2 = T_\beta g) \tag{24}$$

A4. The optimal TSDD:

$$e^{TSDD}_{opt} = \inf_{\substack{g^1, g^2, g^{\#} \\ H(g^1, g^2; \alpha_0, 2\beta_0)}} e^{1+2}_{st}(g^{1\#} = g^{\#}, g^{2\#} = T_\beta g^{\#}) \tag{25}$$

A5. The optimal FSDE:

$$e^{FSDE}_{opt} = \inf_{\substack{g, g^{1\#}, g^{2\#} \\ H_2(g; \alpha_0, \beta_0)}} e^{1+2}_{st}(g^1 = g, g^2 = M_{2\pi\alpha} g) \tag{26}$$

A6. The optimal FSDD:

$$e^{FSDD}_{opt} = \inf_{\substack{g^1, g^2, g^{\#} \\ H(g^1, g^2; \alpha_0, 2\beta_0)}} e^{1+2}_{st}(g^{1\#} = g^{\#}, g^{2\#} = M_{2\pi\alpha} g^{\#}) \tag{27}$$

A7. The optimal TSDED:

$$e^{TSDED}_{opt} = \inf_{\substack{g, g^{\#} \\ H_2(g; \alpha_0, \beta_0)}} e^{1+2}_{st}(g^1 = g, g^2 = T_\beta g, g^{1\#} = g^{\#}, g^{2\#} = T_\beta g^{\#}) \tag{28}$$

A8. The optimal FSDED:

$$e_{opt}^{FSDED} = \inf_{\substack{g,g^{\#} \\ H_2(g;\alpha_0,\beta_0)}} e_{st}^{1+2}(g^1 = g, g^2 = M_{2\pi\alpha}g, g^{1\#} = g^{\#}, g^{2\#} = M_{2\pi\alpha}g^{\#}) \tag{29}$$

B. The Partial Optimal Problems

The partial optimal problems are variations of the following theme: fix either the encoder or the decoder and find the optimal decoder, respectively encoder that minimizes the error. Obviously there are 16 possible problems. Each of them is an optimal problem with respect to a smaller searching space.

C. The Near-Optimal Problems

For the near-optimal problems we need to know first the optimal value for the corresponding problem. Next we give a threshold, say $\delta > 0$. The problem is then to find an encoder and/or decoder that produce an error less than $1 + \delta$ times the optimal error for the corresponding case.

For instance the *near-optimal FSDED* with threshold δ is to find a $g, g^{\#} \in L^2(\mathbf{R})$ such that $e_{st}^{1+2}(g^1 = g, g^2 = M_{2\pi\alpha}g, g^{1\#} = g^{\#}, g^{2\#} = M_{2\pi\alpha}g^{\#}) \leq (1+\delta)e_{opt}^{FSDED}$. We thus have 8 near-optimal problems associated to A1-A8.

In the real world however the total distortion is different from the approximation errors considered before. The assumption made at the beginning of this section, namely that the quantization error is negligible, may not be true. In general we should take into account all sources of error. If we do so, the total distortion has the following form:

$$J^{(1)} = \mathbf{E}\|\mathbf{f} - \sum_{(m,n)\in S} Q_\Delta(<\mathbf{f}, g_{m,n}^1>)g_{m,n}^{1\#}\|_w^2 \tag{30}$$

for the first channel, and similarly for the second channel, where S is the set of coefficients actually encoded. Using the triangle-inequality we obtain:

$$\begin{aligned} J^{(1)} &\leq \mathbf{E}\|\mathbf{f} - \sum_{m,n} <\mathbf{f}, g_{m,n}^1> g_{m,n}^{1\#}\|_w^2 + \mathbf{E}\|\sum_{(m,n)\in S}(Q_\Delta(<\mathbf{f}, g_{m,n}^1>) - <\mathbf{f}, g_{m,n}^1>)g_{m,n}^{1\#}\|_w^2 \\ &+ \mathbf{E}\|\sum_{(m,n)\notin S} <\mathbf{f}, g_{m,n}^1> g_{m,n}^{1\#}\|_w^2 = e_{st}^1 + e_q^1 + e_{tr}^1 \end{aligned} \tag{31}$$

Here: e_{st}^1 is the *stationary error* due to the incompleteness of each channel description; e_q^1 is the *quantization error* due to the quantization; e_{tr}^1 is the *truncation error* and is due to the fact that we send only a subset of the total set of coefficients.

In section 4 we analyze the quantization and truncation errors. In the next section we deal only with the stationary error and the optimization problems stated before.

3 Computations Using Zak Transform

In this section we shall compute the stationary error under the additional hypothesis that the redundancy is a rational number. Consider the general encoding-decoding scheme in Figure 1. Set

$$\alpha\beta = \frac{p}{q} \in [1,2] \tag{32}$$

3.1 Stochastic Errors

First we concentrate on the first channel and estimate the reconstruction error. Then we compute $e_{st}^{1+2} = e_{st}^1 + e_{st}^2$. For the stationary stochastic model (15) we obtain:

$$\begin{aligned}
e_{st}^1 = \mathbf{E}[\int w(x)|\mathbf{f}(x) - \sum_{m,n} < \mathbf{f}, g_{m,n}^1 > g_{m,n}^{1\#}(x)|^2 dx] \\
= \int w(x)R(0)dx - \sum_{m,n}\int\int dx\, dy\, w(x)\overline{g_{m,n}^{1\#}(x)}g_{m,n}^1(y)R(x-y) \\
-\sum_{m,n}\int\int dx\, dy\, w(x)\overline{g_{m,n}^1(y)}g_{m,n}^{1\#}(x)R(y-x) \\
+\sum_{m,n,m',n'}\int\int\int dx\, dy\, dz\, \overline{g_{m,n}^1(y)}g_{m,n}^{1\#}(x)w(x)\overline{g_{m',n'}^{1\#}(x)}g_{m',n'}^1(z)R(y-z)
\end{aligned}$$

Using Parseval's formula (see Appendix C) for the summations over m, m' we get:

$$\begin{aligned}
e_{st}^1 = R(0)\|w\|_1 - \frac{1}{\alpha}\sum_{m,n} R(\frac{m}{\alpha})\int dx\, w(x)g(x-n\beta-\frac{m}{\alpha})\overline{g^{1\#}(x-n\beta)} \\
-\frac{1}{\alpha}\sum_{m,n} R(-\frac{m}{\alpha})\int dx\, w(x)\overline{g^1(x-n\beta-\frac{m}{\alpha})}g^{1\#}(x-n\beta) \\
+\frac{1}{(\alpha)^2}\sum_{m,n,m',n'} R(\frac{m'-m}{\alpha})\int dx\, w(x)\overline{g^1(x-\frac{m}{\alpha}-n\beta)}g^{1\#}(x-n\beta)\overline{g^{1\#}(x-n'\beta)}g(x-\frac{m'}{\alpha}-n'\beta)
\end{aligned} \tag{33}$$

To progress further we use the Zak transform. As we mentioned earlier, we assume $\alpha\beta = \frac{p}{q} \geq 1$ with p and q relatively prime. The Zak transforms of the four windows $g^1, g^2, g^{1\#}$ and $g^{2\#}$ are denoted by capital letters, respectively by $G^1, G^2, G^{1\#}, G^{2\#}$ and are defined similarly to the following:

$$G(t,s) = \sqrt{\beta}\sum_{k\in\mathbf{Z}} e^{2\pi ikt}g(\beta(s+k)) \tag{34}$$

The inversion formulae in time and frequency domain are:

$$g(x) = \frac{1}{\sqrt{\beta}}\int_0^1 G(t,\frac{x}{\beta})dt \quad , \quad \hat{g}(\xi) = \sqrt{\frac{\beta}{2\pi}}\int_0^1 e^{-i\beta s\xi}G(-\frac{\beta\xi}{2\pi}, s)ds \tag{35}$$

We note here our convention regarding the Fourier transform: $\hat{g}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\xi t} g(t)dt$. For more information on the Zak transform we refer the reader to [Jans82],[Jans88]. We recall here two quasi-periodicity relations that will be used throughout this section:

$$G(t+1,s) = G(t,s) \quad , \quad G(t,s+1) = e^{-2\pi i t} G(t,s) \tag{36}$$

We denote by $\Gamma(t,s)$ the $p \times q$ matrix whose (j,k) entry is $G(t+\frac{k}{q}, s+j\frac{q}{p})$, $j = 0,\ldots,p-1$, $k = 0,\ldots,q-1$, i.e.

$$\Gamma(t,s) = \begin{bmatrix} G(t,s) & G(t+\frac{1}{q},s) & \cdots & G(t+\frac{q-1}{q},s) \\ G(t,s+\frac{q}{p}) & G(t+\frac{1}{q},s+\frac{q}{p}) & \cdots & G(t+\frac{q-1}{q},s+\frac{q}{p}) \\ \vdots & \vdots & & \vdots \\ G(t,s+(p-1)\frac{q}{p} & G(t+\frac{1}{q},s+(p-1)\frac{q}{p}) & \cdots & G(t+\frac{q-1}{q},s+(p-1)\frac{q}{p}) \end{bmatrix} \tag{37}$$

We define similarly the matrices $\Gamma^1(t,s), \Gamma^2(t,s), \Gamma^{1\#}(t,s)$ and $\Gamma^{2\#}(t,s)$. We also define the following transforms of the autocovariance function, respectively of the weight function:

$$\rho_r(t) = \sum_m e^{2\pi i m q t} R(\frac{mp+r}{\alpha}) \quad , \quad r = 0,\ldots,p-1 \tag{38}$$

$$\omega(s) = \sum_k w(\beta(s+k)) \tag{39}$$

Let us denote by $\mathbf{M}(t)$ the $p \times p$ matrix whose (r_1,r_2) entry is $\rho_{r_1-r_2}(t)$, i.e.

$$\mathbf{M}(t) = \begin{bmatrix} \rho_0(t) & \cdots & \rho_{-(p-1)}(t) \\ \vdots & & \vdots \\ \rho_{p-1}(t) & \cdots & \rho_0(t) \end{bmatrix} \tag{40}$$

Note the following properties:

$$\begin{array}{ccc} \rho_{r+p}(t) = e^{-2\pi i q t}\rho_r(t) & & \rho_r(t+\frac{1}{q}) = \rho_r(t) \\ \rho_{-r}(t) = \overline{\rho_r(t)} & \Rightarrow & \bar{\mathbf{M}}^T = \mathbf{M} \end{array} \tag{41}$$

i.e. for fixed t, $\mathbf{M}(t)$ is self-adjoint as a matrix (we shall also use $\mathbf{M}^*$ instead of $\bar{\mathbf{M}}^T(t)$). Thus $\mathbf{M}(t)$ is a $p \times p$ self-adjoint Toeplitz matrix.

Using $W(s)$ defined in (39) we construct a $p \times p$ diagonal matrix $\mathbf{W}(s)$ whose (r,r) entry is $\omega(s+r\frac{q}{p})$, i.e.

$$\mathbf{W}(s) = \begin{bmatrix} \omega(s) & & & \\ & \omega(s+\frac{q}{p}) & & 0 \\ & & \ddots & \\ & 0 & & \omega(s+(p-1)\frac{q}{p}) \end{bmatrix} \tag{42}$$

Since $\omega(s)$ is 1-periodic the diagonal of $\mathbf{W}(s)$ contains a permutation of the $\omega(s+\frac{r}{p})$, $r = 0,\ldots,p-1$.

Using these notations and the Parseval identity, the error e_{st}^1 given by (33) turns into:

$$e_{st}^1 = \beta q \int_0^{\frac{1}{p}} ds \int_0^{\frac{1}{q}} dt \;\; trace\{\mathbf{W}(I - \frac{1}{p}\Gamma^{1\#}\Gamma^{1^*})\mathbf{M}(I - \frac{1}{p}\Gamma^1\Gamma^{1\#^*})\} =: e(\mathbf{W}, \mathbf{M}; \Gamma^1, \Gamma^{1\#}; \alpha, \beta) \tag{43}$$

where $\Gamma^1 = \Gamma^1(t,s)$, $\Gamma^{1\#} = \Gamma^{1\#}(t,s)$, $\mathbf{M} = \mathbf{M}(t)$, $\mathbf{W} = \mathbf{W}(s)$ and I is the $p \times p$ identity matrix. In order to make explicit dependence of the error on the window functions g and $g^\#$, we also use the notation

$$e(w, R; g, g^\#; \alpha, \beta) = e(\mathbf{W}, \mathbf{M}; \Gamma, \Gamma^\#; \alpha, \beta). \tag{44}$$

A full account of these computations can be found in [Balan98].

The other channel error is given by:

$$e_{st}^2 = \mathbf{E}\|\mathbf{f} - \sum_{m,n} < \mathbf{f}, g_{m,n}^2 > g_{m,n}^{2\#}\|_w^2$$

Then the same derivation leads to:

$$e_{st}^2 = e(\mathbf{W}, \mathbf{M}, \Gamma^2, \Gamma^{2\#}; \alpha, \beta) \tag{45}$$

We turn now to the special cases of time-shift division and frequency-shift division encoders and decoders. Both cases can be treated by adapting formula (43) to the specific context. We take $\alpha\beta = 2\alpha_0\beta_0 = \frac{p}{q}$.

In the *time-shift division* case the reconstruction operator on channel 1 has the form:

$$S_1^{TSD}\mathbf{f} = \sum_{m,n} < \mathbf{f}, g_{m,2n;\alpha_0,\beta_0}^1 > g_{m,2n;\alpha_0,\beta_0}^{1\#} = \sum_{m,n} < \mathbf{f}, g_{m,n;\alpha_0,2\beta_0}^1 > g_{m,n;\alpha_0,2\beta_0}^{1\#}$$

and thus the reconstruction error is:

$$e_{st}^{1;TSD} = e(\mathbf{W}, \mathbf{M}; \Gamma^1, \Gamma^{1\#}; \alpha_0, 2\beta_0) \tag{46}$$

with $\mathbf{W}, \mathbf{M}, \Gamma^1, \Gamma^{1\#}$ defined as before but for $\beta = 2\beta_0$, $\alpha = \alpha_0$. For the second channel we obtain:

$$S_2^{TSD} f = \sum_{m,n} < \mathbf{f}, g_{m,2n+1;\alpha_0,\beta_0}^2 > g_{m,2n+1;\alpha_0,\beta_0}^{2\#} = T_\beta(\sum_{m,n} < T_{-\beta}\mathbf{f}, g_{m,n;\alpha_0,2\beta_0}^2 > g_{m,n;\alpha_0,2\beta_0}^{2\#})$$

The error is then:

$$e_{st}^{2;TSD} = \mathbf{E}\|\mathbf{f} - S_2^{TSD}\mathbf{f}\|_w^2 = \mathbf{E}\|T_\beta(T_{-\beta}\mathbf{f} - \sum_{m,n} < T_{-\beta}\mathbf{f}, g_{m,n;\alpha_0,2\beta_0}^2 > g_{m,n;\alpha_0,2\beta_0}^{2\#})\|_w^2 =$$
$$\mathbf{E}\|T_{-\beta}\mathbf{f} - \sum_{m,n} < T_{-\beta}\mathbf{f}, g_{m,n;\alpha_0,2\beta_0}^2 > g_{m,n;\alpha_0,2\beta_0}^{2\#}\|_w^2$$

because of the stationarity of the reconstruction error. Next, $T_{-\beta}\mathbf{f}$ has the same second order statistics as $\mathbf{f}$ does, because of the stationarity of the signal itself. Thus $T_{-\beta}$ can be dropped out from the above formula and we obtain:

$$e_{st}^{2;TSD} = e(\mathbf{W}, \mathbf{M}; \Gamma^2, \Gamma^{2\#}; \alpha_0, 2\beta_0) \tag{47}$$

The *frequency-shift division* case is similar, but a bit different. For channel 1 the reconstruction error does not raise any difficulty since:

$$S_1^{FSD}\mathbf{f} = \sum_{m,n} < \mathbf{f}, g^1_{2m,n;\alpha_0,\beta_0} > g^{1\#}_{2m,n;\alpha_0,\beta_0} = \sum_{m,n} < \mathbf{f}, g^1_{m,n;2\alpha_0,\beta_0} > g^{1\#}_{m,n;2\alpha_0,\beta_0}$$

Thus:

$$e_{st}^{1;FSD} = e(\mathbf{W}, \mathbf{M}; \Gamma^1, \Gamma^{1\#}; 2\alpha_0, \beta_0) \tag{48}$$

On the other hand, for the channel 2 we obtain:

$$S_2^{FSD}\mathbf{f} = \sum_{m,n} < \mathbf{f}, g^2_{2m+1,n;\alpha_0,\beta_0} g^{2\#}_{2m+1,n;\alpha_0,\beta_0} = M_{2\pi\alpha_0}(\sum_{m,n} < M_{-2\pi\alpha_0}\mathbf{f}, g^2_{m,n;2\alpha_0,\beta_0} > g^{2\#}_{m,n;2\alpha_0,\beta_0})$$

which implies:

$$\begin{aligned} e_{st}^{2;FSD} &= \mathbf{E}\|\mathbf{f} - S_2^{FSD}\mathbf{f}\|_w^2 = \mathbf{E}\|M_{2\pi\alpha_0}(M_{-2\pi\alpha_0}\mathbf{f} - \sum_{m,n} < M_{-2\pi\alpha_0}\mathbf{f}, g^2_{m,n;2\alpha_0,\beta_0} > g^{2\#}_{m,n;2\alpha_0,\beta_0})\|_w^2 \\ &= \mathbf{E}\|M_{-2\pi\alpha_0}\mathbf{f} - \sum_{m,n} < M_{-2\pi\alpha_0}\mathbf{f}, g^2_{m,n;2\alpha_0,\beta_0} > g^{2\#}_{m,n;2\alpha_0,\beta_0}\|_w^2 \end{aligned}$$

Note now that $M_{-2\pi\alpha_0}\mathbf{f}$ has a different second order statistics than $\mathbf{f}$ has. Indeed:

$$\mathbf{E}[M_{-2\pi\alpha_0}\mathbf{f}(t)\overline{M_{-2\pi\alpha_0}\mathbf{f}(s)}] = e^{2\pi i\alpha_0(s-t)}R(t-s) \tag{49}$$

Thus the reconstruction error becomes:

$$e_{st}^{2;FSD} = e(\mathbf{W}, \mathbf{N}; \Gamma^2, \Gamma^{2\#}; 2\alpha_0, \beta_0) \tag{50}$$

where $\mathbf{N}$ is a $p \times p$ Toeplitz matrix obtained similarly to $\mathbf{M}$ but for the autocovariance function (49). Let us denote by $\sigma_j(t)$ the entries of $\mathbf{N}(t)$. Then an easy computation shows that:

$$\mathbf{N}(t) = \begin{bmatrix} \sigma_0(t) & \cdots & \sigma_{-(p-1)}(t) \\ \vdots & & \vdots \\ \sigma_{p-1}(t) & \cdots & \sigma_0(t) \end{bmatrix} \tag{51}$$

with

$$\sigma_r(t) = (-1)^r \rho_r(t + \alpha_0\beta_0) \tag{52}$$

Note that σ_r and $\mathbf{N}$ have similar properties to those of ρ_r and $\mathbf{M}$ given by (41).

The last two cases we consider are the TSDED and the FSDED. Both schemes are defined by two windows g and $g^{\#}$ only. The total error in the TSDED case is obtained by adding up $e_{st}^{1,TSD}$ and $e_{st}^{2,TSD}$ for $g^1 = g^2 = g$ and $g^{1\#} = g^{2\#} = g^{\#}$. Thus we obtain:

$$e_{st}^{1+2;TSDED} = 2e(\mathbf{W}, \mathbf{M}; \Gamma, \Gamma^{\#}; \alpha_0, 2\beta_0) \tag{53}$$

In the FSDED case we have to add together $e_{st}^{1,FSD}$ and $e_{st}^{2,FSD}$ and we obatin:

$$e_{st}^{1+2;FSDED} = e(\mathbf{W}, \mathbf{M}+\mathbf{N}; \Gamma, \Gamma^{\#}; 2\alpha_0, \beta_0) \tag{54}$$

The expressions (53) and (54) look very similar; note however that, because one of them uses $2\beta_0$ and the other β_0 as the time translation unit, our analysis uses different Zak transforms (in which the translation unit enters), so that $\Gamma, \Gamma^{\#}$ have different forms in the two formulas.

Next we analyze the three hypotheses (19,20,21) stated in the previous section as well as the central decoder construction.

3.2 Biorthogonal and Dual Generators

The hypotheses impose different conditions on a WH set or multiset. The s-Riesz basis and frame conditions on $\mathcal{WH}_{g;\alpha,\beta}$, in terms of the Zak transform, have long been studied (see [Daub90] and [HeWa89]). Similarly one can obtain necessary and sufficient conditions on a WH multiset to become a multiframe (see [ZiZe95]). These conditions can be stated as follows:

THEOREM 3 *A. Consider $\mathcal{WH}_{g;\alpha,\beta}$ a WH set. Suppose $\alpha\beta = \frac{p}{q}$ with p, q relatively prime integers. Let us denote by G, G^1, G^2 the Zak transforms of these windows with respect to the parameter β, and by Γ, Γ^1, Γ^2 the $p \times q$ matrices obtained similarly to (37). Then:*

1. $\mathcal{WH}_{g;\alpha,\beta}$ is a s-Riesz basis with bounds A, B iff for a.e. $(t,s) \in [0, \frac{1}{q}] \times [0, \frac{1}{p}]$

$$A \leq \frac{1}{p}\Gamma^*\Gamma \leq B \tag{55}$$

where the inequalities are understood in the quadratic forms sense (i.e. $M \geq \alpha, \alpha \in \mathbf{R}$ iff $< x, Mx > \geq \alpha\|x\|^2$). Moreover, the standard biorthogonal s-Riesz basis generator $g^{\#}$ is given through the following *relation:*

$$\Gamma^{\#} = p\Gamma(\Gamma^*\Gamma)^{-1} \tag{56}$$

2. $\mathcal{WH}_{g;\alpha,\beta}$ is a frame with bounds A, B iff for a.e. $(t,s) \in [0,\frac{1}{q}] \times [0,\frac{1}{p}]$

$$A \leq \frac{1}{p}\Gamma\Gamma^* \leq B \tag{57}$$

Moreover, the standard dual frame generator $\tilde{g}$ is given by:

$$\tilde{\Gamma} = p(\Gamma\Gamma^*)^{-1}\Gamma \tag{58}$$

B. Consider now $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ a WH multiset. Suppose again $\alpha\beta = \frac{p}{q}$ with p, q integers and consider the same notations as before.

The multiset $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ is a WH multiframe with bounds A, B iff for a.e. $(t,s) \in [0,\frac{1}{q}] \times [0,\frac{1}{p}]$

$$A \leq \frac{1}{p}(\Gamma^1{\Gamma^1}^* + \Gamma^2{\Gamma^2}^*) \leq B \tag{59}$$

In this case the standard dual frame generator $(\tilde{g}^1, \tilde{g}^2)$ is given by:

$$\tilde{\Gamma^1} = p(\Gamma^1{\Gamma^1}^* + \Gamma^2{\Gamma^2}^*)^{-1}\Gamma^1 \tag{60}$$

$$\tilde{\Gamma^2} = p(\Gamma^1{\Gamma^1}^* + \Gamma^2{\Gamma^2}^*)^{-1}\Gamma^2 \tag{61}$$

This result is known in the literature, in various places. Part B is perhaps the least known, since WH multisets have been studied less. In Appendix D we sketch its proof.

REMARK 4 *Note in (55), the s-Riesz basis condition is stated in terms of $\Gamma^*\Gamma$ because $p \geq q$, whereas in (57) the frame condition involves the product $\Gamma\Gamma^*$, since now $q \geq p$.*

This result makes it possible to obtain the central decoder. Indeed, in the TSDE or FSDE cases, the full bit-stream b_{mn} is obtained by encoding the expansion coefficients of the signal f with respect to a frame $\mathcal{WH}_{g;\alpha,\beta}$. On the other hand the standard (minimal) dual frame minimizes the reconstruction error variance when the coefficients are perturbed by an additive independent white noise. Thus, the standard dual frame is a logical choice for the central decoder, and that is what we choose. In the other cases when the encoder is a WH multiframe we choose its standard dual multiframe (60) and (61) for the central decoder (note in the FSDD case the central decoder should preprocess the coefficients of the second channel by shifting them with a constant phase $e^{-2\pi i n\alpha_0\beta_0}$).

3.3 The Partial Optimal Problems

Let us now concentrate on the partial optimal problems. Note first of all the symmetry in terms of the encoding and decoding problem. Indeed, because we can make circular permutations and take adjoints under the trace we have:

$$e(\mathbf{W}, \mathbf{M}; \Gamma^1, \Gamma^{1\#}; \alpha, \beta) = e(\mathbf{M}, \mathbf{W}; \Gamma^{1\#}, \Gamma^1; \alpha, \beta) \tag{62}$$

For the one-channel partial optimal problem with fixed encoder we have to minimize $e(\mathbf{W}, \mathbf{M}; \Gamma, \Gamma^{\#}; \alpha, \beta)$ with respect to $\Gamma^{\#}$, for $\alpha\beta > 1$. This is clearly equivalent with minimizing the following trace:

$$\min_X \ trace\{\mathbf{W}(I - \frac{1}{p}X\Gamma^*)\mathbf{M}(I - \frac{1}{p}\Gamma X^*)\} \tag{63}$$

for $X \in \mathbf{C}^{p\times q}$, with $\mathbf{W}, \mathbf{M} \in C^{p\times p}$ hermitian and $\Gamma \in \mathbf{C}^{p\times q}$ given, because each $X(t,s)$ can be chosen independently from each other for $(t,s) \in [0, \frac{1}{q}] \times [0, \frac{1}{p}]$. The optimal solution is given by:

$$\Gamma^{\#}_{so} = p\mathbf{M}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1} \tag{64}$$

Note it does not depend on $\mathbf{W}$, however the optimal value of the error does:

$$e(\mathbf{W}, \mathbf{M}; \Gamma, \Gamma^{\#}_{so}; \alpha, \beta) = \beta q \int_0^{\frac{1}{p}} ds \int_0^{\frac{1}{q}} dt \ trace\{\mathbf{W}\mathbf{M} - \mathbf{W}\mathbf{M}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1}\Gamma^*\mathbf{M}\} =: e_{so}(\mathbf{W}, \mathbf{M}; \Gamma; \alpha, \beta) \tag{65}$$

The explicit solutions that will be presented assume implicitely the encoding hypotheses (19,20,21) are satisfied. We shall comment on this fact later in the subsequent subsections when the optimal problems are considered.

We introduce also another expression that will be useful next:

$$\begin{aligned} e_{so}^{1,2}(\mathbf{W}, \mathbf{M}_1, \mathbf{M}_2; \Gamma^1, \Gamma^2; \alpha, \beta) = \beta q \int_0^{\frac{1}{p}} ds \int_0^{\frac{1}{q}} dt \ & trace\{\mathbf{W}(\mathbf{M}_1 + \mathbf{M}_2) \\ & - \mathbf{W}(\mathbf{M}_1\Gamma^1 + \mathbf{M}_2\Gamma^2)(\Gamma^{1*}\mathbf{M}_1\Gamma^1 + \Gamma^{2*}\mathbf{M}_2\Gamma^2)^{-1}(\Gamma^{1*}\mathbf{M}_1 + \Gamma^{2*}\mathbf{M}_2)\} \end{aligned}$$

The solutions to the 16 partial optimal problems are given in the following theorem whose proof in given in Appendix D:

THEOREM 5 *Suppose the matrices* $\mathbf{W}(s)$ *and* $\mathbf{M}(t), \mathbf{N}(t)$ *are strictly positive with bounded inverse. Then the solutions of the partial optimal problems are given by:*

B1. *The one-channel partial optimal problem with fixed encoder:*

$$\Gamma^{1\#}{}_{so} = p\mathbf{M}\Gamma^1(\Gamma^{1^*}\mathbf{M}\Gamma^1)^{-1} \tag{66}$$

$$e^1_{st;so} = e_{so}(\mathbf{W}, \mathbf{M}; \Gamma^1; \alpha, \beta) \tag{67}$$

B2. *The one-channel partial optimal problem with fixed decoder:*

$$\Gamma^1{}_{so} = p\mathbf{W}\Gamma^{1\#}(\Gamma^{1\#^*}\mathbf{W}\Gamma^{1\#})^{-1} \tag{68}$$

$$e^{1\#}_{st,so} = e_{so}(\mathbf{M}, \mathbf{W}; \Gamma^{1\#}; \alpha, \beta) \tag{69}$$

B3. *The two-channel partial optimal problem with fixed encoder*

$$\Gamma^{1\#}{}_{so} = p\mathbf{M}\Gamma^1(\Gamma^{1^*}\mathbf{M}\Gamma^1)^{-1} \quad , \quad \Gamma^{2\#}{}_{so} = p\mathbf{M}\Gamma^2(\Gamma^{2^*}\mathbf{M}\Gamma^2)^{-1} \tag{70}$$

$$e^{1+2}_{st,so} = e_{so}(\mathbf{W}, \mathbf{M}; \Gamma^1; \alpha, \beta) + e_{so}(\mathbf{W}, \mathbf{M}; \Gamma^2; \alpha, \beta) \tag{71}$$

B4. *The two-channel partial optimal problem with fixed decoder*

$$\Gamma^1{}_{so} = p\mathbf{W}\Gamma^{1\#}(\Gamma^{1\#^*}\mathbf{W}\Gamma^{1\#})^{-1} \quad , \quad \Gamma^2{}_{so} = p\mathbf{W}\Gamma^{2\#}(\Gamma^{2\#^*}\mathbf{W}\Gamma^{2\#})^{-1} \tag{72}$$

$$e^{1+2\#}_{st,so} = e_{so}(\mathbf{M}, \mathbf{W}; \Gamma^{1\#}; \alpha, \beta) + e_{so}(\mathbf{M}, \mathbf{W}; \Gamma^{2\#}; \alpha, \beta) \tag{73}$$

B5. *The partial optimal TSDE with fixed encoder:*

$$\Gamma^{1\#}{}_{so} = \Gamma^{2\#}{}_{so} = p\mathbf{M}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1} \tag{74}$$

$$e^{1+2,TSDE}_{st,so} = 2e_{so}(\mathbf{W}, \mathbf{M}; \Gamma; \alpha_0, 2\beta_0) \tag{75}$$

B6. *The partial optimal TSDE with fixed decoder:*

$$\Gamma_{so} = p\mathbf{W}(\Gamma^{1\#} + \Gamma^{2\#})(\Gamma^{1\#^*}\mathbf{W}\Gamma^{1\#} + \Gamma^{2\#^*}\mathbf{W}\Gamma^{2\#})^{-1} \tag{76}$$

$$e^{1+2,TSDE\#}_{st,so} = e^{1,2}_{so}(\mathbf{M}, \mathbf{W}, \mathbf{W}; \Gamma^{1\#}, \Gamma^{2\#}; \alpha_0, 2\beta_0) \tag{77}$$

B7. *The partial optimal TSDD with fixed encoder:*

$$\Gamma^{\#}{}_{so} = p\mathbf{M}(\Gamma^1 + \Gamma^2)(\Gamma^{1^*}\mathbf{M}\Gamma^1 + \Gamma^{2^*}\mathbf{M}\Gamma^2)^{-1} \tag{78}$$

$$e^{1+2,TSDD}_{st,so} = e^{1,2}_{so}(\mathbf{W}, \mathbf{M}, \mathbf{M}; \Gamma^1, \Gamma^2; \alpha_0, 2\beta_0) \tag{79}$$

B8. *The partial optimal TSDD with fixed decoder:*

$$\Gamma^1{}_{so} = \Gamma^2{}_{so} = p\mathbf{W}\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-1} \tag{80}$$

$$e^{1+2,TSDD}_{st,so} = 2e_{so}(\mathbf{M}, \mathbf{W}; \Gamma^{\#}; \alpha_0, 2\beta_0) \tag{81}$$

B9. *The partial optimal FSDE with fixed encoder:*

$$\Gamma^{1\#}{}_{so} = p\mathbf{M}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1} \quad , \quad \Gamma^{2\#} = p\mathbf{N}\Gamma(\Gamma^*\mathbf{N}\Gamma)^{-1} \tag{82}$$

$$e^{1+2,FSDE}_{st,so} = e_{so}(\mathbf{W},\mathbf{M};\Gamma,2\alpha_0,\beta_0) + e_{so}(\mathbf{W},\mathbf{N};\Gamma;2\alpha_0,\beta_0) \tag{83}$$

B10. *The partial optimal FSDE with fixed decoder:*

$$\Gamma_{so} = X \text{ solution of the linear system : } \mathbf{M}X\Gamma^{1\#^*} = p(\mathbf{M}\mathbf{W}\Gamma^{1\#} + \mathbf{N}\mathbf{W}\Gamma^{2\#}) \tag{84}$$

$$e^{1+2,FSDE\#}_{st,so} = \beta q \int_0^{\frac{1}{p}} ds \int_0^{\frac{1}{q}} dt\; trace\{2\mathbf{W}\mathbf{M} - \frac{1}{p}X(\Gamma^{1\#^*}\mathbf{W}\mathbf{M} + \Gamma^{2\#^*}\mathbf{W}\mathbf{N})\} \tag{85}$$

B11. *The partial optimal FSDD with fixed encoder:*

$$\Gamma^{\#}{}_{so} = p(\mathbf{M}\Gamma^1 + \mathbf{N}\Gamma^2)(\Gamma^{1^*}\mathbf{M}\Gamma^1 + \Gamma^{2^*}\mathbf{N}\Gamma^2)^{-1} \tag{86}$$

$$e^{1+2,FSDD}_{st,so} = e^{1,2}_{so}(\mathbf{W},\mathbf{M},\mathbf{N};\Gamma^1,\Gamma^2;2\alpha_0,\beta_0) \tag{87}$$

B12. *The partial optimal FSDD with fixed decoder:*

$$\Gamma^1{}_{so} = \Gamma^2{}_{so} = p\mathbf{W}\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-1} \tag{88}$$

$$e^{1+2,FSDD\#}_{st,so} = e_{so}(\mathbf{M}+\mathbf{N},\mathbf{W};\Gamma^{\#};2\alpha_0,\beta_0) \tag{89}$$

B13. *The partial optimal TSDED with fixed encoder:*

$$\Gamma^{\#}{}_{so} = p\mathbf{M}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1} \tag{90}$$

$$e^{1+2,TSDED}_{st,so} = 2e_{so}(\mathbf{W},\mathbf{M};\Gamma;\alpha_0,2\beta_0) \tag{91}$$

B14. *The partial optimal TSDED with fixed decoder:*

$$\Gamma_{so} = p\mathbf{W}\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-1} \tag{92}$$

$$e^{1+2,TSDED\#}_{st,so} = 2e_{so}(\mathbf{M},\mathbf{W};\Gamma^{\#};\alpha_0,2\beta_0) \tag{93}$$

B15. *The partial optimal FSDED with fixed encoder:*

$$\Gamma^{\#}{}_{so} = p(\mathbf{M}+\mathbf{N})\Gamma(\Gamma^*(\mathbf{M}+\mathbf{N})\Gamma)^{-1} \tag{94}$$

$$e^{1+2,FSDED}_{st,so} = e_{so}(\mathbf{W},\mathbf{M}+\mathbf{N};\Gamma;2\alpha_0,\beta_0) \tag{95}$$

B16. *The partial optimal FSDED with fixed decoder:*

$$\Gamma_{so} = p\mathbf{W}\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-1} \tag{96}$$

$$e^{1+2,FSDED\#}_{st,so} = e_{so}(\mathbf{M}+\mathbf{N},\mathbf{W};\Gamma^{\#};2\alpha_0,\beta_0) \tag{97}$$

REMARK 6 *1. We point out that the requirements (such as strict positivity) on* $\mathbf{W}(s), \mathbf{M}(t)$ *and* $\mathbf{N}(t)$ *are not necessary. They are sufficient conditions ensuring that the suboptimal windows* $g^1_{so}, g^2_{so}, g^{1\#}_{so}, g^{2\#}_{so}$ *belong to* $L^2(\mathbf{R})$*; in most cases these conditions can be relaxed. Each case can be dealt with separately, but we shall not go further into this here. However, in practice we are interested in more regularity than simply square integrability; we are interested in smoothness for our windows as well. We return to this issue in the designing step, in section 4 below.*

2. As mentioned before, we assume that each solution satisfies the s-Riesz basis and multiframe condition. Explicit conditions for these hypotheses will be given for the optimal problems.

4 The Optimal Problems

4.1 Spaces of Eigenvalue, Eigenvector and Eigenspace-Valued Maps

The following objects are useful in the analysis of the optimal solutions. The unit 2-dimensional square is symbolized by $\square$, $\square = [0,1] \times [0,1]$. Recall that the Hilber-Schmidt scalar product of two matrices $A, B \in \mathbf{C}^{p \times q}$ is defined by:

$$< A, B >_{HS} = Trace\{A^* B\} = \sum_{i=1}^{p} \sum_{j=1}^{q} \overline{A_{ij}} B_{ij} \tag{98}$$

Similarly, the Hilbert-Schmidt norm of $A \in \mathbf{C}^{p \times q}$ is defined as usual by $\|A\|_{HS} = (< A, A >_{HS})^{1/2}$. Then we can easily define several L^p spaces of matrix-valued functions. In particular we define two spaces:

$$L^2(\square; \mathbf{C}^{p \times q}) \quad := \quad \{A : \square \to \mathbf{C}^{p \times q} \mid \|A\|^2_{L^2(\square;\mathbf{C}^{p\times q})} := \int\int_{\square} \|A(t,s)\|^2_{HS} dt\, ds < \infty\} \tag{99}$$

$$L^\infty(\square; GL_q(\mathbf{C})) \quad := \quad \{A : \square \to \mathbf{C}^{q \times q} \mid \|A\|_{L^\infty(\square;GL_q(\mathbf{C}))} := \sup_{(t,s)\in\square} \|A(t,s)\|_{HS} < \infty \tag{100}$$

$$\text{and} \quad \sup_{(t,s)\in\square} \|A^{-1}(t,s)\|_{HS} < \infty\}$$

Note that $L^\infty(\square; GL_q(\mathbf{C}))$ is not a linear space (for instance the constant zero matrix does not belong to this space), but a group with respect to the matrix multiplication. Consider now the space of $p \times p$ nonnegative symmetric matrices $Sym^+_p(\mathbf{C})$. It is a convex cone and the trace is a pseudometric, $d(S_1, S_2) = |Trace\{S_1 - S_2\}|$, for $S_1, S_2 \in Sym^+_p(\mathbf{C})$. Using this pseudometric we construct the L^1 space of $Sym^+_p(\mathbf{C})$-valued functions as follows:

$$L^1(\square; Sym^+_p(\mathbf{C})) = \{S : \square \to Sym^+_p(\mathbf{C}) \mid \|S\|_{L^1} := \int\int_{\square} Trace\{S(t,s)\}\, dt\, ds < \infty\} \tag{101}$$

$Sym_p^+(\mathbf{C})$ is then extended to the space of $Sym_p^+(\mathbf{C})$-equivalent matrices, denoted $Plus_p(\mathbf{C}) = \{TAT^{-1}\,;\, A \in Sym_p^+(\mathbf{C}), T \in GL_p(\mathbf{C})\}$. Thus a matrix $M \in Plus_p(\mathbf{C})$ if and only if it is diagonalizable and all its eigenvalues are positive real numbers. Accordingly, $L^1(\Box; Sym_p^+(\mathbf{C}))$ extends to:

$$L^1(\Box; Plus_p(\mathbf{C})) = \{B : \Box \to Plus_p(\mathbf{C}) \mid \|B\|_{L^1} := \int\int_{\Box} Trace\{B(t,s)\}dt\,ds < \infty\} \tag{102}$$

Note that $L^1(\Box; Sym_p^+(\mathbf{C}))$ is a subset of the space $L^1(\Box; Plus_p(\mathbf{C}))$ which in turn is a subset of the set of *trace-class* operators over $L^2(\Box; \mathbf{C}^p)$. Moreover, the following property holds true as well:

PROPOSITION 7 *If $S, R \in L^2(\Box; Sym_p^+(\mathbf{C}))$ then $S{\cdot}R \in L^1(\Box; Plus_p(\mathbf{C}))$ and $\|S{\cdot}R\|_{L^1} \leq \|S\|_{L^2_{HS}}\|R\|_{L^2_{HS}}$.*

Proof Suppose $A \in Sym_p^+(\mathbf{C})$ then there is a unique $A^{1/2} \in Sym_p^+(\mathbf{C})$ such that $A^{1/2}A^{1/2} = A$. For every $\varepsilon > 0$ and $B \in Sym_p^+(\mathbf{C})$ we have the following equivalence $(A+\varepsilon I)B \sim (A+\varepsilon I)^{1/2}B(A+\varepsilon I)^{1/2} \in Sym_p^+(\mathbf{C})$. Now, by continuity of the spectrum with respect to (compact) perturbations, we may take the limit $\varepsilon \searrow 0$ and obtain $AB \sim A^{1/2}BA^{1/2}$. Thus $AB \in Plus_p(\mathbf{C})$. Consequently, for S, R satisfying our hypothesis, we have for every point (t,s), $S \cdot R|_{(t,s)} \in Plus_p(\mathbf{C})$. Next, note $Trace\{S^*R\} =< S, R >_{HS} \leq \|S\|_{HS}\|R\|_{HS}$ by the Cauchy-Schwartz inequality and thus $S \cdot R \in L^1(\Box; Plus_p(\mathbf{C}))$ which ends the proof. $\diamond$

Consider now S a matrix-valued function in $L^1(\Box; Plus_p(\mathbf{C}))$. We want to study the eigenproblem solution for $S(t,s)$. We are interested in the eigenvalue and the eigenvector maps.

At each point $(t,s) \in \Box$, the eigenvalues are well-defined and positive. The eigenvectors may not be uniquely defined if one or more of the eigenvalues is degenerate. Let us denote by $\lambda_1(t,s), \ldots, \lambda_p(t,s)$ the p monotonically decrasing ordered eigenvalues of $S(t,s)$ at (t,s). Thus we obtain p real-valued maps over the unit-square, $\lambda_j : \Box \to \mathbf{R}^+$, $1 \leq j \leq p$. The following result characterizes these eigenvalue maps:

THEOREM 8 *Consider $S \in L^1(\Box; Plus_p(\mathbf{C}))$ and $\lambda_1, \ldots, \lambda_p : \Box \to \mathbf{R}^+$ the monotonically decreasing ordered eigenvalue maps as above.*

1. For every $1 \leq j \leq p$, λ_j is measurable (with respect to the standard Lebesgue measure) and $\lambda_j \in L^1(\Box; \mathbf{R}^+)$;

2. If the entries of S are continuous complex-valued functions on $\Box$, then so are $\lambda_j : \Box \to \mathbf{R}^+$, $j = 1, \ldots, p$;

3. Suppose the entries of S are differentiable at some $(t,s) \in \Box$ (i.e. the real and imaginary parts of S_{lk} are differentiable at (t,s), for $1 \leq l,k \leq p$) and $\lambda_j(t,s)$ is nondegenerate, then λ_j is differentiable at

(t,s).

Proof of Theorem 8

2. Part 2 is a standard result in matrix perturbation theory (see for instance Theorem IV.1.3 due to Elsner in [StSu90]) and it is usually proved using complex analysis methods.

3. Part 3 is also standard (see Theorem IV.2.3 in [StSu90]) and it is proved using the Gerschgorin's disks technique.

1. For the first assertion we use first the density of $C(\Box;Plus_p(\mathbf{C}))$ in $L^1(\Box;Plus_p(\mathbf{C}))$. Then we consider a sequence $(S^{(n)})_{n\geq 1}$ in $C(\Box;Plus_p(\mathbf{C}))$ that converges to S in $L^1(\Box;Plus_p(\mathbf{C}))$ sense. Then we extract a subsequence $(n_k)_k$ that converges pointwise almost everywhere on S. Next, the sequences $(\lambda_j^{n_k})_{k\geq a}$, $j=1,\ldots,p$ of the ordered eigenvalues of $S^{(n_k)}$ are bounded and necessarily converge to $\lambda_{\pi(j)}$, some eigenvalue of S, where π is a point-dependent permutation of $\{1,2,\ldots,n\}$. Because of the ordering of $(\lambda_j^{n_k})$ we have $\pi = id$. Thus $\lambda_j(x) = \lim_{k\rightarrow\infty}\lambda_j^{n_k}(x)$ pointwisely and each $\lambda_j^{n_k}$ is continuous, hence measurable and in $L^1(\Box;\mathbf{R}^+)$. Therefore λ_j is measurable as well. Since $0\leq\lambda_j(x)\leq TraceS(x)$ we obtain $\int\int_\Box \lambda_j(t,s)dt\,ds \leq \int\int_\Box Trace\,S(t,s)dt\,ds < \infty$. Thus for every j, $\lambda_j\in L^1(\Box;\mathbf{R}^+)$. End of proof. $\diamond$

We consider now the eigenvectors problem. Unfortunately there are no easy answers to the continuity problem for the eigenvectors. The difficulty arises whenever the eigenvalue is degenerate. For the nondegenerate eigenvalues the problem is relatively easy and the answer is furnished by the spectral theory.

Given a $p\times p$ matrix $A\in\mathbf{C}^{p\times p}$ and a closed curve $\Lambda\in\mathbf{C}$ that does not pass through any eigenvalue of A (i.e. $Spec(A)\cap\Lambda=\emptyset$), then

$$P_\Lambda = \frac{1}{2\pi i}\int_\Lambda (\lambda I - A)^{-1}d\lambda \tag{103}$$

defines a projection onto the spectral space associated to the eigenvalues included in the interior of Λ (see for instance [Kato80]). A *spectral space* associated to some eigenvalue λ is the largest invariant space of A such that the restriction of A to this invariant space has the spectrum made only of λ. If A is selfadjoint, P_Λ is an orthonormal projection and any spectral space is exactly the eigenspace associated to some eigenvalue. In general the eigenspace is included in the spectral space, but the inclusion may be strict. However, when A diagonalizes the two spaces are always of the same dimension.

Suppose now $A = A(t)$ is a $p \times p$ matrix whose entries are continuously parametrized by t, and for $t \in I$, A has a simple eigenvalue at $\lambda_j(t)$; we choose to number the eigenvalues so that λ_j is continuous in t (see above). Then the projection onto e_j (the eigenvector associated to λ_j) is a continuous function in the space of rank one projectors. Next, using a transformation function adapted to P_j (see [Kato80], Cap. II, §4.2) we can construct a continuous map of eigenvectors from I into $\mathbf{C}^p$: $t \mapsto e_j(t)$, associated to the map λ_j. The same argument can be carried over to any spectral projection of constant rank. Thus we get the following result:

PROPOSITION 9 *Let $A : I \to Plus_p(\mathbf{C})$ be a continuous $Plus_p(\mathbf{C})$-valued map defined on an open set I ($I \subset \mathbf{C}^k$, for some $k > 0$) and let $\lambda_1(t) \geq \cdots \geq \lambda_p(t)$ be the monotonically ordered continuous system of eigenvalues. Suppose for some $1 \leq q \leq p$, $\lambda_q(t) > \lambda_{q+1}(t)$ on I, then there are q orthonormal vectors $e_1, \ldots, e_q$, continuously defined on I that form an orthonormal basis for the spectral space of A associated to $\{\lambda_1, \ldots, \lambda_q\}$.* $\diamond$

REMARK 10 *1. Explicitely, $e_1, \ldots, e_q$ have the following property: for every $t \in I$, if E is their q-dimensional span in $\mathbf{C}^p$, then E is an invariant space for A and the spectrum of A restricted to E is exactly $\{\lambda_1, \ldots, \lambda_q\}$, i.e. $Spec(A|_E) = \{\lambda_1, \ldots, \lambda_q\}$.*

2. The result in [Kato80] does not yield directly the orthonormal system, but rather a basis for each t in the spectral space. From there it is straightforward to obtain an orthonormal basis (for instance by Gramm-Schmidt) which will depend continuously on t as well.

In order to deal with the degenerate case we have to be more careful. Consider now $A : \Box \to Plus_p(\mathbf{C})$ a function in $L^\infty(\Box; Plus_p(\mathbf{C}))$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ its system of eigenvalues. Fix q, $1 \leq q \leq p$. Let us denote by $\mathcal{L}_q(\mathbf{C}^p)$ the set of q-dimensional subspaces of $\mathbf{C}^p$. We define the following two important sets of maps:

$$\begin{aligned} eigspacemax(A; p, q) &= \{\mathcal{V} : \Box \to \mathcal{L}_q(\mathbf{C}^p) \mid \text{for every } (t,s) \in \Box, \mathcal{V}(t,s) \text{ is } A(t,s) - \text{invariant} \\ &\quad\; dim\,\mathcal{V}(t,s) = q \;\text{ and }\; Spec(A|_{\mathcal{V}}) = \{\lambda_1, \ldots, \lambda_q\}\ \} \end{aligned} \tag{104}$$

$$\begin{aligned} eigmax(A; p, q) &= \{F : \Box \to \mathbf{C}^{p \times q} \mid F^*(t,s)F(t,s) = I_q\ ,\ F \text{ is measurable and} \\ &\quad\; \exists \mathcal{V} \in eigspacemax(A; p, q) \;\text{ s.t. } Ran\, F = \mathcal{V}\ \} \end{aligned} \tag{105}$$

We shall also use the notations $eigspacemax(A; \alpha, \beta)$ or $eigmax(A; \alpha, \beta)$ with $\alpha\beta = \frac{p}{q}$ to denote the same objects, when there is no danger of confusion. Note that $\mathcal{V}$ is not necessarily a spectral space, because we do not require $\mathcal{V}$ in $eigspacemax$ to be maximal as invariant space (in fact we couldn't have required this because of the dimension constraint). If $\lambda_q > \lambda_{q+1}$ then $\mathcal{V}$ is a spectral space, though. In this case Proposition 9 proves there is a continuous $F \in eigmax(A; p, q)$; moreover $eigspacemax(A; p, q)$ contains only one map, namely the one associates for every $(t, s) \in \Box$ the spectral space of $\{\lambda_1, \dots, \lambda_q\}$. In the case when $\lambda_q = \lambda_{q+1}$ $eigspacemax(A; p, q)$ will contain infinitely many maps. Note the columns of $F \in eigmax(A; p, q)$ form an orthonormal basis for some $\mathcal{V} \in eigspacemax(A; p, q)$.

For a better characterization of these objects we introduce the following indices. For every (t, s) we define the *degeneracy indices* $k_1(t, s)$ and $k_2(t, s)$ as follows:

$$k_1(t, s) = \max_k \{k \mid \lambda_{q-k} = \lambda_q\} \ , \ k_2(t, s) = \max_k \{k \mid \lambda_{q+k} = \lambda_q\} \tag{106}$$

Thus the eigenvalues at (t, s) are ordered as follows:

$$\lambda_{q-k_1-1} > \lambda_{q-k_1} = \dots = \lambda_q = \dots = \lambda_{q+k_2} > \lambda_{q+k_2+1} \tag{107}$$

Then we define:

$$D_{l_1,l_2} = \{(t, s) \in \Box \mid k_1(t, s) = l_1 \ , \ k_2(t, s) = l_2\} \tag{108}$$

and

$$D_l = \cup_{j=0}^{q} D_{j,l} \tag{109}$$

Note that we always have $0 \leq k_1 \leq q$ and $0 \leq k_2 \leq p - q$ always. An easy exercise shows the following properties of these sets:

1) $D_{l_1,l_2} \cap D_{m_1,m_2} = \emptyset$, for every $(l_1, l_2) \neq (m_1, m_2)$;

2) $D_{0,0} = \{(t, s) \in \Box \mid \lambda_{q-1}(t, s) > \lambda_q(t, s) > \lambda_{q+1}(t, s)\}$ is open;

3) $D_0 = \{(t, s) \in \Box \mid \lambda_q(t, s) > \lambda_{q+1}(t, s)\}$ is open;

4) $D_0 \cup D_1 \cup \ldots \cup D_l$ is open;

5) $\cup_{0 \leq l_1 \leq m_1} \cup_{0 \leq l_2 \leq m_2} D_{l_1,l_2}$ is open for every m_1, m_2.

With these notations Proposition 9 implies that on D_0 we can construct (continuous) q-system of orthonormal basis in the spectral space of $\{\lambda_1, \dots, \lambda_q\}$ (assuming A is continuous). Note also that $eigspacemax(A; p, q)$ restricted to D_0 contains only one map. We are now ready to discuss the optimal problems and the localization result.

4.2 The Optimal Problems

Once we have solved the partial optimal problems we can optimize over the remaining freedom in the choice of window functions, i.e. we now concentrate on the eight optimal problems stated in (22)–(29).

We shall solve exactly (i.e. in a closed form) seven of these eight problems. For the eighth problem (the FSDE problem) we provide upper and lower bounds for the approximation error. The exact value of the optimal error in this particular case can be obtained by solving a continuously parametrized finite dimensional optimization problem. For all the cases the optimal solution represents a Karhunen-Loève approximation of the original stochastic signal. For the remainder of this section we assume $\mathbf{M}(t)$ is bounded and invertible for almost every $t \in [0,1]$ (more precisely, $\mathbf{M} \in L^\infty([0,1]; GL_p(\mathbf{C}))$).

We are going to study separately each of the eight optimal problems. In the following we use the notations introduced before.

4.2.1 The One-Channel Optimal Problem

The one-channel optimal problem is the simplest and, in some sense, represents a benchmark for the other optimization problems.

Recall the one-channel structure involves two WH s-Riesz bases $\mathcal{WH}_{g;\alpha,\beta}$, $\mathcal{WH}_{g^\#;\alpha,\beta}$ and the communication structure contains one encoding and one decoding block. If g, respectively $g^\#$, denotes the encoding, respectively decoding window, then the optimal solution is obtained by solving one of the following optimization problems:

$$e^1_{st;o} := \inf_{\substack{g \\ H_1(g;\alpha,\beta)}} \inf_{\substack{g^\# \\ H_1(g^\#;\alpha,\beta)}} e(w,R;g,g^\#;\alpha,\beta) = \inf_{\substack{g^\# \\ H_1(g^\#;\alpha,\beta)}} \inf_{\substack{g \\ H_1(g;\alpha,\beta)}} e(w,R;g,g^\#;\alpha,\beta) \tag{110}$$

where $e(w,R;g,g^\#;\alpha,\beta)$ was defined in (44). Solving the first optimization problem is equivalent to first solving the partial optimal one-channel problem with fixed encoder, and then optimizing over the encoders . The second form in (110) means to optimize the partial optimal one-channel problem with fixed decoder over all admissible decoders, and then to optimize over the encoders. The solution is given by the following:

THEOREM 11 (One-Channel) *The optimal solutions of the one-channel optimal problem are parametrized in the Zak transform domain as follows:*

$$\Gamma_{opt}(t,s) = F(t,s) \cdot L(t,s) \tag{111}$$

$$\Gamma^\#_{opt}(t,s) = F^\#(t,s) \cdot L^\#(t,s) \tag{112}$$

where:

$$F \in eigmax(\mathbf{WM};\alpha,\beta)\ ,\ F^{\#} = \mathbf{M}F(F^*\mathbf{M}^2F)^{-1/2} \in eigmax(\mathbf{MW};\alpha,\beta) \tag{113}$$

and

$$L \in L^\infty(\Box; GL_q(\mathbf{C}))\ ,\ L^{\#} = p(F^*\mathbf{M}^2F)^{1/2}(F^*\mathbf{M}F)^{-1}L^{-*} \in L^\infty(\Box; GL_q(\mathbf{C})) \tag{114}$$

L^{-*} denotes the hermitian conjugate of the inverse: $L^{-*} = (L^{-1})^*$. Recall from §4.1 that $eigmax(S;\alpha,\beta)$ represents the set of $p \times q$-matrices of functions whose columns at each point are orthonormal vectors and span an invariant space of S corresponding to the largest q eigenvalues.

The optimal value of the error turns into:

$$e^1_{st;opt} = e_{opt}(\mathbf{WM};\alpha,\beta) := \beta q \int_0^{1/p} ds \int_0^{1/q} dt \sum_{i=q+1}^{p} \lambda_i(t,s) \tag{115}$$

where $(\lambda_i)_{i=\overline{1,p}}$ are the p real eigenvalues of $\mathbf{WM}|_{(t,s)}$ decreasingly ordered as $\lambda_1 \geq \ldots \geq \lambda_p$.

Proof First note that in the Zak domain the optimization problem decouples into independent finite-dimensional optimization problems continuously parametrized by (t,s). For a fized (t,s) we have to minimize

$$Trace\{\mathbf{WM} - \mathbf{WM}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1}\Gamma^*\mathbf{M}\}$$

over Γ. Assume $\mathbf{M}$ is invertible at (t,s). Notice that $P_\Gamma := \mathbf{M}^{1/2}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1}\Gamma^*\mathbf{M}^{1/2}$ is an orthogonal projection for any choice of Γ for which $\Gamma^*\mathbf{M}\Gamma$ is invertible. Moreover, for every q-dimensional subspace $\mathcal{V}$ of $\mathbf{C}^p$, there is a Γ such that $P_\Gamma = P_\mathcal{V}$, where $P_\mathcal{V}$ is the orthogonal projection onto $\mathcal{V}$. Then the problem reduces to finding a subspace $\mathcal{V}$ that maximizes $Trace\{\mathbf{M}^{1/2}\mathbf{WM}^{1/2}P_\mathcal{V}\}$. It is clear that this subspace should belong to $eigspacemax(\mathbf{M}^{1/2}\mathbf{WM}^{1/2};p,q)$. Next we check that F given in (113) is a solution for Γ. Note first that $Spectrum(\mathbf{WM}) = Spectrum(\mathbf{M}^{1/2}\mathbf{WM}^{1/2})$ (see Proposition 7). Thus F corresponds to the largest eigenvalues of $\mathbf{M}^{1/2}\mathbf{WM}^{1/2}$ as well: $\mathbf{WM}F = F\Lambda$, where $Trace\{\Lambda\} = \sum_{i=1}^q \lambda_i$ with λ_i the ordered eigenvalues of $\mathbf{WM}$. This proves (115). Then

$$\mathbf{M}^{1/2}\mathbf{WM}^{1/2}P_F = \mathbf{M}^{1/2}\mathbf{WM}F(F^*\mathbf{M}F)^{-1}F^*\mathbf{M}^{1/2} = P_F\mathbf{M}^{1/2}\mathbf{WM}^{1/2}$$

which shows that $Ran\,P_F$ is an invariant subspace of $M^{1/2}\mathbf{WM}^{1/2}$ and also $Trace\{\mathbf{M}^{1/2}\mathbf{WM}^{1/2}P_F\} = Trace\{\Lambda\}$. Thus $Ran\,P_F \in eigspacemax(\mathbf{M}^{1/2}\mathbf{WM}^{1/2};p,q)$ which proves (113).

Next we note that any other $\Gamma \in L^\infty(\Box; \mathbf{C}^{p\times q})$ that satisfies the s-Riesz basis condition (55) and is optimal should correspond also to an element of $eigspacemax(\mathbf{WM}; \alpha, \beta)$. This means that $Ran\,\Gamma \in eigspacemax(\mathbf{WM}; \alpha, \beta)$. From (55), it follows that $\Gamma = F \cdot L$ for some $L \in L^\infty(\Box; GL_q(\mathbf{C}))$ and $F \in eigmax(\mathbf{WM}; \alpha, \beta)$, i.e. (111).

Finally, the biorthogonal generator is obtained through (66) which turns into (112) when (111) is used. $\Box$.

An upper bound for the optimal error (115) can be easily obtained by using the following inequality:

$$\frac{1}{p-q}\sum_{i=q+1}^{p}\lambda_i(t,s) \le \frac{1}{p}\sum_{i=0}^{p-1}\lambda_i(t,s)$$

Now, note the right hand side is $\frac{1}{p}trace\{W(s)M(t)\}$. Since $\mathbf{W}(s)$ is diagonal and the diagonal of $\mathbf{M}(t)$ is constant equal to $\rho_0(t)$ we obtain further:

$$\sum_{i=q+1}^{p}\lambda_i(t,s) \le \frac{p-q}{p}\rho_0(t)\sum_{r=0}^{p-1}\omega(s+\frac{r}{p})$$

Next, by integrating for $(t,s) \in [0,\frac{1}{q}] \times [0,\frac{1}{p}]$ we get:

$$e_{opt}(\mathbf{WM}; \alpha, \beta) \le (1 - \frac{1}{\alpha\beta})R(0)\|w\|_{L^1} \tag{116}$$

where $\|w\|_{L^1} = \int w(x)dx$ is the 1-norm of the weight function $w(\cdot)$ and $R(0)$ is the autocovariance function evaluated at lag 0.

REMARK 12 *1. In the case λ_q is nondegenerate, the optimal solution is parametrized only by $L^\infty(\Box; GL_q(\mathbf{C}))$. In the case λ_q is degenerate, the parametrization is more complicated because it takes into account the local degeneracy of λ_q.*

2. In either of the two cases (i.e. λ_q degenerate or not), the approximation error is given by the same formula (115).

3. In the case when $\mathbf{W}(s) = \omega(s)I$ or $\mathbf{M}(t) = \rho_0(t)I$ we can solve this problem explicitely. We postpone this analysis until later.

4.2.2 The Two-Channel Optimal Problem

The two-channel transmission scheme uses two encoding and two decoding blocks, one of each for each channel. It thus represents a union of two one-channel transmission schemes, subject to hypotheses

$H(g^1,g^2)$ and $H(g^{1\#},g^{2\#})$. Without these additional constraints, the optimal two-channel problem would simply reduce to the previously solved one-channel optimal problem. Hence a lower bound for the approximation error is given by twice the optimal value of the optimal error for the one-channel case:

$$e^{1+2}_{st,opt} \geq 2e^1_{opt}(\mathbf{WM};\alpha,\beta) = 2\beta q \int_0^{1/p} ds \int_0^{1/q} dt \sum_{j=q+1}^{p} \lambda_j(t,s) \tag{117}$$

The issue is then whether the lower bound can be achieved or not.

The optimal solution for the one-channel problem is parametrized using the spaces in $eigspacemax(\mathbf{WM};p,q)$; in particular we have $Ran\,\Gamma^1, Ran\,\Gamma^2 \in eigspacemax(\mathbf{WM};p,q)$. On the other hand, the frame condition $H(g^1,g^2)$ requires that in the Zak domain, $Ran\,\Gamma^1 + Ran\,\Gamma^2 = \mathbf{C}^p$, another subspace condition. Now clearly these two conditions are contradictory, unless almost all eigenvalues λ_q are degenerate and there is enough "room" in $eigspacemax(\mathbf{WM};p,q)$ to cover $\mathbf{C}^p$; we can therefore in general not hope to achieve equality in (117). Moreover, (117) cannot be improved: for every $\varepsilon > 0$ there is a near-optimal solution with $e^{1+2}_{st} \leq 2\,e_{opt}(\mathbf{WM};\alpha,\beta) + \varepsilon$. All these facts are proven in the following:

THEOREM 13 (Two-Channel) *For the two-channel optimal problem, the lower bound (117) is sharp. The equality cannot be achieved unless the following two conditions hold for almost every* $(t,s) \in \Box$*:*

1. *The eigenvalue* $\lambda_q(t,s)$ *is degenerate for* $\mathbf{WM}$*;*
2. $1 + k_1(t,s) \geq k_2(t,s) = p - q$.

Proof

We have to prove two statements: one is about the near-optimal solution with bound $2e_{opt}(\mathbf{WM};\alpha,\beta)$, the other concerns conditions to be satisfied if the bound is attained.

For the near-optimal solution we use a perturbative argument as follows. Consider $(g^1,g^{1\#})$ the solution of the optimal one-channel problem. We shall tailor a near-optimal solution $(g^2,g^{2\#})$ for the one-channel problem by perturbing the first one in such a way that the two hypotheses $H(g^1,g^2)$ and $H(g^{1\#},g^{2\#})$ are satisfied.

For each (t,s) dim $Ran\,\Gamma^1(t,s) = q$. Let us construct the orthogonal complement of $Ran\,\Gamma^1(t,s)$ and let $f_1,\ldots,f_{p-q}$ be an orthonormal basis in this complement. Note that $f_1,\ldots,f_{p-q}$ can be chosen to be at least measurable, as vector-valued functions over $\Box$, by a similar argument as in Proposition 9 . Since

$1 \leq \frac{p}{q} \leq 2$ we have $p-q \leq q$. Set:

$$\begin{aligned} \Gamma^2 &= \Gamma^1 + \varepsilon\Delta \\ \Gamma^{2\#} &= \Gamma^{1\#} + \varepsilon\Delta \\ \Delta &= [f_1|\cdots|f_{p-q}|0|\cdots|0] \end{aligned} \tag{118}$$

where the $p \times q$ matrix Δ contains on the first $p-q$ columns, the component of the vectors $f_1, \ldots, f_{p-q}$ in the cannonical basis and then is completed with zero on the remaining $2q-p$ columns. The lower bound of the WH multiset $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ is given by (see (59) $A = \frac{1}{p}\inf_{(t,s)}\min_i \lambda_i(\Gamma^1\Gamma^{1^*} + \Gamma^2\Gamma^{2^*})$. By Theorem 11 $\Gamma^1 = F_1 L_1$ with $L_1 \in L^\infty(\Box; GL_q(\mathbf{C}))$. Then $\Gamma^1\Gamma^{1^*} = F_1 L_1 L_1^* F_1^*$. In the orthonormal basis $f_1, \ldots, f_p$ (which depends on t, s), the matrices of interest are given by:

$$F_1 = \begin{bmatrix} I_q \\ 0 \end{bmatrix} \quad , \quad \Delta = \begin{bmatrix} 0 \\ \tilde{I}_{p-q} \end{bmatrix},$$

where every $p \times q$ matrix is written in two blocks: a $q \times q$ block on top, and a $(p-q) \times q$ block on the bottom; $\tilde{I}_{p-q}$ is the $(p-q) \times q$ matrix of which the first $p-q$ columns form the $(p-q) \times (p-q)$ identity matrix and the remaining $2q-p$ columns are zero. Then

$$\Gamma^1\Gamma^{1^*} + \Gamma^2\Gamma^{2^*} = \begin{bmatrix} L_1 \\ 0 \end{bmatrix} \begin{bmatrix} L_1^* & 0 \end{bmatrix} + \begin{bmatrix} L_1 \\ \varepsilon\tilde{I}_{p-q} \end{bmatrix} \begin{bmatrix} L_1^* & \varepsilon\tilde{I}_{p-q}^* \end{bmatrix} = \begin{bmatrix} 2L_1L_1^* & \varepsilon L_1\tilde{I}_{p-q}^* \\ \varepsilon\tilde{I}_{p-q}L_1^* & \varepsilon^2 I_{p-q} \end{bmatrix}.$$

But:

$$\begin{bmatrix} 2L_1L_1^* & \varepsilon\tilde{L}_1 \\ \varepsilon\tilde{L}_1 & \varepsilon^2 I_{p-q} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}L_1L_1^* & 0 \\ 0 & \frac{\varepsilon^2}{3}I_{p-q} \end{bmatrix} + \begin{bmatrix} \sqrt{\frac{3}{2}}L_1 \\ \varepsilon\sqrt{\frac{2}{3}}\tilde{I}_{p-q} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\frac{3}{2}}L_1^* & \varepsilon\sqrt{\frac{2}{3}}\tilde{I}_{p-q}^* \end{bmatrix} \geq \begin{bmatrix} \frac{1}{2}L_1L_1^* & 0 \\ 0 & \frac{\varepsilon^2}{3}I_{p-q} \end{bmatrix}$$

Hence:

$$\lambda_{min}(\Gamma^1\Gamma^{1^*} + \Gamma^2\Gamma^{2^*}) \geq \min(\frac{1}{3}\varepsilon^2, \frac{1}{2}\|L_1^{-1}\|^{-2})$$

Since $L_1 \in L^\infty(\Box; GL_q(\mathbf{C}))$ it follows that $\Gamma^1\Gamma^{1^*} + \Gamma^2\Gamma^{2^*} \geq \gamma > 0$ for a.e. $(t,s) \in \Box$ and some $\gamma > 0$. Thus $H(g^1, g^{1\#})$ is fulfilled. Similarly $H(g^2, g^{2\#})$ holds true as well.

It remains to check that the approximation error is close to $2e_{opt}(\mathbf{WM}; \alpha, \beta)$. The first channel approximation error is $e_{opt}(\mathbf{WM}; \alpha, \beta)$. For the second channel we use (43) and get:

$$e_{st}^2 - e_{st;o}^1 = e(\mathbf{W}, \mathbf{M}; \Gamma^1 + \varepsilon\Delta, \Gamma^{1\#} + \varepsilon\Delta; \alpha, \beta) - e(\mathbf{W}, \mathbf{M}; \Gamma^1, \Gamma^{1\#}; \alpha, \beta)$$

Notice that $\mathbf{W}, \mathbf{M}, \Gamma^1, \Gamma^{1\#}, \Delta$ are all bounded as functions in $L^\infty(\Box; Sym_p^+(\mathbf{C}))$, $L^\infty(\Box; \mathbf{C}^{p\times q})$ or $L^\infty(\Box; \mathbf{C}^{p\times(p-q)})$. Using again (43) and expanding the above formula, we obtain a fourth order polynomial in ε with zero constant term, $e^2_{st} - e^1_{st;o} = a\varepsilon + b\varepsilon^2 + c\varepsilon^3 + d\varepsilon^4$. Since this can obviously be made arbitrarily small by choosing appropriately small ε, (117) is sharp.

For the second part of the theorem we need to show the two conditions in the hypothesis are equivalent with the existence of two one-channel optimal solutions g^1, g^2 that satisfy $H(g^1, g^2)$. As mentioned before this reduces to the algebraic condition $Ran\,\Gamma^1 + Ran\,\Gamma^2 = \mathbf{C}^p$ a.e., or equivalently, to the existence of two members $\nu_1, \nu_2 \in eigspacemax(\mathbf{WM}; \alpha\beta)$ such that $\nu_1(t,s) + \nu_2(t,s) = \mathbf{C}^p$ for a.e. $(t,s) \in \Box$. Using now the parametrization of $eigspacemax(S;\alpha,\beta)$ developed in 4.1 we obtain the conclusion. $\diamond$.

REMARK 14 *The proof of the theorem suggests how to construct optimal solutions when the two conditions hold. Let us fix some $(t_0, s_0) \in \Box$ and let $k_1 = k_1(t_0, s_0)$, $k_2 = k_2(t_0, s_0)$ and D_{k_1,k_2} as in 4.1. Let $\mathcal{U}(t_0, s_0)$ denote the eigenspace of $\mathbf{W}(s_0)\mathbf{M}(t_0)$ corresponding to the eigenvalues $\{\lambda_j(t_0,s_0), \lambda_j(t_0,s_0) > \lambda_q(t_0,s_0)\}$. If $E_q(t_0,s_0)$ denotes the eigenspace corresponding to $\lambda_q(t_0,s_0)$ then $\mathcal{U}(t_0,s_0) \oplus E_q(t_0,s_0) = \mathbf{C}^p$ and $\dim E_q(t_0,s_0) \geq 2(p-q)$. Let $e_1(t,s), \ldots, e_p(t,s)$ be a measurable system of eigenvectors. Let $\pi_1, \pi_2 : \{1,2,\ldots,1+k_1\} \to \{q-k_1, q-k_1+1, \ldots, q, q+1, \ldots, q+k_2 = p\}$ be two injective selection maps such that $Ran\,\pi_1 \cup Ran\,\pi_2 = \{q-k_1, q-k_1+1, \ldots, q, q+1, \ldots, q+k_2\}$. Then on D_{k_1,k_2} we construct the following objects:*

$$
\begin{aligned}
(t,s) \in D_{k_1,k_2} &\mapsto E^1_q(t,s) = span\{e_{\pi_1(1)}(t,s), \ldots, e_{\pi_1(1+k_1)}(t,s)\} \\
(t,s) \in D_{k_1,k_2} &\mapsto E^2_q(t,s) = span\{e_{\pi_2(1)}(t,s), \ldots, e_{\pi_2(1+k_1)}(t,s)\} \\
(t,s) \in D_{k_1,k_2} &\mapsto \nu_1(t,s) = \mathcal{U}(t,s) \oplus E^1_q(t,s) \\
(t,s) \in D_{k_1,k_2} &\mapsto \nu_2(t,s) = \mathcal{U}(t,s) \oplus E^2_q(t,s) \\
(t,s) \in D_{k_1,k_2} &\mapsto \tilde{F}^1(t,s) = [e_1(t,s) \,|\, \cdots \,|\, e_{q-k_1-1} \,|\, e_{\pi_1(1)}(t,s) \,|\, \cdots \,|\, e_{\pi_1(1+k_1)}(t,s)] \\
(t,s) \in D_{k_1,k_2} &\mapsto \tilde{F}^2(t,s) = [e_1(t,s) \,|\, \cdots \,|\, e_{q-k_1-1} \,|\, e_{\pi_2(1)}(t,s) \,|\, \cdots \,|\, e_{\pi_2(1+k_1)}(t,s)] \\
(t,s) \in D_{k_1,k_2} &\mapsto F^1(t,s) = \tilde{F}^1(\tilde{F}^{1*}\tilde{F}^1)^{-1/2} \\
(t,s) \in D_{k_1,k_2} &\mapsto F^2(t,s) = \tilde{F}^2(\tilde{F}^{2*}\tilde{F}^2)^{-1/2}
\end{aligned}
$$

By construction $E^1_q(t,s) + E^2_q(t,s) = E_q(t,s)$, $\dim \nu_1(t,s) = \dim \nu_2(t,s) = q$ and $\nu_1, \nu_2 \in eigspacemax(\mathbf{W}(s)\mathbf{M}(t); \alpha, \beta)|_{D_{k_1,k_2}}$.

The F^1, F^2 constructed on D_{k_1,k_2} are local optimal solutions for each one-channel transmission problem and for every $(t,s) \in D_{k_1,k_2}$ they form a frame in $\mathbf{C}^p$ with lower bound larger than 1. Therefore by patching together these local frames we get two windows g^1, g^2 that satisfy $H(g^1, g^2)$ and are also optimal *for each one-channel transmission problem.*

4.2.3 The Optimal TSDE

The encoding scheme using a *time-shift division encoder* (TSDE) shown in figure 2 is characterized by:

$$\begin{array}{ll} g^1_{mn} = g_{m,2n;\alpha_0,\beta_0} & g^2_{mn} = g_{m,2n+1;\alpha_0,\beta_0} \\ g^{1\#}_{mn} = g^{1\#}_{m,2n;\alpha_0,\beta_0} & g^{2\#}_{mn} = g^{2\#}_{m,2n+1;\alpha_0,\beta_0} \end{array} \tag{119}$$

where $\alpha_0\beta_0 \in [\frac{1}{2}, 1]$ and $g, g^{1\#}, g^{2\#}$ are the generating windows. The approximation errors have been computed in (46) and (47). In the TSDE case $g^1 = g^2 = g$ and thus the total approximation error is:

$$e^{1+2;TSDE}_{st} = e(\mathbf{W}, \mathbf{M}; \Gamma, \Gamma^{1\#}; \alpha_0, 2\beta_0) + e(\mathbf{W}, \mathbf{M}; \Gamma, \Gamma^{2\#}; \alpha_0, 2\beta_0)$$

The partial optimal TSDE with fixed encoder has been obtained in Theorem 5 case B5, as:

$$\Gamma^{1\#}{}_{so} = \Gamma^{2\#}{}_{so} = p\mathbf{M}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1} \tag{120}$$

and

$$e^{1+2;TSDE}_{st;so} = 2e_{so}(\mathbf{W}, \mathbf{M}; \Gamma; \alpha_0, 2\beta_0) \tag{121}$$

which is perfectly equivalent to the one-channel problem, duplicated modulo a β_0-time shift to the two channels. Clearly the optimal value of the error is bounded below by twice the one-channel optimal error:

$$e^{1+2;TSDE}_{st;opt} \geq 2e_{opt}(\mathbf{WM}; \alpha_0, 2\beta_0) \tag{122}$$

As in the optimal two-channel case, the lower bound in actually achieved if and only if the hypothesis $H_2(g)$ holds true. To be more precise, the only issue is whether there is any optimal solution of the one-channel optimal problem (paramterized by (111)-(114)) that makes also $\mathcal{WH}_{g;\alpha_0,\beta_0}$ a WH frame. A partial answer to this question in given by the following lemma:

LEMMA 15 *Consider $2\alpha_0\beta_0 = \frac{p}{q} \geq 1$ with p, q relatively prime integers and $g \in L^2(\mathbf{R})$ such that $\mathcal{WH}_{g;\alpha_0,2\beta_0}$ is a s-Riesz basis. Then $\mathcal{WH}_{g;\alpha_0,\beta_0}$ is a frame for $L^2(\mathbf{R})$ iff there is an $A > 0$ such that for almost every $(t,s) \in \square$:*

$$\Gamma\Gamma^*|_{(t,s)} + \Gamma\Gamma^*|_{(t,s+\frac{1}{2})} \geq pA \tag{123}$$

(Γ is the $p \times q$ matrix defined in (37)).

Proof Let us denote by $g^{(2)} = T_{\beta_0} g$ the β_0-time shift of g. By sorting the labels (m, n) into those with even n and those with odd n, we have the following decomposition

$$\mathcal{WH}_{g;\alpha_0,\beta_0} = \mathcal{WH}_{g;\alpha_0,2\beta_0} \cup \mathcal{WH}_{g^{(2)};\alpha_0,2\beta_0}$$

Therefore $\mathcal{WH}_{g;\alpha_0,\beta_0}$ is a frame iff $\mathcal{WH}_{(g,g^{(2)});\alpha_0,2\beta_0}$ is a WH multiframe. Next note the Zak transform of $g^{(2)}$ is:

$$G^{(2)}(t,s) = \sqrt{2\beta_0} \sum_k e^{2\pi i k t} g^{(2)}(2\beta_0(s+k)) = G(t, s + \frac{1}{2})$$

The fact that $\mathcal{WH}_{g;\alpha_0,2\beta_0}$ is a WH s-Riesz basis translates into (see Proposition 3, case A.1):

$$pA_0 \leq \Gamma^* \Gamma \leq pB_0 \tag{124}$$

for almost every $(t,s) \in \square$ and some $A_0, B_0 > 0$. By the same Proposition 3, case B.3, the multiset $\mathcal{WH}_{(g,g^{(2)});\alpha_0,2\beta_0}$ is a multiframe (with frame bounds A, B) iff

$$pA \leq \Gamma\Gamma^* + \Gamma^{(2)}(\Gamma)^{(2)*} \leq pB \tag{125}$$

a.e. (t,s). Since $\Gamma^{(2)} = \Gamma(t, s+\frac{1}{2})$ the upper bound comes automatically from (124) (note $\|T^*T\| = \|TT^*\|$). Thus the only condition that remains to be satisfied is the lower bound in (125) which is equivalent to (123). $\diamond$

This Lemma does not solve our problem completely yet. It merely states an equivalent form to the hypothesis $H_2(g)$ when $H_1(g)$ is satisfied (i.e. when $\mathcal{WH}_{g;\alpha_0,2\beta_0}$ is a WH s-Riesz basis). However, it provides an easier verifiable condition. The general solution of the optimal TSDE is furnished by the following theorem:

THEOREM 16 (TSDE) *Consider the encoding scheme 2 using a TSDE. Then the optimal approximation error is*

$$e^{TSDE}_{st;opt} = 2e_{opt}(\mathbf{WM}; \alpha_0, 2\beta_0) \tag{126}$$

This bound is achieved iff for almost every $(t,s) \in \square$,

$$FF^*|_{(t,s)} + FF^*|_{(t,s+\frac{1}{2})} \geq A \tag{127}$$

for $F \in eigmax(\mathbf{WM}; \alpha_0, 2\beta_0)$ and some $A > 0$. In this case the optimal solution in terms of the Zak transform is

$$\Gamma_{opt}(t,s) = F(t,s)L(t,s) \tag{128}$$

$$\Gamma^{1\#}{}_{opt} = \Gamma^{2\#}{}_{opt} = F^{\#}(t,s)L^{\#}(t,s) \tag{129}$$

with

$$F^{\#} = \mathbf{M}F(F^*\mathbf{M}^2F)^{-1/2} \in eigmax(\mathbf{MW}; \alpha_0, 2\beta_0) \tag{130}$$

$$L \in L^\infty(\Box; GL_q(\mathbf{C})) \quad , \quad L^{\#} = p(F^*\mathbf{M}^2F)^{1/2}(F^*\mathbf{M}F)^{-1}L^{-*} \in L^\infty(\Box; GL_q(\mathbf{C})) \tag{131}$$

If (127) is not satisfied, the optimal bound (126) is not achieved, however for every $\varepsilon > 0$ there is a near-optimal solution within ε.

REMARK 17 *1. The condition (127) is generically satisfied. In fact it represents a constraint only on the weight w. However, for every (R, w) (autocovariance and weight functions) such that $\mathbf{W}, \mathbf{M} \in L^\infty(\Box; GL_q(\mathbf{C}))$ but not satisfying (127), and for every $\varepsilon > 0$ there is a weight w' that satisfies (127) and $\|w - w'\|_{L^1} < \varepsilon$. To see this note that $\|w - w'\|_{L^1} \leq \|\mathbf{W} - \mathbf{W}'\|_{L^\infty}$ and (127) can be made to hold true with an arbitrary small perturbation.*

2.There is a particular class of weights for which (127) does not hold true in general (it depends now on R). This class contains the characteristic function of $[0, 2\beta_0]$, or any other weight w such that $\omega(s) \equiv const$. For these weights $\mathbf{W}(s) = \omega(0)I$ and therefore F does not depend on s. The only way for (127) to be satisfied is, in this case, that $\mathbf{M}$ satisfy both conditions of the Theorem 13. Since this is an important case we state it explicitelly in the following corollary:

COROLLARY 18 *Suppose the weight w satisfies $\omega(s) = \sum_{k\in\mathbf{Z}} w(2\beta_0(s+k)) = \omega(0) > 0$ for almost every s. Then the lower bound in the TSDE encoding scheme is achieved iff for almost every t the following two conditions hold true:*

1. The eigenvalue $\lambda_q(t)$ of $\mathbf{M}(t)$ is degenerate;

2. $1 + k_1(t) \geq k_2(t) = p - q$

where $k_1(t), k_2(t)$ are the left and right multiplicities of $\lambda_q(t)$ as defined in 4.1.

Proof of Theorem 16

It is clear that Γ_{opt} and $\Gamma^{1\#}{}_{opt}$ from (127) and (129) achieve the lower bound in (122) because of (121) and the construction of $eigmax(\mathbf{WM}; \alpha_0, 2\beta_0)$. The only issue is to check whether the frame hypothesis on $\mathcal{WH}_{g;\alpha_0,\beta_0}$ is satisfied. Using the previous lemma we have to check whether

$$\Gamma_{opt}\Gamma^*_{opt}|_{(t,s)} + \Gamma_{opt}\Gamma^*_{opt}|_{(t,s+\frac{1}{2})} \geq \tilde{A}$$

for some $\tilde{A} > 0$ and a.e. (t, s). Since $\Gamma = F \cdot L$ and $L \in L^\infty(\Box; GL_q(\mathbf{C}))$ it follows that $\|L^{-1}(t, s)\|$ is uniformly bounded, therefore $L \cdot L^*|_{(t,s)} \geq \gamma > 0$ for a.e. (t, s). Hence:

$$\Gamma_{opt}\Gamma^*_{opt}|_{(t,s)} + \Gamma_{opt}\Gamma^*_{opt}|_{(t,s+\frac{1}{2})} \geq \gamma(FF^*|_{(t,s)} + FF^*_{(t,s+\frac{1}{2})})$$

and (127) is then a sufficient condition.

To show now that it is necessary also, we use the upper bound $L \cdot L^*|_{(t,s)} \leq \Upsilon < \infty$ for some finite $\Upsilon > 0$ and a.e. (t, s). Then if $H_2(g)$ holds true we get:

$$\Upsilon(FF^*|_{(t,s)} + FF^*|_{(t,s+\frac{1}{2})}) \geq \Gamma_{opt}\Gamma^*_{opt}|_{(t,s)} + \Gamma_{opt}\Gamma^*_{opt}|_{(t,s+\frac{1}{2})} \geq p\tilde{A}$$

and thus (127).

The near-optimal solution is easily obtained using the following observation. Condition (127) is structurally stable because for every (t, s), $rank(FF^*|_{(t,s)}) = q \geq \frac{p}{2}$. Thus a small perturbation in $L^\infty(\Box; \mathbf{C}^{p\times q})$-sense of F would make (127) hold true. However by the continuity of the approximation error with respect to the window Γ (see the argument used in the proof of Theorem 13) we get a perturbation that increases the approximation error by no more than ε and makes (127) hold true. This concludes the proof. $\diamond$.

Proof of Corollary 18

If $\mathbf{W}(s) = \omega(0)I_p$ then the columns of F can be chosen from the eigenvectors of $\mathbf{M}$ and the lower bound condition reduces to an algebraic range condition: if $\nu \in eigspacemax(\mathbf{WM}; \alpha_0, 2\beta_0)$ then $\nu(t, s) + \nu(t, s + \frac{1}{2}) = \mathbf{C}^p$ and thus the conclusion. $\diamond$

REMARK 19 *In general the range condition from the proof is a necessary but not sufficient condition for the attainability of the optimal bound. The equation (127) is equivalent to this range condition plus a lower bound of some angle between these spaces. The angle should be defined between the orthogonal complements within each range of their intersection.*

4.2.4 The Optimal TSDD

The encoding scheme using a *time-shift division decoder* (TSDD) is shown in figure 3 and involves the following configuration:

$$\begin{array}{ll} g^1_{mn} = g^1_{m,2n;\alpha_0,\beta_0} & g^2_{mn} = g^2_{m,2n+1;\alpha_0,\beta_0} \\ g^{1\#}_{mn} = g^{\#}_{m,2n;\alpha_0,\beta_0} & g^{2\#}_{mn} = g^{\#}_{m,2n+1;\alpha_0,\beta_0} \end{array} \tag{132}$$

where $\alpha_0\beta_0 \in [\frac{1}{2}, 1]$ and $g^1, g^2, g^\#$ are the generating windows. The approximation error is similar to the TSDE case:

$$e^{1+2;TSDD}_{st} = e(\mathbf{W}, \mathbf{M}; \Gamma^1, \Gamma^\#; \alpha_0, 2\beta_0) + e(\mathbf{W}, \mathbf{M}; \Gamma^2, \Gamma^\#; \alpha_0, 2\beta_0)$$

The partial optimal TSDD with fixed decoder has been obtained in Theorem 5, case B8, as:

$$\Gamma^1{}_{so} = \Gamma^2{}_{so} = p\mathbf{W}\Gamma^\#(\Gamma^{\#^*}\mathbf{W}\Gamma^\#)^{-1} \tag{133}$$

$$e^{1+2;TSDD}_{st} = 2e_{so}(\mathbf{M}, \mathbf{W}; \Gamma^\#; \alpha_0, 2\beta_0) \tag{134}$$

Note the similarity to (120),(121) though $\mathbf{M}$ and $\mathbf{W}$ have switched their places. The problem is formally equivalent to two one-channel problems as in the TSDE case. Since $\mathbf{MW}$ has the same spectrum as $\mathbf{WM}$, it follows immediately the lower bound for the optimal error is:

$$e^{1+2;TSDD}_{st;opt} \geq 2e_{opt}(\mathbf{WM}; \alpha_0, 2\beta_0) \tag{135}$$

Clearly the $\Gamma^\#$ which achieves the lower bound in (135) is given by:

$$\Gamma^\#{}_{opt} = F^\# \cdot L^\# \tag{136}$$

with $F^\# \in eigmax(\mathbf{MW}; \alpha_0, 2\beta_0)$, $L^\# \in L^\infty(\Box; GL_q(\mathbf{C}))$. Using (133) we obtain for Γ

$$\Gamma_{opt} = F \cdot L \tag{137}$$

with

$$\begin{array}{rcccl} F & = & \mathbf{W}F^\#(F^{\#*}\mathbf{W}^2F^\#)^{-1/2} & \in & eigmax(\mathbf{WM}; \alpha_0, 2\beta_0) \\ L & = & p(F^{\#*}\mathbf{W}^2F^\#)^{1/2}(F^{\#*}\mathbf{W}F^\#)^{-1}L^{\#-*} & \in & L^\infty(\Box; GL_q(\mathbf{C})) \end{array} \tag{138}$$

The achievability of the lower bound depends now upon the validity of the frame hypothesis on $\mathcal{WH}_{g;\alpha_0,\beta_0}$ as in the TSDE case. Moreover, the equations (136)-(138) are perfectly equivalent to the solution (127)-(131) of the TSDE case. Thus we have obtained the following

THEOREM 20 (TSDD) *The optimal encoding scheme using a TSDD is identical to the TSDE case. The optimal bound $2e_{opt}(\mathbf{WM};\alpha_0,2\beta_0)$ is achieved under the same conditions and by the same solutions as in the TSDE case.*

REMARK 21 *1. The corollary 18 equally applies to the TSDD case.*

2. One can ask whether the s-Riesz basis and/or frame conditions on the decoder automatically imply the corresponding conditions on the encoder. The answer to this question is analyzed in the next subsubsection, devoted to the TSDED case.

4.2.5 The Optimal TSDED

In the TSDED scheme shown in figure 4, we start with the structure

$$\begin{array}{ll} g^1_{mn} = g_{m,2n;\alpha_0,\beta_0} & g^2_{mn} = g_{m,2n+1;\alpha_0,\beta_0} \\ g^{1\#}_{mn} = g^{\#}_{m,2n;\alpha_0,\beta_0} & g^{2\#}_{mn} = g^{\#}_{m,2n+1;\alpha_0,\beta_0} \end{array} \tag{139}$$

with $\alpha_0\beta_0 \in [\frac{1}{2},1]$ and $g, g^{\#} \in L^2(\mathbf{R})$ the generating windows. Since the partial optimal of TSDE or TSDD schemes involve a TSDED structure anyway, it is straightforward that the optimal TSDED case should be identical to the optimal TSDE and TSDD. Indeed using the same arguments as before, one can easily show that:

$$e^{1+2;TSDED}_{st,opt} \geq 2e_{opt}(\mathbf{WM};\alpha_0,2\beta_0) \tag{140}$$

and the optimal value, if achieved, is attained by (136)-(138). Hence the following theorem:

THEOREM 22 (TSDED) *The optimal TSDED scheme coincides with the optimal TSDE and TSDD schemes. The achievability conditions are the same as in Theorem 16.*

We discuss now two interesting results concerning the behaviour of the encoding and decoding sets. Recall the hypotheses of the optimal problems were stated in terms of the encoding sets. Here we establish the connections with the decoding sets.

PROPOSITION 23 *a) Suppose $R \in L^\infty(\Box; GL_p(\mathbf{C}) \cap Sym^+_p(\mathbf{C}))$ and $g^{\#}$ satisfies $H_1(g^{\#};\alpha,\beta)$, i.e. $\mathcal{WH}_{g^{\#};\alpha,\beta}$ is a s-Riesz basis. Then g defined by*

$$\Gamma = pR\Gamma^{\#}({\Gamma^{\#}}^{*}R\Gamma^{\#})^{-1} \tag{141}$$

satisfies $H_1(g;\alpha,\beta)$ as well, i.e. $\mathcal{WH}_{g;\alpha,\beta}$ is a s-Riesz basis too.

b) Suppose $\mathbf{W}, \mathbf{M} \in L^\infty(\Box; GL_p(\mathbf{C}) \cap Sym_p^+(\mathbf{C}))$ are the standard matrices associated to the weight, respectively the autocovariance function. Suppose $g^\#$ satisfies $H_2(g^\#; \alpha_0, \beta_0)$ and has the form

$$\Gamma^\# = F^\# \cdot L^\# \tag{142}$$

for some $F^\# \in eigmax(\mathbf{MW}; \alpha_0, 2\beta_0)$, $L^\# \in L^\infty(\Box; GL_q(\mathbf{C}))$. Then $g^\#$ satisfies also $H_1(g^\#; \alpha_0, 2\beta_0)$ and g defined by:

$$\Gamma = p\mathbf{W}\Gamma^\#({\Gamma^\#}^*\mathbf{W}\Gamma^\#)^{-1} \tag{143}$$

satisfies both $H_1(g; \alpha_0, 2\beta_0)$ and $H_2(g; \alpha_0, \beta_0)$.

REMARK 24 *One may ask whether is was necessary to assume the special form for $\Gamma^\#$ at part b. The answer is affirmative. Indeed, without assuming $\Gamma^\# = F^\# \cdot L^\#$ as above, the frame conclusion $H_2(g; \alpha_0, \beta_0)$ would not be true in general. Therefore the exact solutions of the partial optimal problems should take into account this phenomenon: while the s-Riesz basis condition on the encoder follows easily from the s-Riesz basis conditions on the decoder, the same thing does not happen for the frame condition.*

Proof of Proposition 23

a) The conclusion follows easily since $H_1(g^\#)$ means:

$$pA_1 \leq {\Gamma^\#}^*\Gamma^\# \leq pB_1 \tag{144}$$

for some $A_1, B_1 > 0$, and $R \in L^\infty(\Box; GL_p(\mathbf{C}))$ means $c_1 \leq R \leq c_2$ a.e. for some $c_1, c_2 > 0$. Hence $pA_1c_1 \leq {\Gamma^\#}^* R \Gamma^\# \leq pB_1c_2$ and

$$\frac{pA_1c_1^2}{B_1^2c_2^2} \leq {\Gamma^\#}^*\Gamma^\# = p^2({\Gamma^\#}^*R\Gamma^\#)^{-1}{\Gamma^\#}^*R^2\Gamma^\#({\Gamma^\#}^*R\Gamma^\#)^{-1} \leq \frac{pB_1c_2^2}{A_1^2c_1^2}$$

and thus $H_1(g)$.

b) The first claim $H_1(g^\#; \alpha_0, 2\beta_0)$ comes from ${\Gamma^\#}^*\Gamma^\# = L^{\#*}L^\# \geq \|{L^\#}^{-1}\|^{-2} > 0$. The conclusion $H_1(g; \alpha_0, 2\beta_0)$ follows from the part a). For the frame condition we use the Theorem 16. Repeating the proof of this theorem we obtain that the only condition we have to check on Γ is the lower bound of the form:

$$K(t,s) = \Gamma \cdot \Gamma^*|_{(t,s)} + \Gamma \cdot \Gamma^*|_{(t,s+\frac{1}{2})} \geq pA$$

for some $A > 0$. We know that $\Gamma^\#$ satisfies (144) and the similar inequality:

$$K^\#(t,s) = \Gamma^\# \cdot {\Gamma^\#}^*|_{(t,s)} + \Gamma^\# \cdot {\Gamma^\#}^*|_{(t,s+\frac{1}{2})} \geq pA_2$$

Using (143) and (142) we get:

$$\begin{aligned} K(t,s) &= p^2[\mathbf{W}\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-2}\Gamma^{\#^*}\mathbf{W}|_{(t,s)} + \mathbf{W}\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-2}\Gamma^{\#^*}\mathbf{W}|_{(t,s+\frac{1}{2})}] \\ &\geq c_3[\mathbf{W}F^{\#}\cdot F^{\#*}\mathbf{W}|_{(t,s)} + \mathbf{W}F^{\#}\cdot F^{\#*}\mathbf{W}|_{(t,s+\frac{1}{2})}] \end{aligned}$$

Now we note that $\mathbf{MW}F^{\#} = F^{\#}\Lambda$ with $\Lambda \in L^{\infty}(\square; GL_q(\mathbf{C}))$. Therefore:

$$\begin{aligned} \mathbf{M}(t)\cdot K(t,s)\cdot\mathbf{M}(t) &\geq c_3[F^{\#}\Lambda^2F^{\#*}|_{(t,s)} + F^{\#}\Lambda^2F^{\#*}|_{(t,s+\frac{1}{2})}] \\ &\geq c_4K^{\#}(t,s) \geq pc_4A_2 \end{aligned}$$

for some positive constants $c_3, c_4 > 0$ depending on $\Gamma^{\#}, \mathbf{W}$ and $\mathbf{M}$. This proves $H_2(g;\alpha_0,\beta_0)$ and hence the Proposition. $\diamond$

4.2.6 The Optimal FSDE

The encoding-decoding scheme for the *frequency-shift division encoder* (FSDE) case is similar to the TSDE case shown in Figure 2. The analog encoder and respectively side decoders are given by:

$$\begin{array}{ll} g^1_{mn} = g_{2m,n;\alpha_0,\beta_0} & g^2_{mn} = g_{2m+1,n;\alpha_0,\beta_0} \\ g^{1\#}_{mn} = g^{1\#}_{2m,n;\alpha_0,\beta_0} & g^{2\#}_{mn} = g^{2\#}_{2m+1,n;\alpha_0,\beta_0} \end{array} \tag{145}$$

with $\alpha_0\beta_0 \in [\frac{1}{2}, 1]$ as before, and $g, g^{1\#}, g^{2\#}$ the generating windows. The approximation error obtained before in (48) and (50) gives:

$$e^{1+2;FSDE}_{st} = e(\mathbf{W},\mathbf{M};\Gamma,\Gamma^{1\#};2\alpha_0,\beta_0) + e(\mathbf{W},\mathbf{N};\Gamma,\Gamma^{2\#};2\alpha_0,\beta_0) \tag{146}$$

The partial optimal FSDE with fixed encoder has been obtained in Theorem 5, case B9, as:

$$\Gamma^{1\#}_{so} = p\mathbf{M}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1} \quad , \quad \Gamma^{2\#} = p\mathbf{N}\Gamma(\Gamma^*\mathbf{N}\Gamma)^{-1} \tag{147}$$

and the error:

$$e^{1+2;FSDE}_{st;so} = e_{so}(\mathbf{W},\mathbf{M};\Gamma;2\alpha_0,\beta_0) + e_{so}(\mathbf{W},\mathbf{N};\Gamma;2\alpha_0,\beta_0) \tag{148}$$

Note that unlike the TSDE case, the two terms are different because $\mathbf{N} \neq \mathbf{M}$ in general. The difference is due to the meaning of stationarity: the signals are assumed stationary in time domain; this makes the TSDE case so easy: both channels have the same $\mathbf{M}$. However, in frequency domain we do not have stationarity which results in different $\mathbf{M}$ matrices.Unfortunately we are not able to obtain a closed form solution for the optimal window in the FSDE case. We still can find lower and upper bounds for the optimal error and a subspace type condition for the optimal window g, though this subspace does not necessary comes from an eigenvalue problem.

For the optimal error we proceed as follows. A lower bound is given as in the previous cases by the two-channel optimal formula with adapted parameters. Thus:

$$e_{opt}^{1+2;FSDE} \geq 2e_{opt}(\mathbf{WM}; 2\alpha_0, \beta_0) \tag{149}$$

An upper bound of the optimal error is obtained by choosing a particular, yet interesting as we shall see later, configuration. Set $\Gamma^{1\#} = \Gamma^{2\#} = \Gamma$ in (146) and obtain:

$$e_{st}^{1+2;FSDE} = e(\mathbf{W}, \mathbf{M} + \mathbf{N}; 2\alpha_0, \beta_0)$$

Thus the optimal value under this constraint is given by a formula similar to the one-channel case. Obviously this will contain an upper bound for the optimal FSDE error:

$$e_{opt}^{1+2;FSDE} \leq e_{opt}(\mathbf{W}(\mathbf{M} + \mathbf{N}); 2\alpha_0, \beta_0) \tag{150}$$

Hence we can bound the optimal FSDE error by:

$$2e_{opt}(\mathbf{WM}; 2\alpha_0, \beta_0) \leq e_{opt}^{1+2;FSDE} \leq e_{opt}(\mathbf{W}(\mathbf{M} + \mathbf{N}); 2\alpha_0, \beta_0) \tag{151}$$

A direct computation shows that $\mathbf{M}$ and $\mathbf{N}$ commute (we shall discuss in §4.3 the exact structure of the spectrum of these two matrices). However they do not commute with $\mathbf{W}$ in general, unless $\mathbf{W}$ (or $\mathbf{M}, \mathbf{N}$) is (are) a multiple of identity. Suppose this is the case, namely $\mathbf{W}$ is a multiple of identity. Then the lower bound is given by the smallest $p - q$ eigenvalues of $\mathbf{M}$, whereas the upper bound is determined by the smallest $p - q$ eigenvalues of $\mathbf{M} + \mathbf{N}$. Unfortunately the $p - q$ eigenvalues of $\mathbf{M}$ that enter in the smallest $p - q$ eigenvalues of $\mathbf{M} + \mathbf{N}$ are, in general, not the $p - q$ smallest eigenvalues of $\mathbf{M}$. Thus, despite of the fact that $e_{opt}(\mathbf{WN}; 2\alpha_0, \beta_0) = e_{opt}(\mathbf{WM}; 2\alpha_0, \beta_0)$ (since $\mathbf{WN}(t) \sim \mathbf{WM}(t + \alpha_0\beta_0)$) we do not obtain an exact formula for the optimal FSDE error.

Let us analyze now the optimal windows. Using (65), (148) turns into the following explicit formula:

$$e_{st;so}^{1+2;FSDE} = \beta_0 q \int_0^{1/p} ds \int_0^{1/q} dt\, trace\{\mathbf{W}(\mathbf{M} + \mathbf{N}) - \mathbf{WM}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1}\Gamma^*\mathbf{M} - \mathbf{WN}\Gamma(\Gamma^*\mathbf{N}\Gamma)^{-1}\Gamma^*\mathbf{N}\}$$

Therefore the optimization problem turns into an infinite number of finite-dimensional optimization problems: for each (t, s), we have to maximize the trace of:

$$I(\Gamma) = trace\{\mathbf{WM}\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1}\Gamma^*\mathbf{M} + \mathbf{WN}\Gamma(\Gamma^*\mathbf{N}\Gamma)^{-1}\Gamma^*\mathbf{N}\} \tag{152}$$

over $\Gamma \in \mathbf{C}^{p\times q}$ subject to the constraints $H_1(g;2\alpha_0,\beta_0)$ and $H_2(g;\alpha_0,\beta_0)$. An analysis of (152) shows that Γ can be factorized in the usual way, $\Gamma = F\cdot L$ where F has orthonormal columns and L is invertible in L^∞. The following Lemma is the first step toward this fact:

LEMMA 25 *Let $T \in GL_q(\mathbf{C})$ be an arbitrary change of coordonates in $\mathbf{C}^q$. Then: $I(\Gamma) = I(\Gamma T)$ and I defines a unique map from $\mathcal{L}_q(\mathbf{C}^p)$, the space of all q-dimensional subspaces of $\mathbf{C}^p$, into $\mathbf{R}^+$ denoted by J as follows:*

$$\nu \in \mathcal{L}_q(\mathbf{C}^p) \longrightarrow J(\nu) := I(\Gamma) \quad , \text{ where } Ran\,\Gamma = \nu \tag{153}$$

REMARK 26 *J can also be defined on the space of rank q orthogonal projectors, since each $\nu \in \mathcal{L}_q(\mathbf{C}^p)$ defines uniquely such a projector.*

Proof The proof is straightforward since $\Gamma(\Gamma^*\mathbf{M}\Gamma)^{-1}\Gamma^*$ and $\Gamma(\Gamma^*\mathbf{N}\Gamma)^{-1}\Gamma^*$ are invariant under the transformation $\Gamma \mapsto \Gamma\cdot T$. Hence, for any two $\Gamma^1,\Gamma^2 \in \mathbf{C}^{p\times q}$ of rank q such that $Ran\,\Gamma^1 = Ran\,\Gamma^2$, $I(\Gamma^1) = I(\Gamma^2)$. $\diamond$

This Lemma shows the optimal window is given by a $\Gamma \in \mathbf{C}^{p\times q}$, $Ran\,\Gamma \subset \mathcal{L}_q(\mathbf{C}^p)$ solution of the following optimization problem:

$$Ran\,\Gamma = arg \max_{\nu\in\mathcal{L}_q(\mathbf{C}^p)} J(\nu) \tag{154}$$

Note that the optimizer may not be unique. However it always can be factorized as $\Gamma = F\cdot L$ with F a $p\times q$-matrix valued function whose columns are orthonormal vectors and L a $q\times q$ matrix valued function. The s-Riesz basis condition $H_1(g;2\alpha_0,\beta_0)$ imposes to consider only those L that are in $L^\infty(\Box;GL_q(\mathbf{C}))$. There still remained to check the frame hypothesis $H_2(g;\alpha_0,\beta_0)$. For this end we need the following Lemma, similar to Lemma 15:

LEMMA 27 *Suppose $2\alpha_0\beta_0 = \frac{p}{q} \geq 1$ with p,q relatively prime integers and $\mathcal{WH}_{g;2\alpha_0,\beta_0}$ a WH s-Riesz basis in $L^2(\mathbf{R})$. Then $\mathcal{WH}_{g;\alpha_0,\beta_0}$ is a frame for $L^2(\mathbf{R})$ iff there is a constant $A>0$ such that for almost every $(t,s)\in\Box$:*

$$\Gamma\Gamma^*|_{(t,s)} + D\Gamma\Gamma^*|_{(t+\alpha_0\beta_0,s)}D \geq pA \tag{155}$$

where D is the $p\times p$ diagonal matrix whose (l,l) element is $D_{l,l} = (-1)^l$ and Γ is the $p\times q$ matrix given by (37).

Proof

Let $g^{(2)} = M_{2\pi\alpha_0} g$ be the $2\pi\alpha_0$-frequency modulation of g. By sorting the labels (m,n) into those with even m and those with odd m, we have the following decomposition

$$\mathcal{WH}_{g;\alpha_0,\beta_0} = \mathcal{WH}_{g;2\alpha_0,\beta_0} \cup \mathcal{WH}_{g^{(2)};2\alpha_0,\beta_0}$$

Therefore $\mathcal{WH}_{g;\alpha_0,\beta_0}$ is a frame iff $\mathcal{WH}_{(g;g^{(2)});2\alpha_0,\beta_0}$ is a WH multiframe. The Zak transform of $g^{(2)}$ is:

$$G^{(2)}(t,s) = \sqrt{\beta_0}\sum_k e^{2\pi ikt} g^{(2)}(\beta_0(s+k)) = e^{2\pi i\alpha_0\beta_0 s} G(t+\alpha_0\beta_0, s)$$

Then the Γ-matrix of $g^{(2)}$ is $\Gamma^{(2)} = e^{2\pi i\alpha_0\beta_0 s} D\Gamma(t+\alpha_0\beta_0, s)$. Since $\mathcal{WH}_{g;2\alpha_0,\beta_0}$ is a WH s-Riesz basis, with Proposition 3, case A.1, we obtain $pA_0 \leq \Gamma^*\Gamma \leq pB_0$ for some $A_0, B_0 > 0$. Using again Proposition 3, case B, $\mathcal{WH}_{(g;g^{(2)});2\alpha_0,\beta_0}$ is a WH multiframe (with frame bounds A, B) iff $pA \leq \Gamma\Gamma^* + \Gamma^{(2)}\Gamma^{(2)*} \leq pB$ for almost every $(t,s) \in \square$ and for some $A, B > 0$. The upper bound is immediate, whereas the lower bound condition is equivalent to (155). This ends the proof of the Lemma. $\diamond$

Thus we have to check (155) for the optimizer of (154). Unfortunately the answer is negative. In general, the optimizers of (154) do not satisfy (155) and the reason is the following: The definition of $\mathbf{N}$ given by (51)-(52) is equivalent to

$$\mathbf{N}(t) = D\mathbf{M}(t+\alpha_0\beta_0)D \tag{156}$$

with D as in the statement of Lemma 27. Then one can easily check that:

$$J(\nu(t+\alpha_0\beta_0, s))|_{(t+\alpha_0\beta_0,s)} = J(D\nu(t+\alpha_0\beta_0,s))|_{(t,s)}$$

Assuming the solution of (154) is unique (in terms of subspaces) and $q < p$ we obtain $DF(t+\alpha_0\beta_0, s) = F(t,s)$ where $\Gamma = F \cdot L$. Then $\Gamma\Gamma^*|_{(t,s)} + D\Gamma\Gamma^*|_{(t+\alpha_0\beta_0,s)}D = F(t,s)(LL^*|_{(t,s)} + LL^*|_{(t+\alpha_0\beta_0,s)})F(t,s)$ which makes (155) impossible. The only way in which (155) can be satisfied is by the existence of two subspace optimizers that cover the entire space $\mathbf{C}^p$ very much like in Theorem 13. In the generic case (when the subspace optimizer is unique) the optimal error is not achieved by admissible windows (i.e. windows that obey the frame and s-Riesz basis hypotheses) but for every $\varepsilon > 0$ there is a near-optimal solution within ε, because of the continuity of the approximation error with respect to the windows. Hence we obtained the following result:

THEOREM 28 (FSDE) *suppose $(t,s) \mapsto F_{opt}(t,s)$ is a measurable solution of (154) whose columns are orthonormal vectors in $\mathbf{C}^p$. Then the optimal error has the lower bound:*

$$e^{1+2;FSDE}_{st;opt} = 2\|w\|_1 R(0) - \beta_0 q \int_0^{1/p} ds \int_0^{1/q} dt\, I(F_{opt}) \tag{157}$$

The optimum is not achieved unless there are two such solutions F_1 and F_2 such that $Ran\, F_1 + Ran\, F_2 = \mathbf{C}^p$ a.e. $(t,s) \in \Box$. In this case the admissible optimizers are parametrized by $\Gamma_{opt} = F_{opt} L$ with $L \in L^\infty(\Box; GL_q(\mathbf{C}))$. The decoding windows are obtained from (147).

In general, for every $\varepsilon > 0$ there is an admissible near-optimum solution g, within ε of (157).

REMARK 29 *The term $2\|w\|_1 R(0)$ in (157) comes from the integral $\beta_0 q\, trace\{\mathbf{W}(\mathbf{M}+\mathbf{N})\}$ over $\Box$. $\|w\|_1 = \int w(x)dx$ stands for the 1-norm of the weight function and $R(0)$ is the variance of the signal (i.e. the autocovariance for lag 0).*

4.2.7 The Optimal FSDD

In the *frequency-shift division decoder* case, the encoding-decoding scheme is similar to the TSDD case shown in Figure 3. The analog encoders and side decoders are described by:

$$\begin{array}{ll} g^1_{mn} = g^1_{2m,n;\alpha_0,\beta_0} & g^2_{mn} = g^2_{2m+1,n;\alpha_0,\beta_0} \\ g^{1\#}_{mn} = g^{\#}_{2m,n;\alpha_0,\beta_0} & g^{2\#}_{mn} = g^{\#}_{2m+1,n;\alpha_0,\beta_0} \end{array} \tag{158}$$

with $\alpha_0\beta_0 \in [\frac{1}{2}, 1]$ and $g^1, g^2, g^\#$ the generating windows. The approximation error obtained through (48) and (50) gives:

$$e^{1+2;FSDD}_{st} = e(\mathbf{W}, \mathbf{M}; \Gamma^1, \Gamma^\#; 2\alpha_0, \beta_0) + e(\mathbf{W}, \mathbf{N}; \Gamma^2, \Gamma^\#; 2\alpha_0, \beta_0) \tag{159}$$

The solution of the partial optimal FSDD with fixed decoder, as found in Theorem 5, case B12, yields:

$$\Gamma^1{}_{so} = \Gamma^2{}_{so} = p\mathbf{W}\Gamma^\#(\Gamma^{\#^*}\mathbf{W}\Gamma^\#)^{-1} \tag{160}$$

and the error:

$$e^{1+2;FSDD\#}_{st;so} = e_{so}(\mathbf{M}+\mathbf{N}, \mathbf{W}; \Gamma^\#; 2\alpha_0, \beta_0) \tag{161}$$

The optimal FSDD sould minimize (161). This is equivalent to the one-channel optimization problem having the matrix-valued autocovariance function $\mathbf{M}+\mathbf{N}$ instead of $\mathbf{M}$. Therefore the lower bound of the optimal approximation error is

$$e^{1+2;FSDD}_{st,opt} \geq e_{opt}((\mathbf{M}+\mathbf{N})\mathbf{W}; 2\alpha_0, \beta_0) \tag{162}$$

and the optimizer is parametrized by $\Gamma^{\#} = F^{\#}L^{\#}$, $F^{\#} \in eigmax((\mathbf{M}+\mathbf{N})\mathbf{W}; 2\alpha_0, \beta_0)$ and $L^{\#} \in L^{\infty}(\Box; GL_q(\mathbf{C}))$ with $2\alpha_0\beta_0 = \frac{p}{q}$. However it remains to check that $g = g^1 = g^2$ satisfies the s-Riesz basis and frame hypotheses $H_1(g; 2\alpha_0, \beta_0)$ and $H_2(g; \alpha_0, \beta_0)$. First we need the following result proved with the help of Lemma 27:

PROPOSITION 30 *Suppose* $\mathbf{W}, \mathbf{M}+\mathbf{N} \in L^{\infty}(\Box; GL_p(\mathbf{C}) \cap Sym_p^+(\mathbf{C}))$ *and* $g^{\#}$ *defined by* $\Gamma^{\#} = F^{\#}L^{\#}$ *for some* $F^{\#} \in eigmax((\mathbf{M}+\mathbf{N})\mathbf{W}; 2\alpha_0, \beta_0)$ *and* $L^{\#} \in L^{\infty}(\Box; GL_q(\mathbf{C}))$*, satisfies* $H_2(g; \alpha_0, \beta_0)$*. Then* $g^{\#}$ *satisfies* $H_1(g^{\#}; 2\alpha_0, \beta_0)$ *and* g *defined by* $\Gamma = p\mathbf{W}\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-1}$ *satisfies both* $H_1(g; 2\alpha_0, \beta_0)$ *and* $H_2(g; \alpha_0, \beta_0)$*.*

Proof

The claim $H_1(g^{\#}; 2\alpha_0, \beta_0)$ comes from $\Gamma^{\#^*}\Gamma^{\#} = L^{\#*}L^{\#} \geq \|L^{*-1}\|^{-2} \geq 0$. The second claim $H_1(g; 2\alpha_0, \beta_0)$ follows from Proposition 23, part a). The frame condition $H_2(g; 2\alpha_0, \beta_0)$ comes from the previous lemma as follows: First, the condition we need to check is (155). Let:

$$K(t,s) = \Gamma\Gamma^*|_{(t,s)} + D\Gamma\Gamma^*|_{(t+\alpha_0\beta_0,s)}D$$

and similarly:

$$K^{\#}(t,s) = \Gamma^{\#}\Gamma^{\#^*}|_{(t,s)} + D\Gamma^{\#}\Gamma^{\#^*}|_{(t+\alpha_0\beta_0,s)}D$$

By hypothesis and Lemma 27, $K^{\#}(t,s) \geq pA$ for some $A > 0$. For $M \in L^{\infty}(\Box; GL_p(\mathbf{C}) \cap Sym_p^+(\mathbf{C}))$ we have $XMX^* \geq \|M^{-1}\|^{-1}XX^*$. Therefore:

$$\begin{aligned} K(t,s) = p^2\mathbf{W}[\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-2}\Gamma^{\#^*}|_{(t,s)} + D\Gamma^{\#}(\Gamma^{\#^*}\mathbf{W}\Gamma^{\#})^{-2}\Gamma^{\#^*}D|_{(t+\alpha_0\beta_0,s)}]\mathbf{W} \\ \geq p^2\|\Gamma^{\#^*}\mathbf{W}\Gamma^{\#}\|^{-2}\|\mathbf{W}^{-1}\|^{-2}K^{\#}(t,s) \geq c_1 A \end{aligned}$$

for some positive constant c_1. This end the proof of the claim and hence of the Proposition. $\diamond$

This Proposition proves the optimizer $g, g^{\#}$ defined above satisfies the s-Riesz basis hypotheses. Unfortunately in the cases of interest, the frame conditions is not satisfied. Indeed, suppose $\Gamma^{\#}$ is given by $\Gamma^{\#} = F^{\#}L^{\#}$ with $F^{\#} \in eigmax((\mathbf{M}+\mathbf{N})\mathbf{W}; 2\alpha_0, \beta_0)$ and $L^{\#} \in L^{\infty}(\Box; GL_q(\mathbf{C}))$. The definition of $\mathbf{N}$ given by (51)-(52) is equivalent to $\mathbf{N}(t) = D\mathbf{M}(t+\alpha_0\beta_0)D$ with D as in the statement of Lemma 27. Then note the following invariance $(\mathbf{M}+\mathbf{N})\mathbf{W}|_{(t+\alpha_0\beta_0,s)} = D(\mathbf{M}+\mathbf{N})\mathbf{W}D|_{(t,s)}$ holds which implies $F^{\#}(t+\alpha_0\beta_0, s) = DF^{\#}(t,s)$, unless the q^{th} eigenvalue λ_q of $(\mathbf{M}+\mathbf{N})\mathbf{W}$ is fully degenerate as in the hypothesis of Theorem 13. This invariance turns into:

$$\Gamma^{\#}\Gamma^{\#^*}|_{(t,s)} + D\Gamma^{\#}\Gamma^{\#^*}|_{(t+\alpha_0\beta_0,s)}D = F^{\#}(t,s)[L^{\#}L^{\#*}|_{(t,s)} + L^{\#}L^{\#*}|_{(t+\alpha_0\beta_0,s)}]F^{\#}(t,s)$$

and clearly the rank of this matrix is $q < p$, when $2\alpha_0\beta_0 > 1$. On the other hand, a similar perturbation argument as in Theorem 13, shows that the approximation error can be arbitray close to the lower bound in (162). Thus we obtained:

THEOREM 31 (FSDD) *In the FSDD case, the optimal approximation error is:*

$$e^{TSDD}_{st;opt} = e_{opt}((\mathbf{M}+\mathbf{N})\mathbf{W}; 2\alpha_0, \beta_0) \tag{163}$$

although in general it is not achieved by any encoder-decoder satisfying the frame hypothesis $H_2(g;\alpha_0,\beta_0)$. For any $\varepsilon > 0$ there are $g^1, g^2, g^{\#}$ that satisfy the s-Riesz basis and frame hypotheses $H_1(g^1; 2\alpha_0, \beta_0)$, $H_1(g^2; 2\alpha_0, \beta_0)$, $H(g^1, g^2; 2\alpha_0, \beta_0)$ and achieve for FSDD scheme an approximation error within ε of the optimal value (163). The optimal value is achieved by an admissible solution (i.e. one that satisfies the above s-Riesz basis and frame hypotheses) only if the conditions in Theorem 13 with $\mathbf{WM}$ replaced by $\mathbf{W}(\mathbf{M}+\mathbf{N})$ are satisfied in which case the optimizers are parametrized by $\Gamma^{\#} = F^{\#}L^{\#}$ with $F^{\#} \in eigmax((\mathbf{M}+\mathbf{N})\mathbf{W}; 2\alpha_0, \beta_0)$ and $L^{\#} \in L^{\infty}(\square; GL_q(\mathbf{C}))$.

4.2.8 The Optimal FSDED

The *frequency-shift division encoder and decoder* case, similar to TSDED shown in Figure 4, has the following equations:

$$\begin{array}{ll} g^1_{mn} = g_{2m,n;\alpha_0,\beta_0} & g^2_{mn} = g_{2m+1,n;\alpha_0,\beta_0} \\ g^{1\#}_{mn} = g^{\#}_{2m,n;\alpha_0,\beta_0} & g^{2\#}_{mn} = g^{\#}_{2m+1,n;\alpha_0,\beta_0} \end{array} \tag{164}$$

leaving only two degrees of freedom g and $g^{\#}$. As in the TSDED case, the FSDED optimal problem reduces to the previous case FSDD. Indeed, as we have seen in (160), the partial optimal FSDD with fixed decoder already requires $g^1 = g^2$. Thus the optimizers of FSDD are also optimizers for FSDED and conversly. Thus we obtained:

THEOREM 32 (FSDED) *The optimal approximation error in the FSDED case is*

$$e^{FSDED}_{st;opt} = e_{opt}((\mathbf{M}+\mathbf{N})\mathbf{W}; 2\alpha_0, \beta_0) \tag{165}$$

and is achieved under the same conditions as in Theorem 31. Moreover for every $\varepsilon > 0$ there is a near-optimal admissible solution $g, g^{\#}$, within ε to (165).

REMARK 33 *The only new thing this Theorem brings, compared to Theorem 31, is that there is near-optimal solutions of the form* $g^1 = g^2 = g$ *for every* $\varepsilon > 0$. *Again the proof of this fact follows the perturbative arguments as shown in Theorem 13.*

4.3 The Case $\mathbf{W}(s) = \omega(s)I$

In this subsection we analyze the optimal solutions obtained before in the case $\mathbf{W}(s)$ is a multiple of the identity matrix. Recall the operator $\mathbf{W}(s)$ is defined in terms of the weight function $w(\cdot)$ by (39) and (42). Thus $\mathbf{W}(s)$ is a multiple of identity if and only if the function $\omega(s)$ is a $\frac{1}{p}$-periodic function. This means the following condition:

$$\sum_{k\in\mathbf{Z}} w(\beta(s+\frac{1}{p}+k)) = \sum_{k\in\mathbf{Z}} w(\beta(s+k)) \tag{166}$$

for almost every s. A particular case is when $w = 1_{[0,\beta]}$ and then the approximation errors (16)-(17) are computed as averages over an interval of length β, the translation step.

Except for the FSDE case, in all the other configurations, the optimal solution involves the computation of the eigenspaces of $\mathbf{WM}$ or $\mathbf{W}(\mathbf{M}+\mathbf{N})$. When $\mathbf{W}$ is a multiple of the identity, the computation reduces to the eigenproblem for $\mathbf{M}$ or $\mathbf{M}+\mathbf{N}$. In the rest of this subsection we find first the eigenvalues and eigenvectors of $\mathbf{M}$ and $\mathbf{M}+\mathbf{N}$, and next we compute the optimal errors for two encoding-decoding schemes (one-channel and FSDD cases).

Recall $\mathbf{M}$ is the $Sym_p(\mathbf{C})$-valued function introduced by (40). We pointed out in (41) some of its properties. In fact $\mathbf{M}(t)$ is not only self-adjoint Toeplitz matrix, but it is also a nonnegative form as we prove below. First note the following "quasi-periodicity" property of $\mathbf{M}(t)$:

$$E(t)\cdot\mathbf{M}(t)\cdot E(t)^* = \mathbf{M}(t) \tag{167}$$

where $E(t)$ is the following $p \times p$ unitary matrix:

$$E(t) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ e^{-2\pi i q t} & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{168}$$

Therefore E and $\mathbf{M}$ commutes, hence they have the same system of eigenvectors. A simple computation shows the eigenvalues of $E(t)$ are the p roots of $e^{-2\pi i q t}$:

$$\varepsilon_r(t) = e^{-2\pi i \frac{q}{p}(t+\frac{r}{q})} \quad , \quad r = 0, 1, \ldots, p-1 \tag{169}$$

corresponding to the eigenvectors:

$$x_r = \frac{1}{\sqrt{p}}[1 \quad \varepsilon_r \quad \varepsilon_r^2 \quad \cdots \quad \varepsilon_r^{p-1}]^T \tag{170}$$

Note at $t = 0$ these eigenvectors are the standard vectors that perform the finite-dimensional discrete Fourier tranform. The eigenvalues of $\mathbf{M}(t)$ are obtained by computing the quadratic form:

$$\mu_r(t) =< x_r, \mathbf{M}(t)x_r >= \frac{1}{p}\sum_{k,l=0}^{p-1} \rho_{k-l}\varepsilon_r^{-(k-l)}(t) = \sum_{k=0}^{p-1} \rho_k(t)\varepsilon_r^{-k}(t)$$

because $\rho_k(t)\varepsilon_r^{k-p} = \rho_{k-p}(t)\varepsilon_r^k$. Furthermore, using the definition (38) of $\rho_r(t)$ we obtain:

$$\mu_r(t) = \sum_{l\in\mathbf{Z}} e^{2\pi i l \frac{q}{p}(t+\frac{r}{q})} R(\frac{l}{\alpha}) \tag{171}$$

Thus $\mu_r(t)$ is the discrete Fourier transform of the sequence $\{R(\frac{l}{\alpha})\}_{l\in\mathbf{Z}}$ evaluated at $\frac{q}{p}(t+\frac{r}{q})$. Note

$$\mu_r(t+\frac{1}{q}) = \mu_{r+1}(t). \tag{172}$$

relation that is useful in the localization problem treated in §4.6.

Suppose now R decays sufficiently fast (for instance $|R(x)| \leq \frac{C}{(1+|x|)^{1+\gamma}}$, for some $\gamma > 0$), then we can apply the Poisson summation formula (see for instance [Gröc96]) and obtain:

$$\mu_r(t) = \sqrt{2\pi}\alpha \sum_{l\in\mathbf{Z}} \hat{R}(2\pi\alpha l - \frac{2\pi\alpha}{p}(qt+r)) \tag{173}$$

where $\hat{R}(\omega) = \frac{1}{\sqrt{2\pi}}\int e^{-i\omega x}R(x)dx$ is the Fourier transform of $R(\cdot)$, and thus the spectral power density of the original signal. The equation (173) shows also that $\mu_r(t) \geq 0$ because the spectral power is always nonnegative. Thus we proved $\mathbf{M}(t)$ is nonnegative definite and the eigenproblem for $\mathbf{M}(t)$ is completely solved.

We study now the eigenproblem for $\mathbf{N}(t)$, defined by (51) and (52). A simple computation shows that $\mathbf{N}(t)$ has the same "quasi-cyclicity" property as $\mathbf{M}(t)$, i.e. $E(t)\mathbf{N}(t)E(t)^* = \mathbf{N}(t)$. Therefore $\mathbf{N}(t)$ has the same system of eigenvectors as $E(t)$, namely (170). In particular this shows that $\mathbf{M}$ and $\mathbf{N}$ commute. The eigenvalues of $\mathbf{N}(t)$ are computed similar to those of $\mathbf{M}(t)$:

$$\nu_r(t) =< x_r, \mathbf{N}(t)x_r >= \sum_{k=0}^{p-1} \sigma_k(t)\varepsilon_r^{-k}(t)$$

Explicitely, this turns into:

$$\nu_r(t) = \sum_{l\in\mathbf{Z}} e^{2\pi i l \frac{q}{p}(t+\frac{r}{q}+\frac{q}{2p})} R(\frac{l}{\alpha}) = \mu_r(t+\frac{p}{2q}) \tag{174}$$

Thus $\mathbf{N}$ has the same eigenvectors as $\mathbf{M}$ has at an argument shifted by $\frac{p}{2q}$. This also proves that $\mathbf{N}$ is nonnegative definite as well. These conclusions could have been obtained also from (156). Indeed, this relation shows that $\mathbf{N}(t)$ has the same eigenvalues as $\mathbf{M}(t+\alpha_0\beta_0)=\mathbf{M}(t+\frac{p}{2q})$. However (174) shows something more, namely these eigenvalues correspond to the same eigenvectors as $\mu_r(t)$'s correspond to for $\mathbf{M}(t)$. This remark allows us to write down immediately the eigenvalues of $\mathbf{M}+\mathbf{N}$:

$$\eta_r(t)=\mu_r(t)+\nu_r(t)=2\sum_{l\in\mathbf{Z}} e^{4\pi i l\frac{q}{p}(t+\frac{r}{q})}R(\frac{2l}{\alpha}) \tag{175}$$

Note $\eta_r(\cdot)$ is $\frac{p}{2q}$-periodic, whereas $\mu_r(\cdot)$ and $\nu_r(\cdot)$ are only $\frac{p}{q}$-periodic.

Now we can obtain explicit forms for some approximations errors. We compute $e_{opt}(\mathbf{WM};\alpha,\beta)$ and $e_{opt}(\mathbf{W}(\mathbf{M}+\mathbf{N});2\alpha_0,\beta_0)$ where $e_{opt}(\cdot;\cdot,\cdot)$ has been introduced in (115). The eigenvalues of $\mathbf{WM}$ are:

$$\lambda_r(\mathbf{WM})(t,s)=\omega(s)\mu_r(t)\quad,\quad r=0,1,\ldots,p-1 \tag{176}$$

To compute the optimal error we need to select the smallest $p-q$ eigenvalues. Let $I(t)$ denote an index set of the smallest $p-q$ eigenvalues. Thus for every $r\in I(t)$ and $j\in\{0,1,\ldots,p-1\}\setminus I(t)$, $\mu_r(t)\leq\mu_j(t)$. The index set $I(t)$ may not be unique if the q^{th} monotonically ordered eigenvalue is degenerate. In any case, $e_{opt}(\mathbf{WM};\alpha,\beta)$ turns into:

$$e_{opt}(\mathbf{WM};\alpha,\beta)=\beta q\int_0^{1/p}\omega(s)\,ds\int_0^{1/q}\sum_{r\in I(t)}\mu_r(t)\,dt$$

Since $\omega(s)$ is $\frac{1}{p}$-periodic, the first integral is $\frac{1}{\beta p}\|w\|_{L^1}$. Thus we obtain:

$$e_{opt}(\mathbf{WM};\alpha,\beta)=\frac{\|w\|_{L^1}}{\alpha\beta}\sum_{l\in\mathbf{Z}}R(\frac{l}{\alpha})\int_0^{1/q}\sum_{r\in I(t)}e^{\frac{2\pi i l}{\alpha\beta}(t+\frac{r}{q})}\,dt \tag{177}$$

For $e_{opt}(\mathbf{W}(\mathbf{M}+\mathbf{N});2\alpha_0,\beta_0)$ analog computations yield:

$$e_{opt}(\mathbf{W}(\mathbf{M}+\mathbf{N});2\alpha_0,\beta_0)=\frac{\|w\|_{L^1}}{\alpha\beta}\sum_{l\in\mathbf{Z}}R(\frac{2l}{\alpha})\int_0^{1/q}\sum_{r\in I(t)}e^{\frac{2\pi i l}{\alpha_0\beta_0}(t+\frac{r}{q})}\,dt \tag{178}$$

REMARK 34 *1. Let us assume now the map $t\mapsto\tilde{R}=\sum_l e^{2\pi i l\frac{q}{p}t}R(\frac{l}{\alpha})$ is continuous and monotonically decreasing on $[0,t^*]$ and increasing on $[t^*,\frac{p}{q}]$, for some $t^*\in[0,\frac{p}{q}]$. Then for a fixed t, the sum of the smallest $p-q$ eigenvalues is given by $\tilde{R}(t+\frac{r_0}{q})+\tilde{R}(t+\frac{r_0+1}{q})+\cdots+\tilde{R}(t+\frac{r_0+p-q-1}{q})$ for some $0\leq r_0\leq q$. Then, integrating over $t\in[0,\frac{1}{q}]$ we obtain:*

$$\int_0^{1/q}\sum_{r\in I(t)}\mu_r(t)\,dt=\int_{I_0}\tilde{R}(t)\,dt$$

where $I_0 = [t_0, t_0 + \frac{p-q}{q}]$ is an interval of length $\frac{p-q}{q}$ containing t^*. The monotonicity insures the contiguity of I_0. Then e_{opt} turns into:

$$e_{opt}(\mathbf{WM}; \alpha, \beta) = \frac{\|w\|_{L^1}}{\alpha\beta} \sum_{l \in \mathbf{Z}} R(\frac{l}{\alpha}) e^{\frac{2\pi i l}{\alpha\beta}(t_0 - \frac{\alpha\beta - 1}{2})} \frac{\sin(\frac{\pi l}{\alpha\beta}(\alpha\beta - 1))}{\frac{\pi l}{\alpha\beta}} \tag{179}$$

Furthermore, if the original signal has the correlation length smaller than $\frac{1}{\alpha}$, i.e. $R(x) = 0$ for $|x| \geq \frac{1}{\alpha}$, then the approximation error turns into a very simple expression:

$$e_{opt}(\mathbf{WM}; \alpha, \beta) = \frac{\alpha\beta - 1}{\alpha\beta} \|w\|_{L^1} R(0) \tag{180}$$

showing the decaying of the error proportional to $1 - \frac{1}{\alpha\beta}$ as plotted in Figure 5, bottom.

2. For the other approximation error $e_{opt}(\mathbf{W}(\mathbf{M} + \mathbf{N}); 2\alpha_0, \beta_0)$ the anlysis can be done similarly.

3. The knowledge of the eigenvectors (170) allows one to explicitely construct the optimizing window(s). However, one has first to find for every t the largest eigenvalues of $\mathbf{M}$ or $\mathbf{M} + \mathbf{N}$ and then to select the right eigenvectors in the columns of Γ. Later on we give some example of optimal windows.

4.4 The case $M(t) = \rho_0(t)I$

The second special case we propose to treat separately is when $\mathbf{M}(t)$ is a diagonal matrix for every t. Since $\mathbf{M}(t)$ is already Toeplitz, it follows that it has to be a multiple of the diagonal matrix, i.e. $\mathbf{M}(t) = \rho_0(t)I$. Such a situation can be achieved when $R(x) = 0$, for $|x| \geq \alpha$, i.e. the signal correlation length is smaller than α. One such a case is when the signal is white noise.

The solution of the optimization problems studied before is controlled in this case by the weight function $w(\cdot)$. The optimizer subspace is given at every (t, s) by the q-dimensional invariant subspaces of $\mathbf{W}(s)$ corrsponding to the largest q eigenvalues, i.e. $eigspacemax(\mathbf{W}(s); p, q)$. One case easily see that the optimization problem (154) reduces also to $eigspacemax(\mathbf{W}(s); p, q)$. The eigenvectors of $\mathbf{W}(s)$ are given by:

$$y_r(s) = [0 \ \ldots \ 0 \ 1 \ 0 \ \ldots \ 0]^T \tag{181}$$

i.e. by the canonical basis of $\mathbf{C}^p$ (the only "1" is on the $r + 1^{st}$ position in (181), $0 \leq r \leq p - 1$). The eigenvalues of $\mathbf{W}(s)$ are:

$$Spec(\mathbf{W}(s)) = \{\omega(s), \omega(s + \frac{q}{p}), \ldots, \omega(s + (p-1)\frac{q}{p})\} \tag{182}$$

or, since p and q are relatively prime:

$$Spec(\mathbf{W}(s)) = \{\omega(s), \omega(s+\frac{1}{p}), \ldots, \omega(s+\frac{p-1}{p})\} \tag{183}$$

The selection of y_r's is based on the largest q eigenvalues, respectively the largest q positive numbers of (182). When the q^{th} eigenvalue is degenerate, the invariant space is not unique. In fact there are infinitely many possible choices, i.e. $eigspacemax(\mathbf{W}; p, q)$ has an infinite cardinal. Let $\omega_r(s) = \omega(s + r\frac{q}{p})$. Note the following "twisting" relation, similar to (172):

$$\omega_r(s+\frac{1}{p}) = \omega_{\sigma(r)}(s) \tag{184}$$

where σ is a permutation of $\{0, 1, \ldots, p-1\}$. This relation will be useful in the localization analysis of the optimal window.

We analyze now the optimal approximation error e_{opt} given in (115). For $M(t) = \rho_0(t)I$ this expression turns into:

$$e_{opt} = \beta q \int_0^{1/q} \rho_0(t)dt \int_0^{1/p} \sum_{r\in I(s)} \omega(s+\frac{r}{p})ds \tag{185}$$

where $I(s)$ is the index set of the $p-q$ labels corresponding to the smallest eigenvalues in (183). Using the definition of $\rho_0(t)$ (38) we obtain:

$$\int_0^{1/q} \rho_0(t)dt = \frac{1}{q}R(0)$$

Thus:

$$e_{opt} = \beta R(0) \int_0^{1/p} \sum_{r\in I(s)} \omega(s+\frac{r}{p})ds \tag{186}$$

REMARK 35 *The upper bound obtained in (116) is more transparent here: first note that:*

$$\frac{1}{p-q} \sum_{r\in I(s)} \omega(s+\frac{r}{p}) \leq \frac{1}{p}\sum_{r=0}^{p-1} \omega(s+\frac{r}{p}) \tag{187}$$

Then the integrand in (186) is bounded above by $\frac{p-q}{p}\int_0^{1/p} trace\{\mathbf{W}(s)\}ds = \frac{p-q}{p}\frac{1}{\beta}\|w\|_{L^1}$. *Thus we get:*

$$e_{opt} \leq (1-\frac{1}{\alpha\beta})R(0)\|w\|_{L^1} \tag{188}$$

The equality in (187) is achieved only for $\mathbf{W}(s) = \omega(s)I$.

4.5 Optimal Windows - Examples

Let us consider the stationary signal f given by the output of a 4-pole Markov process whose transfer function is:

$$H(s) = \frac{158.1s^2(s^2 + 60s + 300^2)}{(s^2 + 20s + 100^2)(s^2 + 200s + 1000^2)} \tag{189}$$

We take the weight function w to be the characteristic function of the interval $[0, \beta]$. Thus:

$$\hat{R}(\xi) = |H(i\xi)|^2 \quad , \quad R(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\xi x} \hat{R}(\xi) d\xi \tag{190}$$

we plot in Figure 6 the autocovariance function in time and frequency domains, respectively.

Let us consider now the one-channel optimal problem. We set $\beta = 0.1$. For every $\alpha \geq 10$ such that $\alpha\beta = \frac{p}{q} \geq 1$, we construct $\mathbf{M}(t)$ using (40). Then the optimal window is parametrized by:

$$\Gamma(t, s) = F(t) \cdot L(t, s) \tag{191}$$

where $F \in eigmax(\mathbf{M}; p, q)$ and $L \in L^\infty(\square; GL_q(\mathbf{C}))$. Let us consider the solution with

$$L(t, s) = \left\{ \begin{array}{lll} I_q & , & t \leq \frac{1}{2q} \\ J_q & , & t > \frac{1}{2q} \end{array} \right. \tag{192}$$

where J_q is the $q \times q$ matrix with 1 on the anti-diagonal ($(J_q)_{l,r} = \delta_{l+r,q-1}$, $0 \leq l, r \leq q - 1$). Then the optimal window g is uniquely determined by $eigmax(\mathbf{M}; p, q)$. We choose the columns in F to be the eigenvectors of $\mathbf{M}$, ordered according to the corresponding eigenvalues $\mu_0(t), \ldots, \mu_p(t)$ of $\mathbf{M}$. Therefore, for those (t, s) where $\mu_0(t) > \mu_1(t) > \cdots > \mu_q(t)$, $F(t)$ is uniquely determined. Suppose this nondegeneracy condition holds for almost every t. Then Γ does not depend on s, for $s \in [0, \frac{1}{p}]$ which in turn implies $G(t, s)$ is piecewise constant for every t and $s \in [0, 1]$. The number of pieces is exactly p. Therefore the optimal window g is piecewise constant on $[0, \beta]$ and the number of pieces on this interval is exactly p.

For our case (189), this situation takes place as it can be seen in the Figures 7–9.

For various α (i.e. p and q) we plot the p eigenvalues of $\mathbf{M}(t)$ as function of t (Figures 10–12), and the optimal window obtained as described before (Figures 7–9). We point out this window is also a Weyl-Heisenberg orthogonal basis generator. Thus the biorthogonal window $g^\#$ coincides with a scaled version of g. The exact normalization is $\frac{1}{\sqrt{p}}g$ to have an orthonormal basis generator. The standard biorthogonal generator to g is $g^\# = \frac{1}{p}g$.

We prove when the signals are real, the optimal solution chosen as before is a real-valued function. First the necessary and sufficient condition for a function to be real in terms of the Zak transform is the following:

LEMMA 36 *$g \in L^2(\mathbf{R})$ is a real-valued function if and only if*

$$\overline{G(t,s)} = G(1-t,s) \tag{193}$$

or, equivalently, iff:

$$\overline{\Gamma(t,s)} = \Gamma(\frac{1}{q}-t,s)\cdot J_q \tag{194}$$

where $G = Zak(g)$ is the Zak transform of g, Γ is the matrix representation (37) and J_q is the $q \times q$ matrix with 1 on the anti-diagonal.

This lemma can be checked directly from the definition of the Zak transform (34) and of Γ in (37). Next, we check the symmetry properties of the eigenvectors (170). Since in $F(t)$ defined above, the eigenvectors are ordered according to the eigenvalues, we have to find the symmetry relations of the eigenvalues as well. Note the eigenvalues are proportional with $\mu_r(t)$ given in (171) or $\mu_r(t)+\nu_r(t)$ with $\nu_r(t)$ given in (174). Assume the signal is real. This implies the autocovariance function is real too and then:

$$R(-x) = R(x) \tag{195}$$

This symmetry of the autocovariance function implies immediatelly:

$$\mu_r(\frac{1}{q}-t) = \mu_{p-r-1}(t) \quad , \quad \nu_r(\frac{1}{q}-t) = \mu_{p-r-1}(t) \tag{196}$$

Therefore, if the 0-indexed column (i.e. the first one) of F at t is given by the eigenvector $x_r(t)$, then at $\frac{1}{p}-t$ the first column is given by $x_l(\frac{1}{q}-t)$ such that $\mu_l(\frac{1}{q}-t) = \mu_r(t)$. Hence $l = p-r-1$. Similar for the other columns. Thus, if $r_0(t),\ldots,r_{q-1}(t)$ are indices of the eigenvectors appearing in the columns of $F(t)$, at $\frac{1}{q}-t$, the indices are $r_0(\frac{1}{q}-t) = p-r_0(t)-1$, $r_1(\frac{1}{q}-t) = p-r_1-1,\ldots,r_{q-1}(\frac{1}{q}-t) = p-r_{q-1}-1$:

$$F_{:,l}(t) = x_{r_l(t)}(t) \quad , \quad F_{:,l}(\frac{1}{q}-t) = x_{p-r_l(t)-1}(\frac{1}{q}-t)$$

Note that $\varepsilon_r(t)$ defined in (169) obeies:

$$\overline{\varepsilon_r(\frac{1}{q}-t)} = \varepsilon_{p-1-r}(t)$$

hence:

$$\overline{x_r(\frac{1}{q}-t)} = x_{p-1-r}(t)$$

Thus:

$$F_{:,l}(\frac{1}{q}-t) = \overline{x_{r_l(t)}(t)} = \overline{F_{:,l}(t)}$$

or:

$$F(\frac{1}{q}-t) = \overline{F(t)} \tag{197}$$

(Alternatively, (197) could have been obtained also by noting that $\mathbf{M}(\frac{1}{q}-t) = \overline{\mathbf{M}(t)}$.) Note $L(t,s)$ has been chosen in (192) in such a way that we obtain:

$$\Gamma(\frac{1}{q}-t,s) = \overline{\Gamma(t,s)} \cdot J_q$$

which proves that g is real. The following result contains our findings so far:

THEOREM 37 *When the autocovariance function is real, the optimal window can be chosen to be real-valued.*

4.6 Localization of the optimal windows

We continue in this subsection the anlysis of the optimal solution of the one-channel problem. The other encoding-decoding schemes, except for the FSDE case, reduce to an equivalent one-channel optimal problem. The examples shown in the previous subsection suggest a "bad" localization in the time-frequency domain. For a given autocovariance function $R(\cdot)$ and a weight $w(\cdot)$, the optimizer is parametrized via (191), by some $L \in L^\infty(\Box; GL_q(\mathbf{C}))$, in general. In the examples shown in Figures 7–9, we made a particular choice for L to obtain real-valued windows, namely (192). One can ask whether by appropriately choosing L, the "bad" time-frequency localization phenomenon observed before can be avoided. The purpose of this subsection is to show for any choice of L, when the data satisfies some topological condition, the optimizer g is not well localized in time-frequeny domain, in a sense that is made more precise below.

Our method of proving requires the exact knowledge of the eigenvectors and eigenvalues. Thus we shall assume either $\mathbf{W}(s) = \omega(s) I_p$ or $\mathbf{M}(t) = \rho_0(t) I_p$. The general case still remains an open problem, though some perturbative arguments may extend the nonlocalization results that we obtain, a bit further than the mentioned cases.

For our purposes a window g is said *well-localized in time-frequency domain* if $g \in Q^{1,1} \cup C(L^\infty, l^1)$ or $\hat{g} \in Q^{1,1} \cup C(L^\infty, l^1)$. The space $Q^{1,1}$ and Wiener amalgam space $C(L^\infty, l^1)$ has been introduced in A, by (236), respectively (237). As it can be easily checked, if a function g is well-localized in time-frequency domain (as defined above), then the Zak transform G is either a continuous function (if $g \in C(L^\infty, l^1)$ or $\hat{g} \in C(L^\infty, l^1)$), i.e. $G \in C(\square; \mathbf{C})$, or the derivatives of G with respect to t and s are square integrable (if $g \in Q^{1,1}$), i.e. $G \in W^{1,2}(\square, \mathbf{C})$. Thus, in order to prove g is *not* well-localized (or, equivalently, is *bad* localized) in TF domain, we have to show that $G \notin C(\square; \mathbf{C}) \cup W^{1,2}(\square; \mathbf{C})$, i.e. G is discontinuous and $\int\int_\square |\frac{\partial G}{\partial t}|^2 dt\, ds = \infty$ or $\int\int_\square |\frac{\partial G}{\partial s}|^2 dt\, ds = \infty$ (see [BeHeWa95]).

Consider now the data $(R, w; p, q)$ (the autocovariance and weight functions and $\alpha\beta = \frac{p}{q} \geq 1$) for the one-channel optimal problem. With this data we construct the matrix-valued functions $\mathbf{M}$ and $\mathbf{W}$ as in (40) and (42). For every (t, s) we denote by $\lambda_j(t, s)$, $j = 0, \ldots, p-1$, the p real eigenvalues of $\mathbf{W}(s)\mathbf{M}(t)$ ordered monotonically decreasing. Consider now the following sets of points of $\square = [0, \frac{1}{q}] \times [0, \frac{1}{p}]$:

$$D_0 = \{(t, s) \in \square \mid \lambda_q(t, s) > \lambda_{q+1}(t, s)\} \tag{198}$$

$$D_\infty = D_0 \setminus D_0 \tag{199}$$

Note D_0 agrees with the definition in (109), and $D_\infty = \cup_{l=1}^{p-q} D_l$ with the same notations as in (109).

The data $(R, w; p, q)$ is said *nondegenerate on a dense subset* if D_0 is dense in $\square$.

The data is said *sufficiently regular* if the spectral projector associated to the eigenvalues $\lambda > \lambda_{q+1}(t, s)$ is continuous on D_0.

For instance is enough that $\mathbf{M}$ and $\mathbf{W}$ are continuous for the data $(R, w; p, q)$ to be sufficiently regular (see Proposition 9). Hence, when the data is sufficiently regular and nondegenerate on a dense subset, D_∞ is a closed subset with empty interior in $\square$. This also shows the following property:

LEMMA 38 *Suppose $(R, w; p, q)$ is a data sufficiently regular and nondegenerate on a dense subset. Let $f : \square \to \mathbf{R}$ be a function such that on $D_0 \subset \square$ it takes only values on a discrete set (for instance $\{0, 1, \ldots, p-1\}$). Then, if f is continuous it follows that f should be constant on $\square$.*

The next lemma extends the previous result to multi-valued maps (or selection maps):

LEMMA 39 *Suppose the data $(R, w; p, q)$ is sufficiently regular and nondegenerate on a dense subset. Let $f_0, \ldots, f_{p-1} : \Box \to \mathbf{C}^p$ be p continuous vector-valued maps on $\Box$ such that at every point $(t,s) \in \Box$, $\{f_0(t,s), \ldots, f_{p-1}(t,s)\}$ form a basis in $\mathbf{C}^p$. Let $\mathcal{V} : \Box \to \mathcal{L}_q(\mathbf{C}^p)$ be a q-dimensional subspace-valued map on $\Box$ such that on D_0, $\mathcal{V}$ is spanned by some q vectors from $\{f_0, \ldots, f_{p-1}\}$, i.e. there is a selection map $\pi : D_0 \times \{0, 1, \ldots, q-1\} \to \{0, 1, \ldots, p-1\}$ such that:*

$$\mathcal{V}(t,s) = span\{f_{\pi(t,s;0)}(t,s), \ldots, f_{\pi(t,s;q-1)}(t,s)\} \quad , \quad \forall (t,s) \in D_0 \tag{200}$$

Then, if $\mathcal{V}$ is continuous on $\Box$ with respect to the graph topology (or, equivalently, the orthonormal projection onto $\mathcal{V}$ at every (t,s) is continuous with respect to the norm topology), then the selection map is a permutation of the same index set I, i.e. $Ran\,\pi(t_1, s_1; \cdot) = Ran\,\pi(t_2, s_2; \cdot) = I$ for every $(t_1, s_1), (t_2, s_2) \in D_0$. Moreover, in this case $\mathcal{V}(t,s) = span\{f_i(t,s)\ ;\ i \in I\}$ for every $(t,s) \in \Box$.

Proof Despite of its rather long statement, the proof of this lemma is relatively simple. Suppose $D_0 = \cup_{a \in A} H_a$ is the decomposition of D_0 into its arcwise connected components. Thus each H_a is an open arcwise connected subset of $\Box$. The continuity of $\mathcal{V}$ implies that $\pi|_{H_a}$ is a permutation of the same index set I_a, $Ran\,\pi(t,s;\cdot) = I_a$, for every $(t,s) \in H_a$. Next, consider two neighborhing subspaces H_{a_1} and H_{a_2} such that $\overline{H_{a_1}} \cap \overline{H_{a_2}} \neq \emptyset$, where $\overline{H_a}$ is the closure (with respect to the usual euclidian topology) of H_a in $\Box$. This is always possible since there is no isolated subset H_a in D_0 (recall D_0 is dense). Now, the same continuity argument implies that $I_{a_1} = I_{a_2}$ and, furthermore, on the common boundary:

$$\mathcal{V}(t,s) = span\{f_i(t,s)\ ,\ i \in I_{a_1} = I_{a_2}\} \quad , \quad (t,s) \in \partial H_{a_1} \cap \partial H_{a_2}$$

where $\partial H_a = \overline{H_a} \setminus H_a$ is the boundary of H_a. This ends the proof of the lemma. $\Box$.

REMARK 40 *The two essential ingredients in this lemma are the continuity of $\mathcal{V}$ and the fact that the arcwise connected components of D_0 are not isolated. Thus, for every continuous curve $\gamma : [0,1] \to \Box$, the intersection of its image with D_∞ is either empty or is made out of isolated points.*

Now let us return to our optimization solution Γ. In the one-channel case, Γ is chosen in such a way that $Ran\,\Gamma|_{(t,s)}$ is a q-dimensional subset of a family of q-dimensional subspaces. Moreover, at $(t,s) \in D_0$ $Ran\,\Gamma|_{(t,s)}$ is constrained to a unique subspace. Thus we are in a position to check whether $\mathcal{V}(t,s) = Ran\,\Gamma|_{(t,s)}$ can be a continuous map on $\Box$. Under regularity and density conditions of the initial data, if $\mathcal{V}$ is continuous it follows the selection map is a permutation of the some index set of eigenvectors.

But this is not possible when $\mathcal{V}$ is chosen to be spanned by the eigenvectors corresponding to the largest eigenvalues because of twisting relations (172) and (184) as we prove next:

LEMMA 41 (The case $\mathbf{W}(s) = \omega(s)I_p$) *Suppose the data $(R, w; p, q)$ is sufficiently regular and non-degenerate on a dense subset, and, additionally, $\mathbf{W}(s) = \omega(s)I_p$. Then, if $(t_0, s_0) \in D_0$, (t_0, s_0) and $(t_0 + \frac{1}{q}, s_0 + \frac{1}{p})$ belong to different connected components of D_0, and every $\mathcal{V} \in eigspacemax(\mathbf{WM}; p, q)$ is discontinuous.*

LEMMA 42 (The case $\mathbf{M}(t) = \rho_0(t)I_p$) *Suppose the data $(R, w; p, q)$ is sufficiently regular and non-degenerate on a dense subset, and, additionally, $\mathbf{M}(t) = \rho_0(t)I_p$. Then, if $(t_0, s_0) \in D_0$, (t_0, s_0) and $(t_0 + \frac{1}{q}, s_0 + \frac{1}{p})$ belong to different connected components of D_0, and every $\mathcal{V} \in eigspacemax(\mathbf{WM}; p, q)$ is discontinuous.*

REMARK 43 *The conclusions of these lemmas also imply that for any continuous curve γ connecting (t_0, s_0) to $(t_0 + \frac{1}{q}, s_0 + \frac{1}{p})$, there is at least one transversal intersection of the curves λ_q and λ_{q+1}, or, since $\lambda_q \geq \lambda_{q+1}$, λ_q is not differentiable along that curve γ.*

Proof of Lemma 41

In the case $\mathbf{W}(s) = \omega(s)I_p$ the eigenvectors are given by $x_r(t)$, $r = 0, 1, \ldots, p-1$, given in (170). Let $\mu_r(t)$ be the associated eigenvalue for $M(t)$ as given in (171). Note $\lambda_j(t, s) = \omega(s)\mu_{\tau(j)}(t)$ for some permutation τ depending on t. we prove the claim by contradiction. Suppose $\gamma : [0, 1] \to \square$ is a continuous curve connecting (t_0, s_0) to $(t_0 + \frac{1}{q}, s_0 + \frac{1}{p})$. First note that $M(t_0 + \frac{1}{q}) = M(t_0)$, hence from $(t_0, s_0) \in D_0$ it follows $(t_0 + \frac{1}{q}, s_0 + \frac{1}{p}) \in D_0$ as well. Using Lemma 39, it follows that $\mathcal{V}(t_0, s_0)$ and $\mathcal{V}(t_0 + \frac{1}{q}, s_0 + \frac{1}{p})$ are spanned by the same eigenvector labels, say $\{r_0, r_1, \ldots, r_{q-1}\} \subset \{0, 1, \ldots, p-1\}$. On the other hand $\mu_r(t_0 + \frac{1}{q}) = \mu_{r+1}(t_0)$ (see (172)). Thus it necessarily holds true:

$$\sum_{l=0}^{q-1} \mu_{r_l}(t_0 + \frac{1}{q}) \neq \sum_{l=0}^{q-1} \mu_{r_l}(t_0) \tag{201}$$

On the other hand $\mathcal{V} \in eigspacemax(\mathbf{WM}; p, q)$ that implies the eigenvalues of $\mathbf{WM}|_{\mathcal{V}}$ should be the largest q eigenvalues of $\mathbf{WM}$. In particular, because of periodicity, we should have $trace\{\mathbf{WM}|_{\mathcal{V}(t_0,s_0)}\} = trace\{\mathbf{WM}|_{\mathcal{V}(t_0+\frac{1}{q},s_0+\frac{1}{p})}\}$. But this is a contradiction with (201). This end the proof of this lemma. $\diamond$.

Proof of Lemma 42

Note first that since $\mathbf{W}(s + \frac{1}{p})$ is equivalent to $\mathbf{W}(s)$, if $(t_0, s_0) \in D_0$ then $(t_0 + \frac{1}{q}, s_0 + \frac{1}{p}) \in D_0$ as well. On D_0, $\mathcal{V}$ is spanned by some q vectors of $\{y_0, y_1, \ldots, y_{p-1}\}$, where y_r is defined in (181).

Note the correspinding eigenvalues of $\mathbf{W}$ are $\omega_r(s) = \omega(s + r\frac{q}{p})$. Suppose the conclusion of this lemma is fale. Then there is a continuous curve γ connecting (t_0, s_0) to $(t_0 + \frac{1}{q}, s_0 + \frac{1}{p})$ such that $\mathcal{V}|_\gamma$ is continuous. In particular, using Lemma 39, it follows the eigenvalues of $\mathbf{W}|_{\mathcal{V}(t_0+\frac{1}{q}, s_0+\frac{1}{p})}(t_0 + \frac{1}{q}, s_0 + \frac{1}{p})$ are $\omega_{r_0}(s_0 + \frac{1}{p}), \ldots, \omega_{r_{q-1}}(s_0 + \frac{1}{p})$ where $r_0, \ldots, r_{q-1}$ are the indices of the largest q eigenvalues of $\mathbf{W}$ at (t_0, s_0). Because of (184):

$$\sum_{l=0}^{q-1} \omega_{r_l}(s_0 + \frac{1}{p}) \neq \sum_{l=0}^{q-1} \omega_{r_l}(s_0)$$

which is a contradiction with the maximality of the eigenvalues of $\mathbf{W}|_{\mathcal{V}}$. $\diamond$

These two lemmas have the following conclusion:

THEOREM 44 *Suppose the data $(R, w; p, q)$ is nondegenerate on a dense subset. Additionally suppose that either $\mathbf{W}(s) = \omega(s)I_p$ or $\mathbf{M}(t) = \rho_0(t)I_p$. Then any window defined by $\Gamma(t,s) = F(t,s)L(t,s)$ with $F \in eigmax(\mathbf{WM}; p, q)$ and $L \in L^\infty(\Box; GL_q(\mathbf{C}))$ is bad localized in time-frequency domain, in that $g \notin C(L^\infty, l^1) \cup Q^{1,1}$ and $\hat{g} \notin C(L^\infty, l^1) \cup Q^{1,1}$.*

Proof

If the data $(R, w; p, q)$ is not sufficiently regular, then the spectral projector onto the eigenvectors corresponding to the largest q eigenvalues is discontinue and then g will have even less regularity than in the case when the data is sufficently regular. Thus we can assume the data is sufficiently regular.

We have to prove two conditions: G is discontinuous and $\frac{\partial G}{\partial t}$ or $\frac{\partial G}{\partial s}$ is not square integrable over $\Box$.

The discontinuity of G comes directly from the previous two lemmas because otherwise $\mathcal{V} = Ran\,\Gamma \in eigspacemax(\mathbf{WM}; p, q)$ would be continuous.

The nonintegrability condition is proved as follows. First note in any of the two cases $\mathbf{W}(s) = \omega(s)I_p$ or $\mathbf{M}(t) = \rho_0(t)I_p$, G turns out discontinuous on at least a straight segment $\{(t_0, s)\,,\, |s - s_0| < \varepsilon\}$ or $\{(t, s_0)\,,\, |t - t_0| < \varepsilon\}$ parallel with one of the axes. Then by Fubini theorem:

$$\int\!\!\int_{\substack{|t-t_0|<\varepsilon \\ |s-s_0|<\varepsilon}} dt\, ds\, |\frac{\partial G}{\partial n}|^2 = \int_{t_0-\varepsilon}^{t_0+\varepsilon} dt \int_{s_0-\varepsilon}^{s_0+\varepsilon} ds\, |\frac{\partial G}{\partial n}|^2 = \int_{s_0-\varepsilon}^{s_0+\varepsilon} ds \int_{t_0-\varepsilon}^{t_0+\varepsilon} dt\, |\frac{\partial G}{\partial n}|^2$$

where n is the normal direction to the segment (t in the first case, and s in the second case). Then one of the inner integrals $\int_{s_0-\varepsilon}^{s_0+\varepsilon} |\frac{\partial G}{\partial n}|^2 ds$ or $\int_{t_0-\varepsilon}^{t_0+\varepsilon} |\frac{\partial G}{\partial n}|^2 dt$ is infinite (because a discontinuous function on the line cannot have a square-integrable derivative) and therefore the left-hand side is infinite as well. In conclusion we obtain that $\frac{\partial G}{\partial t}$ and $\frac{\partial G}{\partial s}$ cannot both be in $L^2(\Box; \mathbf{C})$ which is equivalent to $g \notin Q^{1,1}$. This ends the proof of the Theorem. $\diamond$

REMARK 45 *1. The discontinuity of G comes as a result of the discontinuity of $\mathcal{V} \in eigspacemax(\mathbf{WM}; p, q)$. This means the rank q projector-valued function associated to $\mathcal{V}$ that is uniquelly defined on D_0, cannot be extended to a continuous projector-valued function on $\Box$.*

2. As in the standard Balian-Low Theorem, the second part of non-localization (i.e. $g \notin Q^{1,1}$) does not come automatically from G discontinuous via a Sobolev embedding argument, because Sobolev embedding theorem in dimension 2 does not require continuity for functions in $W^{1,2}$.

5 Near-Optimal Solution and Distortion-Rate Estimates

The previous section showed the optimal solution is not well-localized in the time-frequency domain. Moreover, in the Frequency-Shift Division cases, the optimal solution does not satisfy in general the frame condition $H_2(g; \alpha, \beta)$, unless some very strong geometric conditions are satisfied. Naturally, in such cases we can ask to find a near-optimal solution. In this section we show in a case study how to design or choose a near-optimal solution and next we obtain some asymptotic estimates regarding the transmission rate when a "nice" window is used (i.e. a well-localized in time-frequency domain). In fact the TF localization of the window is a first important factor in determining the rate. Thus for better-localized windows, the filter lengths used to implement the encoders and decoders are smaller and the number of coefficients in frequency label (i.e. m label) is smaller than in the non well-localized case.

5.1 Near-Optimal Solutions: A Case Study

Let us study closer the example introduced in subsection 4.5. Thus consider a signal f whose autocovariance function is given by (190) and for the weight we choose $w = 1_{[0,2\beta]}$. Consider $\beta = 0.05$ (note is half the value considered in §4.5) and let us concentrate on the Time-Shift Division Optimal Problems. More specific we shall consider the TSDED problem where the encoder is characterized by a window g and the decoder is determined by a window $g^{\#}$. As proved in Theorem 22 the optimal problem reduces to a one-channel problem with $\beta_{equiv} = 2\beta = 0.1$. This latter problem has been analyzed in §4.5. The Figures 7-9 show some real-valued optimal windows. In general any optimal solution is bad-localized in TF plane as we proved in Theorem 44. On the other hand we know for any $\varepsilon > 0$ and α, there is a near-optimal solution $(g, g^{\#})$ in that the criterion $J(g, g^{\#}; R, w; \alpha, \beta)$ is not larger than $1 + \varepsilon$ times the optimal value $J_{st,opt}^{1+2;TSED}(R, w; \alpha, \beta)$ given in (140). Our problem here is to find well-localized near-optimal solutions. Suppose $\varepsilon_{max} = 5\%$ is the allowed tolerance to the optimal criterion.

We start by choosing a particular window for the encoder and computing the best associated decoding window. For this we use the solution of the partial optimal problem TSDED given in Theorem 5, case B 13. We choose a gaussian function for the encoding window. More specific, consider:

$$g(x) = e^{-1000x^2} \tag{202}$$

for the encoder. For the partial optimal $g^{\#}$ we use (90) and for the criterion (91). The numerical results are compared in Table 1 (note that $J_{max} = 2\|w\|_{L^1} R(0) = 26.88$ would be obtained for $\alpha\beta = \infty$).

A_{Riesz} and B_{Riesz} are the Riesz basis bounds obtained for $\{g_{mn;\alpha,2\beta}\}$, whereas A_{frame} and B_{frame} are the frame bounds obtained for $\{g_{mn;\alpha,\beta}\}$ with $2\alpha\beta = \frac{p}{q}$ and $\varepsilon = (e^{1+2;TSDED}_{st,so} - e^{1+2;TSDED}_{st,opt})/e^{1+2;TSDED}_{st,opt}$. The decoding window $g^{\#}$ is plotted in Figures 13–16 for each case. Note how well-localized in TF domain each window is. Basically we can very well approximate the window by compactly supported functions in both time and frequency domain. Practically, this implies short filters for both the analog encoder and analog decoder.

We would like to have the criterion smaller than 1.05 times the optimal value. Notice the gaussian window (202) satisfies this condition for all the redundancy values except $\frac{p}{q} = \frac{6}{5}$ and $\frac{p}{q} = \frac{7}{6}$. For these particular valuee we have to choose different encoding windows. Here is how we proceed for these two cases. In the first step we analyze the optimal window and find some frequency conditions. In the second step we design an encoding window satisfying these constraints Finally, in the third step we adjust, if needed, the decoding window to have a better time-frequency localization satisfying, at the same time the near-optimality condition.

The optimal encoding windows given by:

$$\Gamma = F(t) \cdot L(t,s)$$

with $F \in eigmax(\mathbf{M};\alpha,2\beta)$ (recall $\mathbf{W} = I_p$ in our case) and $L \in L^\infty(\square; GL_q(\mathbf{C}))$ as in (192), are represented in Figure 17 and 18. Note the symmetry axis in time domain has an offset from the origin. Thus the Fourier transform of the window has an imaginary component too. In any case, the frequency plots suggest to choose encoding windows that have no DC components. And that what we do next.

The previous frequency constraint imposes to choose a window of zero mean. The s-Riesz basis and frame constraints require the measure of the window support to be at least $\frac{1}{\alpha}$. One simple choice satisfying these constraints is:

$$g(x) = \left\{ \begin{array}{rl} -1\ , & x \in [-\beta, 0] \\ 1\ , & x \in [0, \beta] \end{array} \right. \tag{203}$$

shown in Figure 19. For this choice we obtain the results of Table 2.

The approximation error is less than 3% larger than the optimal value. The time domain plots of the partial optimal duals for (203) are shown in Figure 20. Note the time localization of both the encoding and decoding windows is very good. The frequency localization is not so good, but this is due to the temporal discontinuities of these windows. For practical purposes, their form is good enough. Hence we do not need a third step, to filter the decoding window function. Other example is given in [Balan98], where all these steps are effectively applied to obtain windows well localized in TF plane.

5.2 Rate and Distortion Estimation

Let us return to our transmision scheme in Figure 1. The digital encoders will encode only those meaningful coefficients. By meaningful we mean those coefficients whose variance is greater than a threshold. The labels of these coefficients can be determined a priori, based on the autocovariance function R and the encoding window g. Suppose $\frac{\Delta^2}{12}$ is this threshold. (This corresponds to a uniform quantizer with interlevel Δ). As lower the threshold, as larger the number of coefficients to be encoded and transmitted, and hence as higher the rate. At the other limit, as higher the threshold as larger the distortion (i.e. the reconstruction error). Thus there is a trade-off between distortion and rate realized by this threshold. However, in our case we want to keep the threshold fixed and realize the trade-off by changing the redundancy of the encoding scheme. In the following, the threshold is assumed fixed and the redundancy is the free parameter. However, to obtain analytic expressions, we need to consider the asymptotic limit $\Delta \rightarrow 0$.

The analysis is done in the following steps: first we compute the variance of the coefficients $< f, g^1_{mn} >$ and $< f, g^2_{mn} >$ that are outputed from the analog encoder. Next, using a Gaussian model for the signal, we can estimate the number of bits (in average) needed to encode this coefficient when the entropic encoder is used. Thus we obtain an exact formula for the rate. Also we can obtain an upper bound for the distortion when only a finite number of coefficients are sent. However, to obtain the qualitative behaviour of these we shall make the asymptotic analysis for $\Delta \rightarrow 0$. We also make the assumption the signal is real. This implies the autocovariance is a real and even function.

Here we make the analysis for one channel. The same result holds for all the two-channel cases, although some adjustment of the formula might be necessary because of different encoding windows. We assume the encoding window is sufficiently well localized in time frequency domain. For instance the frequency band of the window is much smaller than the frequency range of the stochastic process (i.e. of the spectral power).

First, let us evaluate the variance of the coefficents. We denote by g the window that defines the encoder. Thus we are interested in estimating $\mathbf{E}[|c_{mn}|^2], \mathbf{E}[|Re(c_{mn})|^2], \mathbf{E}[|Im(c_{mn})|^2]$ where $c_{mn} =< f, g_{mn;\alpha,\beta} >$. Assuming $R \in L^2(\mathbf{R})$ and $g \in L^{4/3}(\mathbf{R}) \cap L^2(\mathbf{R})$ these variances are easily obtained as (for details see [Balan98]):

$$\sigma_{mn}^2 = \mathbf{E}[|c_{mn}|^2] = \sqrt{2\pi} \int_{-\infty}^{\infty} d\xi \hat{R}(\xi) |\hat{g}(\xi - 2\pi m\alpha)|^2 \leq \sqrt{2\pi} \|R\|_{L^2} \|\hat{g}\|_{L^4}^2 \leq \sqrt{2\pi} \|R\|_{L^2} \|g\|_{L^{4/3}}^2 \tag{204}$$

$$\begin{aligned} \mu_{mn}^2 = \mathbf{E}[|Re(c_{mn})|^2] = & \frac{\sqrt{2\pi}}{4} \int_{-\infty}^{\infty} d\xi \hat{R}(\xi) |e^{2\pi i m n \alpha\beta} \hat{g}(\xi - 2\pi m\alpha) \\ & + e^{-2\pi i m n \alpha\beta} \overline{\hat{g}(-\xi - 2\pi m\alpha)}|^2 \end{aligned} \tag{205}$$

$$\begin{aligned} \nu_{mn}^2 = \mathbf{E}[|Im(c_{mn})|^2] = & \frac{\sqrt{2\pi}}{4} \int_{-\infty}^{\infty} d\xi \hat{R}(\xi) |e^{2\pi i m n \alpha\beta} \hat{g}(\xi - 2\pi m\alpha) \\ & - e^{-2\pi i m n \alpha\beta} \overline{\hat{g}(-\xi - 2\pi m\alpha)}|^2 \end{aligned} \tag{206}$$

Note in general the variances of the real and imaginary parts may depend on n. In any case, this dependency is q-periodic.

The scheme works in the following way: in a $q\beta$ time interval, say $[N\beta, (N+q+1)\beta)$, the transmitter has to send the meaningful coefficients $c_{m,n}$ (or $c_{m,n-d}$ for some fixed delay $d > 0$) for $n \in \{N, N+1, \dots, N+q\}$. The meaningful coefficients are those given by $\mu_{mn}^2 \geq \frac{\Delta^2}{12}$ or $\nu_{mn}^2 \geq \frac{\Delta^2}{12}$ (when each real and imaginary part is quantized separately). Using the Riemann-Lebesgue lemma (see [Rudin74]) there exists a $M > 0$ such that for every $|m| > M$, $\sigma_{mn}^2 \leq \frac{\Delta^2}{12}$. Thus we have to send only a finite number of quantized values.

In the second step of our analysis we have to assume a particular distribution for the signal. Suppose the signal is gaussian. Then, when the entropic encoder is used, for the threshold Δ the number of bits R_x needed to quantize a gaussian random variable x with zero-mean and variance $\mathbf{E}[|x|^2]$ is given by:

$$\mathbf{E}[|x|^2] = \frac{\Delta^2}{12} 2^{2R_x} \tag{207}$$

(see [Davi72]). This yields the following rate:

$$Rate = \frac{1}{2q\beta} \sum_{\substack{m,n \\ \mu_{mn}^2 \geq \frac{\Delta^2}{12}}} \log_2(\frac{12}{\Delta^2}\mu_{mn}^2) + \frac{1}{2q\beta} \sum_{\substack{m,n \\ \nu_{mn}^2 \geq \frac{\Delta^2}{12}}} \log_2(\frac{12}{\Delta^2}\nu_{mn}^2) \tag{208}$$

where we assume we encode independently the real and imaginary parts of c_{mn}.

Let us analyze now the distortion obtain through this scheme. In (31) we obtained an upper bound for the distortion in this transmission scheme. We now analyze further the terms in that formula.

Let $g^{\#}$ the decoding window. Let $\mathcal{S}_1 = \{(m,n)|\mu_{mn}^2 \geq \frac{\Delta^2}{12}\}$, $\mathcal{S}_2 = \{(m,n)|\nu_{mn}^2 \geq \frac{\Delta^2}{12}\}$. Then the reconstructed signal has the following form:

$$f^{recon} = \sum_{(m,n)\in\mathcal{S}_1} Q_\Delta(Re(< f, g_{mn} >))g_{mn}^{\#} + i \sum_{(m,n)\in\mathcal{S}_2} Q_\Delta(Im(< f, g_{mn} >))g_{mn}^{\#}$$

Then:

$$\begin{aligned}
\sqrt{Distortion} \quad &\leq \quad (\mathbf{E}\|f - \sum_{m,n} c_{mn} g_{mn}^{\#}\|_{L_w^2}^2)^{1/2} \\
&+ \quad (\mathbf{E}\| \sum_{(m,n)\notin\mathcal{S}_1} Re(< f, g_{mn} >)g_{mn}^{\#} + i \sum_{(m,n)\notin\mathcal{S}_2} Im(< f, g_{mn} >)g_{mn}^{\#}\|_{L_w^2}^2)^{1/2} \\
&+ \quad (\mathbf{E}\| \sum_{(m,n)\in\mathcal{S}_1} (Re(< f, g_{mn} >) - Q_\Delta(Re(< f, g_{mn} >)))g_{mn}^{\#} \\
&+ \quad i \sum_{(m,n)\in\mathcal{S}_2} (Im(< f, g_{mn} >) - Q_\Delta(Im(< f, g_{mn} >)))g_{mn}^{\#}\|_{L_w^2}^2)^{1/2} \\
&= \quad \sqrt{J} + \sqrt{J_\varepsilon} + \sqrt{J_q}
\end{aligned} \tag{209}$$

where J represents *the stochastic approximation error* due to the incompleteness of the set $\{g_{mn}; m,n \in \mathbf{Z}\}$ in $L^2(\mathbf{R})$; J_ε is *the truncation error* and represents those coefficients that are excluded from encoding; J_q is *the quantization error* and is due to the uncertainty introduced by the quantizer. Our problem is to bound and control each term.

Now recall the upper bound given in Lemma 59, Appendix B. Let $B^{\#}$ denote the norm of the reconstruction operator $T_{g^{\#}}^* : l^{2,\infty} \to L_w^2$. By the lemma above, an upper bound for this norm is:

$$B^{\#} \leq B^{2,\infty} := \sum_n \| \sum_k w(\cdot + \frac{k}{\alpha})|g^{\#}(\cdot + \frac{k}{\alpha} - n\beta)|^2\|_{L^\infty(0,\frac{1}{\alpha})}$$

For an arbitrary distribution of $< f, g_{mn} >$, the difference $|Re(< f, g_{mn} >) - Q_\Delta(Re(< f, g_{mn} >))| \leq \frac{\Delta}{2}$ which implies $\mathbf{E}[|Re(< f, g_{mn} >) - Q_\Delta(Re(< f, g_{mn} >))|^2] \leq \frac{\Delta^2}{4}$. The same relation holds true for the imaginary part too. However, if we assume the signal f is gaussian, the upper bound becomes $\frac{\Delta^2}{12}$ instead of $\frac{\Delta^2}{4}$. The same thing is obtained if we assume the $< f, g_{mn} >$ is uniformly distributed on each quantization interlevel. Anyway in general we obtain:

$$J_q \leq B^{\#} \frac{\Delta^2}{4} \sup_n (\#\mathcal{S}_{1n} + \#\mathcal{S}_{2n}) \tag{210}$$

where $\mathcal{S}_{1n} = \{(m,n) \in \mathcal{S}_1\}$, $\mathcal{S}_{2n} = \{(m,n) \in \mathcal{S}_2\}$. Assuming symmetry between the distribution of real and imaginary parts of the coefficients c_{mn} we get:

$$J_q \leq 2B^{2,\infty} \frac{\Delta^2}{4} (\#\mathcal{S}) \tag{211}$$

where $\mathcal{S} = \{m | \mathbf{E}[|c_{mn}|^2] \geq \Delta^2/12\}$. We give now a rough evaluation of the cardinality of $\mathcal{S}$ based on (204) and the following assumptions: $\hat{R}(\xi)$ is concentrated in a band of size $2b_R$ (2 because $\hat{R}$ is even in frequency domain - recall we assumed real-valued signals) and the support of $\hat{g}$ is much narrower than $2b_R$. Then the number of coefficients is roughly constant and it is given by:

$$\#\mathcal{S} \approx \frac{2b_R}{2\pi\alpha} = \frac{b_R}{\pi\alpha}$$

Thus:

$$J_q \approx \frac{b_R B^{2,\infty}}{2\pi\alpha} \Delta^2 \sim C \ \Delta^2 \tag{212}$$

which says that J_q decays to 0 as Δ^2 when $\Delta \to 0$.

For the truncation error, using again Lemma 59 (see Appendix A) we obtain a first estimate of the form:

$$J_\varepsilon \leq B^{2,\infty} (\sup_n \sum_{m \notin \mathcal{S}_{1n}} \mathbf{E}[|Re(c_{mn})|^2] + \sup_n \sum_{m \notin \mathcal{S}_{2n}} \mathbf{E}[|Im(c_{mn})|^2])$$

Next, assuming again a symmetry in the distribution of the real and imaginary part we obtain:

$$J_\varepsilon \leq 2B^{2,\infty} \sup_n \sum_{m \notin \mathcal{S}} \mathbf{E}[|c_{mn}|^2] = 2B^{2,\infty} \sum_{m \notin \mathcal{S}} \sigma_{mn}^2$$

with $\mathcal{S} = \{m | \sigma_{mn}^2 > \frac{\Delta^2}{12}\}$. Assuming $\mathcal{S} = [-M, M]$ we obtain:

$$J_\varepsilon \leq 2\sqrt{2\pi} B^{2,\infty} \sum_{|m| \geq M} \int_{-\infty}^{\infty} \hat{R}(\xi) |\hat{g}(\xi - 2\pi m\alpha)|^2 d\xi \tag{213}$$

The assumptions made before to obtain (212) would now give $J_\varepsilon = 0$. Thus if we assume that both the autocovariance function and the window are band-limited, we get rid of the truncation error provided we take into account all the (finite) non-zero coefficients.

Another (more realistic) model of R and g is to assume that both decay in frequency domain as:

$$|\hat{R}(\xi)| \leq \frac{C_1}{(1+|\xi|)^a} \quad , \quad |\hat{g}(\xi)| \leq \frac{C_2}{(1+|\xi|)^b} \tag{214}$$

with $a, b > 1$. The assumption on $\hat{R}$ is particularily useful when we assume that our signal is the output of a linear system excited by white noise. Then $\hat{R}(\xi) = |H(i\xi)|^2$ where $H(s)$ is the linear system transfer function. We shall give an asymptotic estimation of the rate and the truncation and quantization errors.

We start by estimating the variance σ_{mn}^2:

$$\begin{aligned}
\sigma_{mn}^2 &= \sqrt{2\pi}\int_{-\infty}^{\infty} |\hat{R}(\xi+\pi m\alpha)| \cdot |\hat{g}(\xi-\pi m\alpha)|^2 d\xi \\
&\leq C'\int_{-\infty}^{\infty} \frac{d\xi}{(1+|\xi+\pi m\alpha|)^a(1+|\xi-\pi m\alpha|)^{2b}} \\
&\leq \frac{C'}{(\pi m\alpha)^{2b}}\int_{-\infty}^{0} \frac{d\xi}{(1+|\xi+\pi m\alpha|)^a} + \frac{C'}{(\pi m\alpha)^a}\int_{0}^{\infty} \frac{d\xi}{(1+|\xi-\pi m\alpha|)^{2b}} \\
&\leq \frac{C'}{(\pi m\alpha)^{2b}}\int_{-\infty}^{\infty} \frac{d\xi}{(1+|\xi|)^a} + \frac{C'}{(\pi m\alpha)^a}\int_{-\infty}^{\infty} \frac{d\xi}{(1+|\xi|)^{2b}} \leq \frac{C}{m^r}
\end{aligned} \tag{215}$$

where $r = \min(a, 2b)$ and an estimate of C is:

$$C = 2C_1C_2\sqrt{2\pi}\left(\frac{1}{(a-1)(\pi\alpha)^{2b}} + \frac{1}{(2b-1)(\pi\alpha)^a}\right) \sim \frac{C_0}{\alpha^r}$$

Next we estimate M_Δ such that for $|m| > M_\Delta$, $\sigma_{mn}^2 < \frac{\Delta^2}{12}$. Using (214) we obtain for M_Δ the following estimate:

$$M_\Delta = \frac{(12C)^{1/r}}{\Delta^{2/r}} \tag{216}$$

Therefore we have to encode at most $2M_\Delta + 1$ coefficients. This gives the following estimate for the quantization error J_q (see (211):

$$J_q \leq 2B^{2,\infty}\frac{\Delta^2}{4}(2M_\Delta + 1) \approx C_q\Delta^{2(1-\frac{1}{r})} \tag{217}$$

with an estimate of C_q given by $C_q = (12C)^{1/r}B^{2,\infty}$.

For the truncation error we use the following estimate (see (213)):

$$J_\varepsilon \leq 2\sqrt{2\pi}B^{2,\infty}\sum_{|m|\geq M_\Delta}\frac{C}{m^r} \approx \tilde{C}''\int_{M_\Delta}^{\infty}\frac{dx}{x^r} = \frac{C''}{M_\Delta^{r-1}}$$

Using now (216) we obtain:

$$J_\varepsilon \leq C_\varepsilon\Delta^{2(1-\frac{1}{r})} \tag{218}$$

with an estimate of C_ε given by $C_\varepsilon = 2C\sqrt{2\pi}B^{2,\infty}/(r-1)$. We notice that J_ε and J_q are both of the same order in Δ. Moreover, for $a > 1$ and $b > \frac{1}{2}$ they both decay to zero as $\Delta \to 0$. Thus by choosing a sufficiently small Δ we can make $J_\varepsilon + J_q < J$. The moral of this computation was to show that asymptotically (i.e. for $\Delta \to 0$), the dominant term in the distortion (209) is given by the stochastic approximation error J which depends on the redundancy proportional to $1 - \frac{1}{\alpha\beta}$ (see (116).

We analyze next an asymptotic approximation of the rate, under the same assumptions as before. We use (208) and again we replace μ_{mn} and ν_{mn} by σ_{mn} and we get:

$$Rate \leq \frac{1}{\beta} \sum_{|m| \leq M_\Delta} \log_2(\frac{12}{\Delta^2}\sigma_{mn}^2) = \frac{2}{\beta} \sum_{1 \leq m \leq M_\Delta} (\log_2 \frac{12C}{\Delta^2} - r \log_2 m)$$

Note that M_Δ has been chosen so that $\log_2 \frac{12C}{\Delta^2} = r \log_2 M_\Delta$. Then, when we approximate the sum by an integral we get:

$$Rate \leq \frac{2r}{\beta} \int_1^{M_\Delta} (\log_2 M_\Delta - \log_2 x) dx \approx \frac{2r}{\beta \ln 2} M_\Delta$$

Thus:

$$Rate \leq \frac{(12C)^{1/r} 2r}{\beta \ln 2} \Delta^{-2/r} \sim \frac{\Delta^{-2/r}}{\alpha\beta} \tag{219}$$

Note the upper bound of the rate goes to ∞ when $\Delta \to 0$, a very natural conclusion since we are going to send more and more coefficients. For a fixed Δ, the rate is (essentialy) proportional to the redundancy $\frac{1}{\alpha\beta}$, whereas, for sufficiently small Δ, the distortion is given by the stochastic part which is proportional to $1 - \frac{1}{\alpha\beta}$. Thus the redundancy $\frac{1}{\alpha\beta}$ parametrizes both the distortion and the rate plots, realizing thus a trade-off between two two quantities in the distortion-rate characteristics.

We end this section by evaluating the rate for the stochastic process studied in Subsection 5.1. We analyze the rate for two encoding windows: the gaussian (202) and the step function (203), both after norming. We took $\Delta = 1.0$. The numerical results are given in the Table 3, where M_Δ^q represents the number of coefficients sent in a $q\beta$ time interval ($\beta = 0.1$ for each channel and $\alpha\beta = \frac{p}{q}$)) counting separately the real and immaginary parts of each c_{mn}, M_Δ is the average on each β time interval, i.e. $M_\Delta = M_\Delta^q / q$, and the rate *Rate* is computed with (208). We computed these values for both the gaussian window and the step function. In Figure 21 we represent the typical distribution of the variance coefficients for the two windows. On the same plot is shown the coefficient variance for different n. Note for gaussian window there is no difference for different n's. For the step function, for some values of m

the variance is different for different n (in particular at $m = 1$). In Figure 22 the rate-redundancy and rate-distortion characteristics are shown. The redundancy is $Red = \frac{q}{p}$. Note the almost linear dependence of the rate to the redundancy. We computed only the stochastic part of the distortion when the decoding window is the partial optimal dual to g. In Figure 22, the distortion is measured as the ratio J/J_{max} where $J_{max} = \mathbf{E}\|f\|^2_{L^2_w} = R(0)\|w\|_{L_1} = 13.4$. These results are for one channel. For two channels, one should multiply the numbers by a factor of two. However the dependence rate-distortion is essentially the same.

6 Conclusions

In this paper we analyzed the multiple representation transmission scheme when windowed Fourier encoders and decoders are used. A windowed Fourier encoder is obtained by computing first the Fourier coefficients of the signal multiplied by a translated window and next by quantizing and encoding these coefficients. The inverse of the product of the modulation parameter and window translation factor represents the redundancy of the encoder. When this number is subunital, the encoder gives only a partial description of the signal, i.e. ignoring the quantization effects, the encoder is not invertible, or, which is the same, any decoder would not be able to perfectly reconstruct the original signal. The windowed Fourier decoder is obtained by an inverse operation (even though, as we mentioned before, it may not be the inverse operator), namely by making a linear combination of the quantized coefficients with translates and modulates of a certain window function. Thus a windowed Fourier encoding-decoding channel is characterized by two distinct windows: the encoding window and the decoding window. Some natural invariance properties meke us to choose respectively the same translation and modulation parameters for the encoder and decoder.

In the multiple description scheme, two channels are used to transmite the signal (see Figure 1). Each channel carries a partial description of the signal, but together they form an (over)complete representation of the original signal. The side decoders have the task to estimate the original signal based on the partial description that each channel carries. When the signal is assumed stationary with a known autocovariance function and the error is measured as a weighted L^2 norm, an exact analysis of the approximation error is possible and this is what we did in this paper.

In Section 2 we introduced several configuration schemes. More specific, the two encoders can be realized a a time-shift or a frequency-shift of the other. This means that one of the windows is, essentially, either a time-shift or a frequency-shift of the other. We call the former scheme a Time-Shift Division Encoder (TSDE), and the latter scheme a Frequency-Shift Division Encoder (FSDE). The same terminology is appliable to the decoder: Time-Shift Division Decoder (TDED) and Frequency-Shift Division Decoder (FSDD). When the division is performed at both the encoder and decoder we have a TSDED or FSDED transmission schemes. These cases are shown in Figures 2–4, for Time-Shift Division.

The one-channel distortion, in the absence of quantization effects, is simply a weighted 2-norm of the approximation error. In Section 3 we analyzed this error by using the Zak transform. In order to do that, we assume the redundancy parameter is rational. This allows us to obtain reduce the problem to a finite dimension matrix algebra problem. The minimization of this criterion (i.e. the purely stochastic distortion) yields an interesting optimization problem. In Section 4 we analyzed all the eight optimization problems and we obtained explicit parametrization in all cases, except the FSDD case. The parametrization is a Karhunen-Loève-type formula': the window is obtained by solving an eigenvalue-eigenvector problem for a self-adjoint matrix of functions. The distortion is obtained by summing and then integrating the lowest eigenvalues of this matrix (the exact formuae are obtained in the Subsection 4.2). In the FSD cases, the optimal solution is shown not to satisfy generically the completeness hypothesis on the encoders. More specific, the encoders are optimal, the two channels do not give an (over)complete description of the signal. Thus, even though each channel would have the lowest distortion, when both channels work, the reconstruction stochastic error (i.e. neglecting the quantization effects) is nonzero. However, a near-optimal solution always exists and satisfis the (over)completeness requirement. A second negative property of the optimizer is established in Subsection 4.6 where a non-localization result (whether for TSD or FSD case) of the optimal window is proved. These non-localization results are of the type of "no-go theorems" of Balian-Low and Heil-Walnut. The case FSDD is still open, though.

A case study is presented in Subsection 5.1. It is shown that well-localized windows can achieve near-optimality. The optimal window, despite of being poorly localized, gives interesting information about certain "frequency bands" that a near-optimal window has to avoid. This suggests an algorithm of designing such windows (and therefore encoding-decoding schemes).

Finally, in Subsection 5.2, we analyzed the total distortion and the rate needed for transmission. Asymptotic formulae with respect to the quantizing interlevel are obtained in that section. The redundancy parameter is shown to trade-off between distortion and rate.

Some problems are left for a further study. We mention here two such open problems. First is the non-localization phenomenon proved in Subsection 4.6. There, we proved this result assuming some extra conditions (either $\mathbf{M}(t)$ is diagonal of $\mathbf{W}(s)$ is Toeplitz). Thus we naturally ask whether thsi result holds in the general case. The second problem concerns the near-optimal case. A better solution to the designing problem would be to minimize a criterion containing not only the approximation error measure but also a time-frequency localization norm of the window (for instance some mixed Sobolev norm).

Appendices

A Known Results on Weyl-Heisenberg Frames and Riesz Bases

The abstract concept of *frame* has been introduced in the seminal paper [DuSch52] by R.J.Duffin and A.C.Schaeffer. Consider H a Hilbert space and $\mathbf{I}$ a countable index set. Then:

DEFINITION 46 *A set of vectors of H, $\mathcal{F} = \{f_i\}_{i\in\mathbf{I}}$ is caled a* frame *(for H) if there are constants $A, B > 0$ such that for every $f \in H$ we have:*

$$A\|f\|^2 \le \sum_{i\in\mathbf{I}} |<f, f_i>|^2 \le B\|f\|^2 \tag{220}$$

The constants A, B are called *frame bounds*. If we can choose $A = B$ the frame is called *tight*. Note that (220) immediately implies $\mathcal{F}$ is a complete set in H (i.e. the set of finite linear combinations of f_i's is dense in H). Indeed if this is not so, there would exist a nontrivial $f \in H$ orthogonal to all of f_i's. But the first inequality would imply $\|f\| = 0$, leading to a contradiction.

The Riesz basis for its span concept generalizes the notion of an orthonormal set:

DEFINITION 47 *A set of vectors of H, $\mathcal{F} = \{f_i\}_{i\in\mathbf{I}}$ is called a* Riesz basis for its span *(or a* s-Riesz basis*) if there are constants $A, B > 0$ such that for every finite sequence of complex numbers $c = \{c_i\}_{i\in\mathbf{I}}$ we have:*

$$A\sum_{i\in\mathbf{I}} |c_i|^2 \le \|\sum_{i\in\mathbf{I}} c_i f_i\|^2 \le B\sum_{i\in\mathbf{I}} |c_i|^2 \tag{221}$$

The constants A, B are called *s-Riesz basis bounds*. If we can choose $A = B$, the s-Riesz basis is an *orthogonal equi-norm set*. The *span* of $\mathcal{F}$ is defined as the closure of its linear span (i.e. of the finite linear combinations of f_i's). If the span is H, we simply call $\mathcal{F}$ a *Riesz basis*. Note that (221) implies a l^2-form of linear independence of f_i's. Moreover, the restriction that c be a finite sequence can be dropped; we can let c run through $l^2(\mathbf{I})$.

We introduce now several operators associated to frames and s-Riesz bases:

$$\text{the } \textit{analysis operator} \ \ T : H \to l^2(\mathbf{I}) \ \ , \ \ T(f) = \{<f, f_i>\}_{i\in\mathbf{I}} \tag{222}$$

$$\text{the } \textit{synthesis operator} \ \ T^* : l^2(\mathbf{I}) \to H \ \ , \ \ T^*(c) = \sum_{i\in\mathbf{I}} c_i f_i \tag{223}$$

$$\text{the } \textit{frame operator} \ \ S : H \to H \ \ , \ \ S = T^*T \ \ , \ \ S(f) = \sum_{i\in\mathbf{I}} <f, f_i> f_i \tag{224}$$

$$\text{the } \textit{grammian operator} \ \ G : l^2(\mathbf{I}) \to l^2(\mathbf{I}) \ \ , \ \ G = TT^* \ \ , \ \ G(c) = \{<\sum_{j\in\mathbf{I}} c_j f_j, f_i>\}_{i\in\mathbf{I}} \tag{225}$$

The right inequalities in (220) and (221) show that the analysis and synthesis operators are bounded and well-defined on their definition domains. Moreover, notice they are dual to one another (hence the $*$ notation). Thus the frame and grammian operators are well defined as well. The frame condition (220) is equivalent to $A \leq S \leq B$, where the inequalities are in the sense of quadratic forms. Similarly, (221) is equivalent to $A \leq G \leq B$ in the same sense. Note that $A > 0$ implies S, respectively G, is invertible. Then let us define the following vectors. For a frame $\mathcal{F}$:

$$\tilde{f}_i = S^{-1} f_i \quad , \quad i \in \mathbf{I} \tag{226}$$

and respectively, for a s-Riesz basis $\mathcal{F}$:

$$f_i^{\#} = T^* G^{-1} \delta_i \tag{227}$$

(where δ_i is the canonical basis of $l^2(\mathbf{I})$, $(\delta_i)_j = \delta_{ij}$ the Kronecker symbol). Straightforward computations show the following result:

PROPOSITION 48 *A. Suppose $\mathcal{F}$ is a frame for H and consider the vectors (226). Then:*

1. The set $\tilde{\mathcal{F}} = \{\tilde{f}_i, i \in \mathbf{I}\}$ is a frame for H with bounds $\frac{1}{B}, \frac{1}{A}$;

2. The synthesis operator associated to $\tilde{\mathcal{F}}$ is a left inverse of the analysis operator associated to $\mathcal{F}$. Similarly, the synthesis operator associated to $\mathcal{F}$ is a left inverse of the analysis operator $\tilde{T}$ associated to $\tilde{\mathcal{F}}$. Explicitly this means for every $f \in H$,

$$\sum_{i \in \mathbf{I}} < f, f_i > \tilde{f}_i = \sum_{i \in \mathbf{I}} < f, \tilde{f}_i > f_i = f \tag{228}$$

3. Let E denote the range of T in $l^2(\mathbf{I})$. Then:

$$P_E := T\tilde{T}^* \tag{229}$$

is the orthogonal projection onto E. Thus $P_E = \tilde{T}T^*$ and $\tilde{T}$ has the same range E as T.

B. Suppose $\mathcal{F}$ is a s-Riesz basis in H and consider the vectors $f_i^{\#}$ defined by (227).

1. The set $\mathcal{F}^{\#} = \{f_i^{\#}, i \in \mathbf{I}\}$ is a s-Riesz basis in H with bounds $\frac{1}{B}, \frac{1}{A}$ and with the same span $\mathcal{E}$ as $\mathcal{F}$ in H.

2. The analysis operator $T^{\#}$ associated to $\mathcal{F}^{\#}$ is a left inverse of the synthesis operator T^* of $\mathcal{F}$, and the analysis operator T is a left inverse of the synthesis operator $T^{\#*}$ associated to $\mathcal{F}^{\#}$. Explicitly, for every $c = (c_i)_{i \in \mathbf{I}} \in l^2(\mathbf{I})$,

$$< \sum_{j \in \mathbf{I}} c_j f_j, \tilde{f}_i > = < \sum_{j \in \mathbf{I}} c_j \tilde{f}_j, f_i > = c_i \tag{230}$$

Thus the following biorthogonality relations hold:

$$< f_i, \tilde{f}_j >= \delta_{ij} \tag{231}$$

3. The operator:

$$P_{\mathcal{E}} = T^* T^{\#} \tag{232}$$

is the orthogonal projection onto the span $\mathcal{E}$ of $\mathcal{F}$. Hence the following identity holds true as well:

$$P_{\mathcal{E}} = T^{\#*} T \tag{233}$$

4. $\mathcal{F}$ is a frame for $\mathcal{E}$. Moreover the formula (226) gives the same vectors as (227), i.e.

$$(S|_{\mathcal{E}})^{-1} f_i = T^* G^{-1} \delta_i \tag{234}$$

The set $\tilde{\mathcal{F}}$ defined by (228) is called the *standard dual frame* (associated to $\mathcal{F}$). The remarkable property (228) represents a *discrete resolution of identity*, called also a *reconstruction formula*. Note that (228) does not uniquely define $\tilde{\mathcal{F}}$. In other words, there may exist many *dual* frames which yield reconstruction formulae as (228). In general there are infinity many such duals, unless $\mathcal{F}$ is a Riesz basis, in which case the dual is unique. Each such alternate dual frame gives an oblique (i.e. non-orthogonal) projection onto E via (229).

The set $\mathcal{F}^{\#}$ defined by (227) is called the *standard biorthogonal s-Riesz basis* (associated to $\mathcal{F}$). (230) represents a reconstruction formula in the space of coefficients and follows immediately from the biorthogonality relations (231). Note that in general there are many alternate biorthogonal s-Riesz bases that satisfy (231). Each of them will yield a reconstruction formula of type (230), although (232) will give only an oblique projection onto $\mathcal{E}$ and therefore (233) is no longer true. These alternate biorthogonals will have different spans.

Weyl-Heisenberg frames and s-Riesz bases are simply WH sets that are frames or s-Riesz bases, according to the previous definitions. Note that the definitions in Definition 1 are simply particular instances of Definitions 46 and 47, applied to the WH context.

WH sets enjoy the remarkable property (which wavelet sets do not have, for instance) that the standard dual frame or the standard biorthogonal s-Riesz basis is also a WH set. This follows from the commutation relations:

$$ST_\beta = T_\beta S \quad , \quad SM_{2\pi\alpha} = M_{2\pi\alpha} S \tag{235}$$

Therefore, in the WH frame case, if $\tilde{g} = S^{-1}g$, the WH set $\mathcal{WH}_{\tilde{g};\alpha,\beta}$ is the standard dual frame, whereas in the WH s-Riesz basis case, if $g^{\#} = T^*G^{-1}\delta_{(0,0)} = (S|_{\overline{span\mathcal{WH}_{g;\alpha,\beta}}})^{-1}g$ then $\mathcal{WH}_{g^{\#};\alpha,\beta}$ is the standard biorthogonal s-Riesz basis.

The window $\tilde{g}$ is called the *standard dual frame window*, whereas $g^{\#}$ is called the *standard biorthogonal window*. Note that in general there are many dual frame generators (respectively biorthogonal windows) that give rise to alternate WH dual frames (respectively alternate WH biorthogonal s-Riesz bases). The only case when the dual (or biorthogonal) is unique is when $\mathcal{WH}_{g;\alpha,\beta}$ is a Riesz basis, in which case the standard dual frame is also the standard biorthogonal s-Riesz basis.

Beside this dual/biorthogonal construction, there are also results dealing with the density of the lattice $\{(m\alpha, n\beta); m, n \in \mathbf{Z}\}$ and the localization of the generator g in the basis case. We summarize these results in the following theorem. The interested reader may find the density results proved for the lattice case in [Rief81], [DaLaLa96] or [Jans95], and for nonuniform sets in [RaSte95]. The localization theorem due in its original form to Balian ([Balian81]) and Low ([Low85]) has been rigorously proved by Coifman and Semmes (see [Daub90]) and differently (and much more simply) by Battle in [Batt88]. Later on it was extended to a different space of functions by Heil and Walnut (see [BeHeWa95]). Statements of these results are summarized in the theorem below. The following spaces of functions turn out to be very useful in the statement:

$$Q^{1,1} = \{f \in L^2 | \int x^2|f(x)|^2dx < \infty \ , \ \int \xi^2|\hat{f}(\xi)|^2d\xi < \infty\} \tag{236}$$

$$C(L^\infty, l^1) = \{f \mid \|f\|_{W(L^\infty,l^1)} := \sum_{n\in\mathbf{Z}} \|f \cdot 1_{[n,n+1]}\|_\infty < \infty\} \tag{237}$$

THEOREM 49 *Consider $\mathcal{WH}_{g;\alpha,\beta}$ a WH Riesz basis. Then:*

1. The lattice $\{(m\alpha, n\beta); m, n \in \mathbf{Z}\}$ has uniform density 1, i.e. $\alpha = \beta = 1$. Moreover, if $\mathcal{WH}_{g;\alpha,\beta}$ is a frame then $\alpha\beta \leq 1$, whereas if $\mathcal{WH}_{g;\alpha,\beta}$ is a s-Riesz basis then $\alpha\beta \geq 1$.

2.1 (Balian-Low) The generator g has an infinite uncertainty product, i.e. $g \notin Q^{1,1}$.

2.2 (Heil-Walnut) The generator is spread in time-frequency domain, i.e. $g \notin C(L^\infty, l^1)$ and $\hat{g} \notin C(L^\infty, l^1)$.

□

Moreover, when we have a union of WH sets, the following density and nonlocalization results apply:

THEOREM 50 *Suppose $\mathcal{WH}_{g^1;\alpha,\beta} \cup \mathcal{WH}_{g^2;\alpha,\beta}$ is a Riesz basis of $L^2(\mathbf{R})$. Then:*

1.(see [ChDeHe99]) $\alpha\beta = \frac{1}{2}$;

2.(see [ZiZe97]) Either $g^1 \notin Q^{1,1}$, *or* $g^2 \notin Q^{1,1}$;

3.(see the proof of the previous result and the Heil-Walnut's proof in [BeHeWa95]) Either $g^1 \notin C(L^\infty, l^1)$ *and* $\hat{g^1} \notin C(L^\infty, l^1)$, *or* $g^2 \notin C(L^\infty, l^1)$ *and* $\hat{g^2} \notin C(L^\infty, l^1)$.

B Approximation of Stochastic Processes by Weyl-Heisenberg Sets

Consider two WH sets $\mathcal{WH}_{g^1;\alpha,\beta}$, $\mathcal{WH}_{g^2;\alpha,\beta}$ and a stationary stochastic signal $\mathbf{f}$ of zero-mean and second-order statistics (i.e. autocovariance) given by $R(t) = \mathbf{E}[\mathbf{f}(\cdot)\overline{\mathbf{f}(\cdot - t)}]$. Assume the analysis operators of the two WH sets are bounded on $L^2(\mathbf{R})$ (this means they are *Bessel sequences*). For such pairs of WH sets we define the *frame operator* denoted $S_{g^1,g^2;\alpha,\beta}$ by:

$$S_{g^1,g^2;\alpha,\beta}(f) = \sum_{m,n} < f, g^1_{mn;\alpha,\beta} > g^2_{mn;\alpha,\beta} \tag{238}$$

Note that $S_{g^1,g^2;\alpha,\beta}$ is bounded on $L^2(\mathbf{R})$ and its norm is bounded by the product of the two analysis operator norms $||T_{g^1;\alpha,\beta}|| \cdot ||T_{g^2;\alpha,\beta}||$. The problem is to give sense to and study the boundedness of this frame operator when applied to the stochastic signal $\mathbf{f}$.

First we introduce a couple of function spaces that are useful in the following. the general *Wiener amalgam space* is defined for by

$$W(L^p, l^q) = \{f \mid ||f||_{W(L^p,l^q)} := ||\{||f \cdot 1_{[n,n+1]}||_p\}_{n\in\mathbf{Z}}||_q < \infty\} \tag{239}$$

where the (p, q) norm for $p, q \neq \infty$ is defined as

$$||\{||f \cdot 1_{[n,n+1]}||_p\}_{n\in\mathbf{Z}}||_q = (\sum_{n\in\mathbf{Z}} (\int_n^{n+1} |f(x)|^p dx)^{\frac{q}{p}})^{\frac{1}{1}q}. \tag{240}$$

For $p = \infty$ or $q = \infty$ the definitions have to be adapted in the obvious way. For instance $W(L^\infty, l^1)$ has been defined earlier in (237). In particular we are concerned with $W(L^2, l^\infty), W(L^\infty, l^1)$ and $W(L^1, l^\infty)$. For more properties of these spaces we refer the reader to [FoSt85]. Note that these spaces are translation and dilation invariant. In particular the space does not change (although the norm does) if instead of translation step 1 we consider the translation step β.

Another useful space is the *weighted* L^2_w space for some nonnegative function $w \geq 0$:

$$L^2_w = \{f \mid ||f||_w := (\int_{-\infty}^{\infty} w(x)|f(x)|^2 dx)^{1/2} < \infty \tag{241}$$

With this weight w we construct also a weighted version of the Wiener amalgam space:

$$W_\beta(L^2_w, l^\infty) = \{f \mid \|f\|_{W_\beta(L^2_w,l^\infty)} := \sup_{n\in\mathbf{Z}} \int_{-\infty}^{\infty} w(x)|f(x-n\beta)|^2 dx \tag{242}$$

L^2_w and $W_\beta(L^2_w, l^\infty)$ are in general no longer translation invariant, but for appropriate w they may be (see below). Finally, we introduce also the following notion:

DEFINITION 51 *A function $f : \mathbf{R} \to \mathbf{C}$ has the* persistency length a *if there is a $\delta > 0$ and a compact set K congruent to $[0, a] \bmod a$ such that for every $x \in K$, $|f(x)| \geq \delta$.*

The following results are proved in [Balan98].

THEOREM 52 *Suppose $g^1, g^2 \in W(L^\infty, l^1)$.*

a) Let $f \in W(L^2, l^\infty)$ and $\alpha, \beta > 0$. Then $\sum_{m,n} < f, g^1_{mn} > g^2_{mn}$ converges unconditionally in the L^2_{loc} topology, i.e. for every $\varepsilon > 0$ and compact set K there are $N_\varepsilon, M_\varepsilon > 0$ such that for every finite set $S \subset \mathbf{Z}^2 \setminus ([-M_\varepsilon, M_\varepsilon] \times [-N_\varepsilon, N_\varepsilon])$,

$$\| \sum_{(m,n)\in S} < f, g^1_{mn} > g^2_{mn} \|_{L^2(\mathbf{R})} < \varepsilon; \tag{243}$$

moreover (243) converges also in the weak-$*$ topology of $W(L^2, l^\infty)$, i.e. for every $h \in W(L^2, l^1)$ and $\varepsilon > 0$ there are $M_\varepsilon, N_\varepsilon > 0$ such that for every $N > N_\varepsilon$, $M > M_\varepsilon$,

$$| < h, f - \sum_{|m|\leq M} \sum_{|n|\leq N} < f, g^1_{mn} > g^2_{mn} > | < \varepsilon$$

b) For every $\alpha, \beta > 0$ there is some constant $C = C(g^1, g^2; \alpha, \beta)$ such that for every $f \in W(L^2, l^\infty)$ the function defined by (238) is in $W(L^2, l^\infty)$ and $\|S_{g^1,g^2;\alpha,\beta} f\|_{W(L^2,l^\infty)} \leq C\|f\|_{W(L^2,l^\infty)}$. Therefore $S_{g^1,g^2;\alpha,\beta}$ is a well-defined and bounded operator on $W(L^2, l^\infty)$. Moreover, the constant C can be chosen as $C(g^1, g^2; \alpha, \beta) = C_{\alpha,\beta}\|g^1\|_{W(L^\infty,l^1)} \cdot \|g^2\|_{W(L^\infty,l^1)}$.

REMARK 53 *It is well-known (see [HeWa89]) that if $g^1, g^2 \in W(L^\infty, l^1)$ then g^1 and g^2 are Bessel sequence generators; therefore $S_{g^1,g^2;\alpha,\beta}$ is well-defined and bounded on $L^2(\mathbf{R})$. However, in general even if $S_{g^1,g^2;\alpha,\beta}$ is well-defined and bounded on $L^2(\mathbf{R})$, it does not need to be bounded on $W(L^2, l^\infty)$.*

REMARK 54 *Similar results has been proved in [FeiGrö97] but under a stronger requirement, namely the generators to belong to the Segal algebra S_0 which is a subspace of $W(L^\infty, l^1)$ (for exact definitions see [FeiGrö97]).*

REMARK 55 *The series (238) that locally defines $S_{g^1,g^2;\alpha,\beta}$ is not strongly convergent in $W(L^2,l^\infty)$-norm in general.*

The condition $g^1,g^2 \in W(L^\infty,l^1)$ in Theorem 52 is not necessary for the boundedness of $S_{g^1,g^2;\alpha,\beta}$ on $W(L^2,l^\infty)$ (an example is given in [Balan98]). However the following result shows that it is a necessary condition for $g^1 \in W(L^\infty,l^2)$:

THEOREM 56 *Let $(g^1,g^2;\alpha,\beta)$ be the given data and suppose the following:*

1. *For every $f \in W(L^2,l^\infty)$, the series $\sum_{m,n} < f,g^1_{mn} > g^2_{mn}$ converges unconditionally in L^2_{loc};*
2. *The frame operator is bounded on $W(L^2,l^\infty)$;*
3. *g^2 has persistency length $\frac{1}{\alpha}$.*

Then $g^1 \in W(L^\infty,l^2)$. □

The above results refer specifically to $W(L^2,l^\infty)$. We are interested to measure the error in L^2_w though. The transition toward this space is given by the following result:

PROPOSITION 57 *Suppose the nonnegative weight w has persistency length β and $w \in W(L^\infty,l^1)$. Then the norm $\|\cdot\|_{W(L^2,l^\infty)}$ is equivalent to $\|\cdot\|_{W_\beta(L^2_w,l^\infty)}$ and thus the two Banach spaces are identical: $W(L^2,l^\infty) = W_\beta(L^2_w,l^\infty)$.* □

Note that by equivalence we mean there are constants $A,B>0$ such that for every f,

$$A\|f\|_{W(L^2,l^\infty)} \leq \|f\|_{W_\beta(L^2_w,l^\infty)} \leq B\|f\|_{W(L^2,l^\infty)}$$

All these results show that if $g^1,g^2 \in W(L^\infty,l^1)$ and the weight w has persistency length β and belongs to the space $W(L^\infty,l^1)$, then $S_{g^1,g^2;\alpha,\beta}$ is well-defined and bounded on $W_\beta(L^2_w,l^\infty)$.

The picture is now the following. We would like to work with $f \in L^2_w$ because $f \in L^2(\mathbf{R})$ is not possible for stationary signals. However, extending $S_{g^1,g^2;\alpha,\beta}$ to L^2_w is tricky because L^2_w is not well-adapted to the study of translations (in general it is not even a normed space). Therefore we introduce $W(L^2,l^\infty) = W_\beta(L^2_w,l^\infty)$ under some mild conditions on the weight. On this space $S_{g^1,g^2;\alpha,\beta}$ is well-defined and bounded, provided the conditions of Theorem 52 are satisfied. Now, our stochastic signal is given by an element $\mathbf{f}$ of $L^2(\Omega;W(L^2,l^\infty))$ where (Ω,Σ,μ) is a probability space and $L^2(\Omega;W(L^2,l^\infty))$ is the space of $W(L^2,l^\infty)$-valued functions on Ω that are square integrable with respect to the probability measure μ. Thus $\omega \in \Omega \mapsto \mathbf{f}_\omega = \mathbf{f}(\omega) \in W(L^2,l^\infty)$ is a realization of this stochastic process. The goal is to approximate $\mathbf{f}$ by the coherent stochastic signal $S_{g^1,g^2;\alpha,\beta}\mathbf{f}$. We still measure the approximation error in L^2_w:

$$J(g^1,g^2;\alpha,\beta)=\mathbf{E}\|\mathbf{f}-S_{g^1,g^2;\alpha,\beta}\mathbf{f}\|^2_w$$

which is finite and bounded as follows:

$$\begin{aligned}J(g^1,g^2;\alpha,\beta)\leq\int_\Omega d\mu(\omega)\ \sup_n\|T^n_\beta(\mathbf{1}-S_{g^1,g^2;\alpha,\beta})\mathbf{f}_\omega\|^2_w\\ \leq(1+\|S_{g^1,g^2;\alpha,\beta}\|_{B(W_\beta(L^2_w,l^\infty))})\|\mathbf{f}\|_{L^2(\Omega;W_\beta(L^2_w,l^\infty))}\end{aligned}$$

because $S_{g^1,g^2;\alpha,\beta}$ commutes with the translation of β. Note that

$$\|S_{g^1,g^2;\alpha,\beta}\|_{B(W_\beta(L^2_w,l^\infty))}\leq C_{\alpha,\beta,w}\|g^1\|_{W(L^\infty,l^1)}\|g^2\|_{W(L^\infty,l^1)}$$

which turns the previous relation into:

$$J(g^1,g^2;\alpha,\beta)\leq(1+C_{\alpha,\beta,w}\|g^1\|_{W(L^\infty,l^1)}\|g^2\|_{W(L^\infty,l^1)})\|\mathbf{f}\|_{L^2(\Omega;W(L^2,l^\infty))}\tag{244}$$

All the above are summarized by the following theorem:

THEOREM 58 *Suppose $g^1,g^2\in W(L^\infty,l^1)$ and the nonnegative weight w has persistency length β and $w\in W(L^\infty,l^1)$. Then for every stochastic signal $\mathbf{f}\in L^2(\Omega;W(L^2,l^\infty))$ the approximation error given by the WH pair $(g^1,g^2;\alpha,\beta)$ is bounded above as in (244).* $\square$

In the asymptotic analysis of the distortion we shall need the following result (whose proof is in [Balan98]):

LEMMA 59 *Suppose $g,w\in W(L^\infty,l^1)$. Then $T^*_g:l^{2,\infty}\to L^2_w$ defined by $T^*_g(c)=\sum_{m,n\in\mathbf{Z}}c_{mn}g_{mn}$ is well defined and bounded by:*

$$\begin{aligned}\|T^*_g\|^2_{B(l^{2,\infty},L^2_w)}&\leq\frac{1}{\alpha}\sum_n\|\sum_k w(\cdot+\frac{k}{\alpha})|g(\cdot+\frac{k}{\alpha}-n\beta)|^2\|_{L^\infty(0,\frac{1}{\alpha})}\\&\leq C_{\alpha,\beta}\|w\|_{W(L^\infty,l^1)}\|g\|_{W(L^\infty,l^1)}\|g\|_\infty\end{aligned}\tag{245}$$

C Note on the Poisson Summation Formula

The computations made in subsection 3.1 used a special form of the Poisson summation formula. Actually, despite its similarity, the formula we use is in fact the Parseval identity. We call it the *weak form of* *the Poisson Summation Formula.* It has been proved and used by many authors before (see the proof of Theorem 4.1.5 in [HeWa89], or Theorem 2 in [ChuiShi93], or Lemma 3.2 in [DaLaLa96]). Below we prove it for a different set of functions:

LEMMA 60 (Weak Poisson Summation Formula) *Suppose $f_1, f_2 \in W(L^2, l^\infty)$ and $g^1, g^2 \in W(L^\infty, l^1)$. Then:*

$$\sum_m \int\int dx\, dy\, f_1(x) g^1(x) f_2(y) g^2(y) e^{2\pi i m\alpha(x-y)} = \frac{1}{\alpha}\sum_m \int dx\, f_1(x) g^1(x) f_2(x+\frac{m}{\alpha}) g^2(x+\frac{m}{\alpha}) \quad (246)$$

and the integrals converge absolutely.

Proof

The "trick" is to periodize each integral in the left hand side. Notice $f_1 g^1, f_2 g^2 \in L^1(\mathbf{R}) \cap L^2(\mathbf{R})$. Then:

$$c_1(m) := \int dx\, f_1(x) g^1(x) e^{-2\pi i m\alpha x} = \int_0^{\frac{1}{\alpha}} dx [\sum_{l\in\mathbf{Z}} f_1(x+\frac{l}{\alpha}) g^1(x+\frac{l}{\alpha})] e^{-2\pi i m\alpha x}$$

and the $\frac{2\pi}{\alpha}$-periodic function $x \mapsto \sum_{l\in\mathbf{Z}} f_1(x+\frac{l}{\alpha}) g^1(x+\frac{l}{\alpha})$ is in $L^1[0, \frac{2\pi}{\alpha}] \cap L^2[0, \frac{2\pi}{\alpha}]$ because of the following:

$$\int_0^{\frac{1}{\alpha}} |\sum_{l\in\mathbf{Z}} f_1(x+\frac{l}{\alpha}) g^1(x+\frac{l}{\alpha})| dx \le \sum_{l\in\mathbf{Z}} \int_0^{\frac{1}{\alpha}} |f_1(x+\frac{l}{\alpha})| \cdot |g^1(x+\frac{l}{\alpha})| dx = \int_{-\infty}^{\infty} |f_1(x)| \cdot |g^1(x)| dx = \|f_1 g^1\|_{L^1}$$

$$\int_0^{\frac{1}{\alpha}} |\sum_{l\in\mathbf{Z}} f_1(x+\frac{l}{\alpha}) g^1(x+\frac{l}{\alpha})|^2 dx \le [\sum_l \int_0^{\frac{1}{\alpha}} |f_1(x+\frac{l}{\alpha})|^2 \cdot |g^1(x+\frac{l}{\alpha})| dx] \cdot [\sum_l \int_0^{\frac{1}{\alpha}} |g^1(x+\frac{l}{\alpha})| dx] =$$

$$= \int_{-\infty}^{\infty} |f_1(x)|^2 \cdot |g^1(x)| dx \cdot \|g^1\|_{L^1} \le \|f_1\|^2_{W(L^2,l^\infty)} \|g^1\|^2_{W(L^\infty,l^1)}$$

We denote by $c_2(m)$ a similar expression as for $c_1(m)$ where the product $f_1 g^1$ is replaced by $f_2 g^2$. Thus, using the Parseval identity, the left hand side in (246) becomes:

$$\sum_m c_1(m) c_2(m) = \frac{1}{\alpha} \int_0^{\frac{1}{\alpha}} [\sum_{l\in\mathbf{Z}} f_1(x+\frac{l}{\alpha}) g^1(x+\frac{l}{\alpha})] \cdot [\sum_{l\in\mathbf{Z}} f_2(x+\frac{l}{\alpha}) g^2(x+\frac{l}{\alpha})] dx$$

Now commuting all the summation symbols with the integral (allowed because of the absolute convergence of the integrals) and extending again the integral over the entire real line, we obtain the right hand side in (246) and thus the conclusion. □

This lemma has the following corollary:

COROLLARY 61 *Suppose* $\mathbf{f} \in L^2(\Omega; W(L^2, l^\infty))$ *is a stationary stochastic signal of autocovariance function* R*, and* $g^1, g^2 \in W(L^\infty, l^1)$*, then:*

$$\sum_m \int\int dx\, dy\, w(x) R(x-y) e^{2\pi i m\alpha(y-x)} g^1(y-n\beta)\overline{g^2(x-n\beta)} =$$

$$= \frac{1}{\alpha}\sum_m R(\frac{m}{\alpha}) \int dx\, w(x) g^1(x-n\beta-\frac{m}{\alpha})\overline{g^2(x-n\beta)} \quad (247)$$

□

D Proof of Theorems 3 and 5

Proof of Theorem 3

The idea is to consider the frame operator associated to WH multisets $\mathcal{WH}_{(g^1,g^2);\alpha,\beta}$ and $\mathcal{WH}_{(\tilde{g^1},\tilde{g^2});\alpha,\beta}$, $S_{1,\tilde{1}} + S_{2,\tilde{2}}$. For the first two cases we set $g^1 = g, g^2 = 0$, $\tilde{g^1} = g^\#$ or $\tilde{g}$, $\tilde{g^2} = 0$. In the first case $S_{1,\tilde{1}}$ should be the orthogonal projector onto the range of $S_{1,1}$ (which is the span of the s-Riesz basis); in the second case $S_{1,\tilde{1}} = \mathbf{1}$ and in the third case $S_{1,\tilde{1}} + S_{2,\tilde{2}} = \mathbf{1}$, the unit operator.

These conditions are more transparent if we change the representation. Instead of $L^2(\mathbf{R})$ we use $L^2([0,1]\times[0,\frac{1}{p}];\mathbf{C}^p)$ via the unitary transformation:

$$f \mapsto \Phi(t,s) = [F(t,s)\ F(t,s+\frac{q}{p})\ \ldots\ F(t,s+(p-1)\frac{q}{p})]^*$$

where F is the Zak transform (34) of f and $*$ denotes the hermitian conjugation. Note also that:

$$< g^1, g^2 >= \int_0^{\frac{1}{p}} ds \int_0^{\frac{1}{q}} dt\ trace\{\Gamma^1 \Gamma^{2^*}\} =:< \Gamma^1, \Gamma^2 >_{HS} \quad (248)$$

Then standard computations show that, for every $f_1, f_2 \in L^2(\mathbf{R})$ (or, equivalently, for every $\Phi_1, \Phi_2 \in L^2([0,1]\times[0,\frac{1}{p}];\mathbf{C}^p)$):

$$< f_1, S_{1,\tilde{1}} f_2 >= \sum_{m,n} < f_1, g^1_{m,n;\alpha,\beta} >< g^2_{m,n;\alpha,\beta}, f_2 >= \frac{1}{p}\int_0^{\frac{1}{p}} ds \int_0^1 dt \Phi_1^* \Gamma^2 \Gamma^{1^*} \Phi_2$$

Thus the frame operator $S_{1,\tilde{1}} + S_{2,\tilde{2}}$ acting on $L^2([0,1]\times[0,\frac{1}{p}];\mathbf{C}^p)$ is given simply by the matrix multiplication with $\frac{1}{p}(\tilde{\Gamma^1}\Gamma^{1^*} + \tilde{\Gamma^2}\Gamma^{2^*})$ for every (t,s). This easily implies (57) and (59). Note that if any of (55), (57) or (59) holds true for $t \in [0,\frac{1}{q}]$ then it automatically holds true for every t by the 1-periodicity of the Zak transform in t. Note that a s-Riesz basis can always be viewed as a frame when one restricts oneself to the span of the s-Riesz basis; the upper s-Riesz basis bound is then identical to the upper frame bound of this (restricted) frame. Hence if $\mathcal{WH}_{g;\alpha,\beta}$ is a s-Riesz basis then $B \geq \frac{1}{p}\lambda_{max}(\Gamma\Gamma^*) = \frac{1}{p}\lambda_{max}(\Gamma^*\Gamma) \geq \frac{1}{p}\Gamma^*\Gamma$ where $\lambda_{max}(M)$ is the largest eigenvalue of the matrix M. Thus we obtain the upper bound in (55).

Next we compute the standard biorthogonal generator, respectively the standard dual generator. In the case A1 we impose $\frac{1}{p}\tilde{\Gamma}\Gamma^*$ to be the othogonal projection onto the range of $\frac{1}{p}\Gamma\Gamma^*$. This clearly implies (56).

In the case A2 we require for $\frac{1}{p}\tilde{\Gamma}\Gamma^* = I$. Clearly (58) is a solution. Now any other solution will be given by $\Gamma^{\#} = \tilde{\Gamma} + \Delta\Gamma$ for some $\Delta\Gamma$ with $< \tilde{\Gamma}, \Delta\Gamma >= \int_0^{\frac{1}{p}} ds \int_0^{\frac{1}{q}} dt\, trace\{\tilde{\Gamma}\Delta\Gamma^*\} = 0$. Thus $\|\Gamma^{\#}\|_{HS} \geq \|\tilde{\Gamma}\|_{HS}$ which can be drawn back to g's via (248). Since the standard dual frame generator is the dual frame generator with the smallest norm (see [DaLaLa96]) we obtain that $\Delta\Gamma = 0$ and hence (58).

Similarly, in case B we obtain the equation $\frac{1}{p}(\tilde{\Gamma^1}{\Gamma^1}^* + \tilde{\Gamma^2}{\Gamma^2}^*) = I$ or $\frac{1}{p}\tilde{\Gamma}\Gamma^* = I$ with $\tilde{\Gamma} = [\tilde{\Gamma^1}\ \tilde{\Gamma^2}]$ and $\Gamma = [\Gamma^1\ \Gamma^2]$. The discussion follows the previous case and we thus obtain (60) and (61).

The lower bounds in (55), (57) and (59) are obtained by noting that the standard biorthogonal s-Riesz basis, respectively standard dual frame, have $\frac{1}{A}$ as upper bound. $\Box$.

Proof of Theorem 5.

First a Lemma:

LEMMA 62 *Let W_1, W_2, M_1, M_2 be $p \times p$ nonegative symmetric matrices and Γ^1, Γ^2 be $p \times q$ matrices. Suppose further that either W_1, M_1 are invertible and Γ^1 is of full rank, or W_2, M_2 are invertible and Γ^2 has maximal rank. Then the solution of the following optimization problem:*

$$e^* = \min_{X \in \mathbf{C}^{p\times q}} J(X) := trace\{W_1(I - X\Gamma^{1*})M_1(I - \Gamma^1X^*) + W_2(I - X\Gamma^{2*})M_2(I - \Gamma^2X^*)\} \qquad (249)$$

is unique and given by the solution X_0 of the linear system:

$$W_1M_1\Gamma^1 + W_2M_2\Gamma^2 = W_1X_0\Gamma^{1*}M_1\Gamma^1 + W_2X_0\Gamma^{2*}M_2\Gamma^2 \qquad (250)$$

and the optimum in (249) is:

$$e^* = trace\{W_1(I - X_0\Gamma^{1*})M_1 + W_2(I - X_0\Gamma^{2*})M_2\} \qquad (251)$$

Proof of Lemma

The variation of the criterion (249) due to a variation δX is:

$$\delta J = trace\{-W_1(I - X\Gamma^{1*})M_1\Gamma^1\delta X^* - W_2(I - X\Gamma^{2*})M_2\Gamma^2\delta X^*\} + c.c.$$
$$= -trace\{[W_1(I - X\Gamma^{1*})M_1\Gamma^1 + W_2(I - X\Gamma^{2*})M_2\Gamma^2]\delta X^*\} + c.c.$$

where c.c. stands for *complex conjugated term.*

By the variational principle δJ must vanish for every $\delta X, \delta X^*$. Thus we get (250). We have now to prove that (250) has a unique solution. Consider the linear endomorphism on $\mathbf{C}^{p\times q}$ defined by $X \mapsto T(X) = W_1 X \Gamma^{1*} M_1 \Gamma^1 + W_2 X \Gamma^{2*} M_2 \Gamma^2$. This linear map defines also a quadratic form on the Hilbert space $(\mathbf{C}^{p\times q}, <,>_{HS})$ of $p \times q$ complex matrices with scalar product $< A, B >_{HS} = trace\{A^* B\}$. It is straightforward to see that:

$$< X, T(X) >_{HS} = trace\{W_1^{1/2} X \Gamma^{1*} M_1 \Gamma^1 X^* W_1^{1/2} + W_2^{1/2} X \Gamma^{2*} M_2 \Gamma^2 X^* W_2^{1/2}\} \geq 0$$

The hypothesis guarantees at least one of the two terms is strictly positive. Thus

$$< X, T(X) >_{HS} \geq \alpha < X, X >_{HS}$$

with $\alpha > 0$, and then T is invertible.

The only issue that remains is to check that X_0 defines the minimum for (249). To do this consider $X = X_0 + \Delta$. Then, an easy computation using (249) shows that:

$$J(X) = J(X_0) + trace\{W_1^{1/2} \Delta \Gamma^{1*} M_1 \Gamma^1 \Delta^* W_1^{1/2} + W_2^{1/2} \Delta \Gamma^{2*} M_2 \Gamma^2 \Delta^* W_2^{1/2}\} \geq J(X_0)$$

Thus X_0 is a global minimum and (251) follows. □.

Proof of the Theorem 5

Each of the partial optimal problems is solved by choosing appropriate $W_1, W_2, M_1, M_2, \Gamma^1, \Gamma^2$ in (249) as follows:

B1. $W_1 = \mathbf{W}$, $M_1 = \mathbf{M}$, $\Gamma^1 = \frac{1}{p}\Gamma^1$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$.

B2. $W_1 = \mathbf{M}$, $M_1 = \mathbf{W}$, $\Gamma^1 = \frac{1}{p}\Gamma^{1\#}$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$.

B3. $W_1 = \mathbf{W}$, $M_1 = \mathbf{M}$, $\Gamma^1 = \frac{1}{p}\Gamma^1$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$ (for channel 1),

$W_1 = M_1 = 0$, $\Gamma^1 = 0$, $W_2 = \mathbf{W}$, $M_2 = \mathbf{M}$, $\Gamma^2 = \frac{1}{p}\Gamma^2$ (for channel 2).

B4. $W_1 = \mathbf{M}$, $M_1 = \mathbf{W}$, $\Gamma^1 = \frac{1}{p}\Gamma^{1\#}$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$ (for channel 1),

$W_1 = M_1 = 0$, $\Gamma^1 = 0$, $W_2 = \mathbf{M}$, $M_2 = \mathbf{W}$, $\Gamma^2 = \frac{1}{p}\Gamma^{2\#}$ (for channel 2).

B5. $W_1 = \mathbf{W}$, $M_1 = \mathbf{M}$, $\Gamma^1 = \frac{1}{p}\Gamma$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$ (for channel 1),

$W_1 = M_1 = 0$, $\Gamma^1 = 0$, $W_2 = \mathbf{W}$, $M_2 = \mathbf{M}$, $\Gamma^2 = \frac{1}{p}\Gamma$ (for channel 2).

B6. $W_1 = W_2 = \mathbf{W}$, $M_1 = M_2 = \mathbf{M}$, $\Gamma^1 = \frac{1}{p}\Gamma^{1\#}$, $\Gamma^2 = \frac{1}{p}\Gamma^{2\#}$.

B7. $W_1 = W_2 = \mathbf{W}$, $M_1 = M_2 = \mathbf{M}$, $\Gamma^1 = \frac{1}{p}\Gamma^1$, $\Gamma^2 = \frac{1}{p}\Gamma^2$.

B8. $W_1 = \mathbf{M}$, $M_1 = \mathbf{W}$, $\Gamma^1 = \frac{1}{p}\Gamma^{\#}$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$ (for channel 1),

$W_1 = M_1 = 0$, $\Gamma^1 = 0$, $W_2 = \mathbf{M}$, $M_2 = \mathbf{W}$, $\Gamma^2 = \frac{1}{p}\Gamma^{\#}$ (for channel 2).

B9. $W_1 = \mathbf{W}$, $M_1 = \mathbf{M}$, $\Gamma^1 = \frac{1}{p}\Gamma$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$ (for channel 1),

$W_1 = M_1 = 0$, $\Gamma^1 = 0$, $W_2 = \mathbf{W}$, $M_2 = \mathbf{N}$, $\Gamma^2 = \frac{1}{p}\Gamma$ (for channel 2).

B10. $W_1 = \mathbf{M}$, $W_2 = \mathbf{N}$, $M_1 = M_2 = \mathbf{W}$, $\Gamma^1 = \frac{1}{p}\Gamma^{1\#}$, $\Gamma^2 = \frac{1}{p}\Gamma^{2\#}$.

B11. $W_1 = W_2 = \mathbf{W}$, $M_1 = \mathbf{M}$, $M_2 = \mathbf{N}$, $\Gamma^1 = \frac{1}{p}\Gamma^1$, $\Gamma^2 = \frac{1}{p}\Gamma^2$.

B12. $W_1 = \mathbf{M}$, $M_1 = \mathbf{W}$, $\Gamma^1 = \frac{1}{p}\Gamma^{\#}$, $W_2 = M_2 = 0$, $\Gamma^2 = 0$ (for channel 1),

$W_1 = M_1 = 0$, $\Gamma^1 = 0$, $W_2 = \mathbf{N}$, $M_2 = \mathbf{W}$, $\Gamma^2 = \frac{1}{p}\Gamma^{\#}$ (for channel 2).

B13. $W_1 = W_2 = \mathbf{W}$, $M_1 = M_2 = \mathbf{M}$, $\Gamma^1 = \Gamma^2 = \frac{1}{p}\Gamma$.

B14. $W_1 = W_2 = \mathbf{M}$, $M_1 = M_2 = \mathbf{W}$, $\Gamma^1 = \Gamma^2 = \frac{1}{p}\Gamma^{\#}$.

B15. $W_1 = W_2 = \mathbf{W}$, $M_1 = \mathbf{M}$, $M_2 = \mathbf{N}$, $\Gamma^1 = \Gamma^2 = \frac{1}{p}\Gamma$.

B16. $W_1 = \mathbf{M}$, $W_2 = \mathbf{N}$, $M_1 = M_2 = \mathbf{W}$, $\Gamma^1 = \Gamma^2 = \frac{1}{p}\Gamma^{\#}$.

This ends the proof of the Theorem. $\square$

References

[Ahls85] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inform. Th.*, vol. IT-31, pp. 721–726, November 1985.

[Balan98] R. Balan, *A Study of Weyl-Heisenberg and Wavelet Frames*, Ph.D. thesis, Princeton University 1998.

[Balian81] R. Balian, "Un principe d'incertitude fort en théorie du signal on en mécanique quantique", *C.R.Acad.Sci. Paris* , vol. 292, pp.1357–1362, 1981.

[BaVa94] J.-C. Batllo and V. A. Vaishampayan, "Multiple-description transform codes with an application to packetized speech," in *Proceedings of the 1994 IEEE International Symposium on Information Theory*, Trondhiem, Norway, June 1994.

[Batt88] G. Battle, "Heisenberg Proof of the Balian-Low Theorem", *Letters in Mathematical Physics* , vol. 15, pp. 175–177, 1988.

[BeHeWa95] J. J. Benedetto, C. Heil, D. F. Walnut, "Differentiation and the Balian-Low Theorem", *J.Fourier Anal.Applic.* vol. 1, no.4, pp. 355–402, 1995.

[BerZha83] T. Berger and Z. Zhang, "Minimum breakdown degradation in binary source encoding," *IEEE Trans. Inform. Th.*, vol. IT-29, pp. 807–814, November 1983.

[Bhat87] R. Bhatia, **Perturbation bounds for matrix eigenvalues**, L.G. U.K. 1987.

[BolHla97] H. Bolcskei, F. Hlawatsch, "Oversampled filter banks: Optimal noise shaping, design freedom, and noise analysis", IEEE ICASSP-97, Munich, Germany, vol.3, pp.2453–2456, April 1997

[ChMeWa99] P. A. Chou, S. Mehrotra, A. Wang, "Multiple Description Decoding of Overcomplete Expansions Using Projections on Convex Sets", *Proceedings of the Data Compression Conference*, Snowbird, Utah, pp.72–81, March 29–31, 1999.

[ChDeHe99] O. Christensen, B. Deng, C. Heil , *Density for Gabor Frames*, Appl.Comp.Harm.Anal., vol.7 no.3, pp.292–304, 1999

[ChuiShi93] C. K. Chui, X. Shi, " Inequalities of Littlewood-Paley Type for Frames and Wavelets", *SIAM J.Math.Anal.*, vol. 24, no.1, pp.263–277, 1993.

[Daub90] I. Daubechies, " The Wavelet Transform, Time-Frequency Localizatin and Signal Analysis", *IEEE Trans.Inform.Theory*, vol. 36, no.5, pp.961–1005, 1990.

[DaLaLa96] I. Daubechies, H. J. Landau, Z. Landau, "Gabor Time-Frequency Lattices and the Wexler-Raz Identity", *J.Fourier Anal.Applic.*, vol. 1, no.4, pp.437–478, 1995.

[Davi72] L. D. Davisson, " Rate Distortion Theory and Application", *Proc.IEEE* , vol. 60 , pp.800–808 , 1972, reprinted in *Waveform Quantization and Coding*, N.S.Jayant Ed., IEEE Press, New York 1976.

[DuSch52] R. J. Duffin, A. C. Schaeffer, " A Class of Nonharmonic Fourier Series", *Trans.Amer.Math.Soc.*, vol. 72, , pp.341–366, 1952.

[EquCov91] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Th.*, vol. 37, pp. 269–275, March 1991.

[FeiGrö97] H. G. Feichtinger, K. Gröchenig, " Gabor Frames and Time-Frequency Analysis of Distributions", *J.Func.Anal.*, vol. 146, pp.464–495, 1997.

[FeiStr98] "Gabor Analysis and Algorithms. Theory and Applications", H. .G. Feichtinger, T. Strohmer Eds., Birkhäuser 1998.

[FlEf99] M. Fleming, M. Effros, "Generalized Multiple Description Vector Quantization", *Proceedings of the Data Compression Conference*, Snowbird, Utah, pp.3–12, March 29–31, 1999.

[FoSt85] J. J. F. Fournier, J. Stewart, " Amalgams of L^p and l^q", *Bull.AMS*, vol. 13 , no.1, pp.1–21, 1985.

[ElGaCo82] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Th.*, vol. IT-28, pp. 851–857, November 1982.

[GoKoVe99] V. K. Goyal, J. Kovacevic, M. Vetterli, "Quantized Frame Expansions as Source-Channel Codes for Erasure Channels", *Proceedings of the Data Compression Conference*, Snowbird, Utah, pp.326–335, March 29–31, 1999.

[Gröc96] K. H. Gröchenig, " An uncertainty principle related to the Poisson summation formula", *Stud.Math.*, vol. 121, no.1, pp.87–104, 1996.

[HeWa89] C. Heil, D. Walnut, " Continuous and Discrete Wavelet Transforms", *SIAM Review*, vol. 31, no.4, pp.628–666, 1989.

[HoJo85] R.A.Horn, C.R.Johnson, **Matrix Analysis**, Cambridge Univ. Press 1985.

[IngVai95] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 48–58, January 1995.

[Jans82] A. J. E. M. Janssen, " Bargmann Transform, Zak Transform, and Coherent States", *J.Math.Phys.* , vol. 23, pp.720–731, 1982.

[Jans88] A. J. E. M. Janssen, " The Zak Transform: a Signal Transform for Sampled Time-Continuous Signals", *Philips J.Res.*, vol. 43, pp.23–69, 1988.

[Jans95] A. J. E. M. Janssen, " Duality and Biorthogonality for Weyl-Heisenberg Frames", *J.Fourier Anal.Appl.*, vol. 1, no.4, pp.403–436, 1995.

[JayChr81] N. S. Jayant and S. W. Christensen, "Effects of packet losses in waveform coded speech and improvements due to an odd-even sample interpolation procedure," *IEEE Trans. Commun.*, vol. COM-29, pp. 101–109, February 1981.

[Kato80] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag 2nd Ed. 1980.

[Li95] S. Li, "On General Frame Decompositions", Num.Funct.Anal.and Optim. 16(9&10), pp. 1181-1191, 1995

[Low85] F. Low, " Complete sets of wave packets", *A Passion for Physics - Essays in Honor of Geoffrey Chew*, World Scientific, Singapore, pp.17–22, 1985.

[MoRiLa99] A. E. Mohr, E. A. Riskin, R. E. Ladner, "Graceful Degradation Over Packet Erasure Channels Through Forward Error Correlation",*Proceedings of the Data Compression Conference*, Snowbird, Utah, pp.92–101, March 29–31, 1999.

[Munch92] N. J. Munch, " Noise Reduction in Tight Weyl-Heisenberg Frames", *IEEE Trans.Info.Th.*, vol. 38, no.2, pp.608–616, 1992.

[OrWaVaRe97] M. Orchard, Y. Wang, V. Vaishampayan, and A. Reibman, "Redundancy rate distortion analysis of multiple description coding using pairwise correlating transforms," in *Proceedings of the 1997 International Conference on Image Processing*, Oct. 1997.

[Ozarow80] L. Ozarow, "On a source coding problem with two channels and three receivers," *The Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, December 1980.

[RaSte95] J. Ramanathan, T. Steger, " Incompleteness of Sparse Coherent States", *Appl.Comp.Harm.Anal.* , vol. 2, no.2, pp.148–153, 1995.

[Rief81] M. A. Rieffel, " Von Neumann Algebras Associated with Pairs of Lattices in Lie Groups", *Math.Anal.* , vol. 257, pp.403–418, 1981.

[Rudin74] W. Rudin, **Real and Complex Analysis**, 2nd.Ed. McGraw-Hill, NY, 1974.

[SeVaSl99] S. D. Servetto, V. A. Vaishampayan, N. J. A. Sloan, "Multiple Description Lattice Vector Quantization", *Proceedings of the Data Compression Conference*, Snowbird, Utah, pp.13–22, March 29–31, 1999.

[Srin99] M. Srinivasan, "Iterative Decoding of Multiple Descriptions", *Proceedings of the Data Compression Conference*, Snowbird, Utah, pp.29–31, March 29–31, 1999.

[StSu90] G. W. Stewart, J. G. Sun, **Matrix Perturbation Theory**, Academic Press 1990.

[Vaish93] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[Vaish96] V. A. Vaishampayan, "Application of multiple description codes to image and video transmission over lossy networks," in *Proc. 7th International Workshop on Packet Video*, pp. 55–60, Brisbane, Australia, March 1996.

[VaDo94] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple description scalar quantizers," *IEEE Trans. Inform. Th.*, pp. 245–250, January 1994.

[VaSi95] V. Vaishampayan and A. A. Siddiqui, "Speech predictor design for diversity communication systems," in *Proceedings of the 1995 IEEE Speech Coding Workshop*, 20-22 September 1995.

[VaBa98] V. Vaishampayan and J.-C. Batllo, "Asymptotic Analysis of Multiple Description Quantizers," *IEEE Trans. Inform. Th.*, vol. 44, No. 1, pp. 278–284, Jan. 1998

[Waln92] D. F. Walnut, " Continuity Properties of the Gabor Frame Operator", *J.Math.Anal.Appl.*, vol. 165, no.2, pp.608–616, 1992.

[WiWy81] H. S. Witsenhausen and A. D. Wyner, "Source coding for multiple descriptions II: A binary source," *Bell Syst. Tech. J.*, vol. 60, pp. 2281–2292, December 1981.

[WoWyZi80] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, pp. 1417–1426, October 1980.

[ZhBe87] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Th.*, vol. IT-33, pp. 502–521, July 1987.

[ZiZe95] M. Zibulski, Y. Y. Zeevi, "Analysis of Multi-Window Gabor-Type Schemes by Frame Methods", *CC Pub.*, # 101, Techion Israel Institute of Technology, Haifa, Israel, 1995.

[ZiZe97] M. Zibulski, Y. Y. Zeevi, "Analysis of Multi-Window Gabor-Type Schemes by Frame Methods", Appl.Comp.Harm.Anal., vol.4, pp.188–221, 1997

Captions for Figures and Tables

Figure 1: The Multidescription Transmission Block Diagram.

Figure 2: The Encoding-Decoding Scheme for the Time-Shift Division Encoder (TSDE): g^1, g^2 are derived from the same g.

Figure 3: The Encoding-Decoding Scheme for the Time-Shift Division Decoder (TSDD): $g^{1\#}, g^{2\#}$ are derived from the same $g^{\#}$.

Figure 4: The Encoding-Decoding Scheme for the Time-Shift Division Encoder and Decoder (TS-DED): g^1, g^2 are derived from the same g and $g^{1\#}, g^{2\#}$ are derived from the same $g^{\#}$.

Figure 5: The optimal error in the deterministic model (top plot) and in the generic stochastic model (bottom plot).

Figure 6: The autocovariance function of the stationary process (189) in time domain - left plot - and the spectral power (i.e. frequency domain) - right plot.

Figure 7: The optimal solution for $p = 2$ and $q = 1$.

Figure 8: The optimal solution for $p = 5$ and $q = 1$.

Figure 9: The optimal solution for $p = 3$ and $q = 2$.

Figure 10: The eigenvalue maps for the case $p = 2$ and $q = 1$.

Figure 11: The eigenvalue maps for the case $p = 5$ and $q = 1$.

Figure 12: The eigenvalue maps for the case $p = 3$ and $q = 2$.

Table 1: Numerical results for the partial optimal problem with gaussian window (202).

Figure 13: The gaussian g given by (202) (top) and the partial optimal $g^{\#}$ found in time domain for various choices of p and q: Bottom left: $p = 6, q = 5$; Bottom right: $p = 5, q = 4$.

Figure 14: The partial optimal $g^{\#}$ found in time domain for various choices of p and q: Top left: $p = 4, q = 3$; Top right: $p = 3, q = 2$; Bottom left: $p = 5, q = 3$; Bottom right: $p = 2, q = 1$.

Figure 15: The gaussian g given by (202) (top) and the partial optimal $g^{\#}$ found in frequency domain for various choices of p and q: Bottom left: $p = 6, q = 5$; Bottom right: $p = 5, q = 4$.

Figure 16: The partial optimal $g^{\#}$ found in frequency domain for various choices of p and q: Top left: $p = 4, q = 3$; Top right: $p = 3, q = 2$; Bottom left: $p = 5, q = 3$; Bottom right: $p = 2, q = 1$.

Figure 17: The optimal encoding window for $p = 6$ and $q = 5$.

Figure 18: The optimal encoding window for $p = 7$ and $q = 6$.

Table 2: Numerical results for the partial optimal problem with the characteristic function window

(203).

Figure 19: The encoding window g given in (203) in time (left plot) and frequency (right plot) domains.

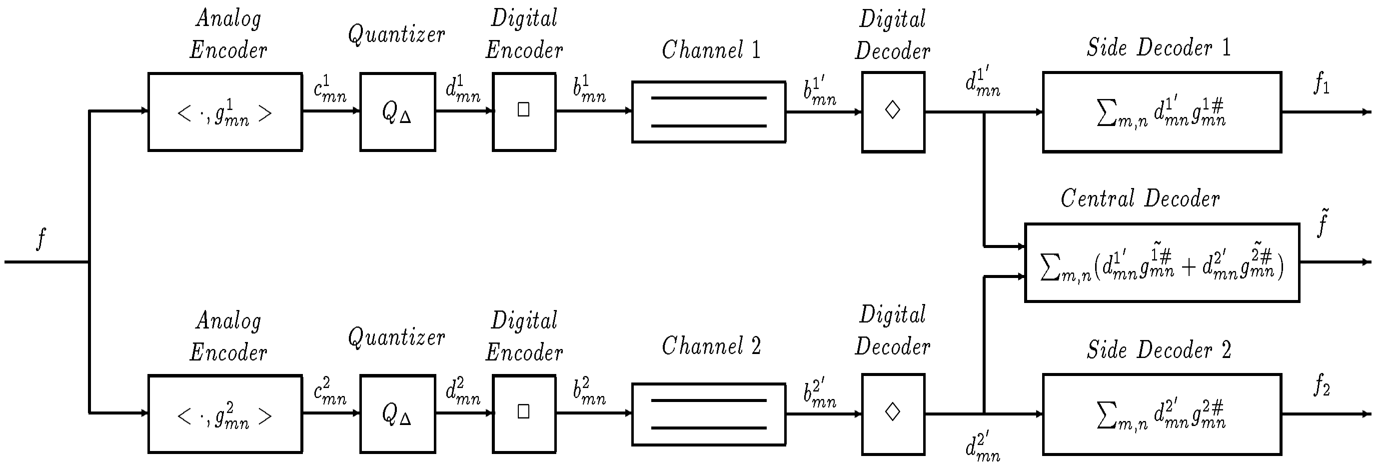
Figure 20: The decoding window $g^{\#}$ obtained as the partial optimal dual to g via (90) for $p = 6$ and $q = 5$ (left plot) and $p = 7$ and $q = 6$ (right plot).

Table 3: Numerical results for the Rate Estimation.

Figure 21: The distribution of the variance of $Re(c_{mn})$ for $p = 4$ and $q = 3$ when g is the gaussian (left plot) or the step function (right plot).

Figure 22: The rate-redundancy (left plot) and rate-distortion (right plot) characteristics; o is for the gaussian and $+$ is for the step function.

List of Figures and Tables



Analog Encoder
$< \cdot, g^1_{mn} >$
c^1_{mn}
Quantizer
Q_Δ
d^1_{mn}
Digital Encoder
$\square$
b^1_{mn}
Channel 1
$b^{1'}_{mn}$
Digital Decoder
$\diamond$
$d^{1'}_{mn}$
Side Decoder 1
$\sum_{m,n} d^{1'}_{mn} g^{1\#}_{mn}$
f_1
f
Central Decoder
$\sum_{m,n}(d^{1'}_{mn} g^{\tilde{1}\#}_{mn} + d^{2'}_{mn} g^{\tilde{2}\#}_{mn})$
$\tilde{f}$
Analog Encoder
$< \cdot, g^2_{mn} >$
c^2_{mn}
Quantizer
Q_Δ
d^2_{mn}
Digital Encoder
$\square$
b^2_{mn}
Channel 2
$b^{2'}_{mn}$
Digital Decoder
$\diamond$
$d^{2'}_{mn}$
Side Decoder 2
$\sum_{m,n} d^{2'}_{mn} g^{2\#}_{mn}$
f_2

Figure 1:

p/q	$A_{Riesz} \cdot 10^2$	$B_{Riesz} \cdot 10^2$	$A_{frame} \cdot 10^2$	$B_{frame} \cdot 10^2$	$\frac{A_{frame}}{B_{frame}}$	$e_{st,so}^{1+2;TSDED}$	$e_{st,opt}^{1+2;TSDED}$	ε
$7/6$	0.4309	8.5715	4.919	9.4805	1.92	3.9264	3.679	6.72%
$6/5$	0.5149	8.333	4.7907	9.6833	2.02	4.542	4.308	5.43%
$5/4$	0.6509	8	4.6195	10.207	2.21	5.396	5.218	3.41%
$4/3$	0.9002	7.5	4.397	10.981	2.50	6.644	6.492	2.34%
$7/5$	1.1139	7.1435	4.2834	11.365	2.65	7.533	7.3528	2.45%
$3/2$	1.444	6.669	4.2325	11.621	2.75	8.736	8.524	2.49%
$5/3$	1.983	6.01	4.4729	11.447	2.56	10.47	10.24	2.24%
$7/4$	2.233	5.7323	4.721	11.188	2.37	11.208	10.86	3.20%
$2/1$	2.827	5.136	5.653	10.271	1.82	13.254	12.984	2.08%

Table 1:

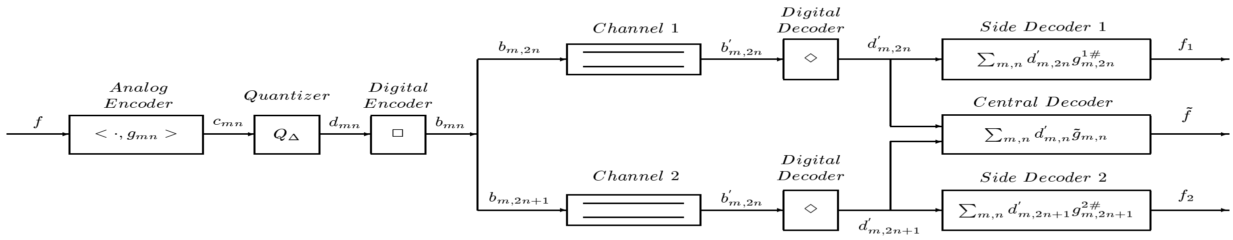


Figure 2:

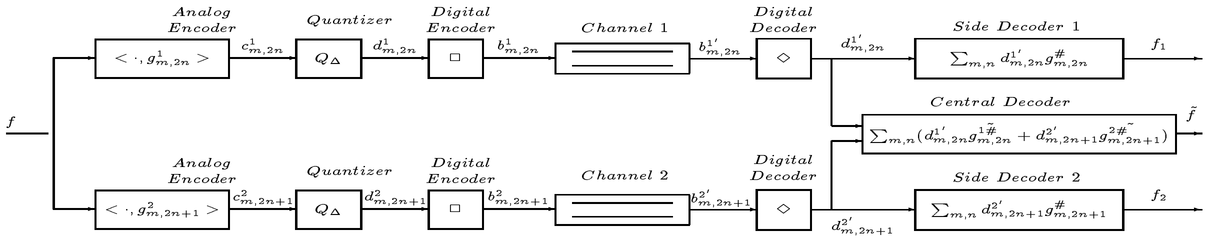


Figure 3:

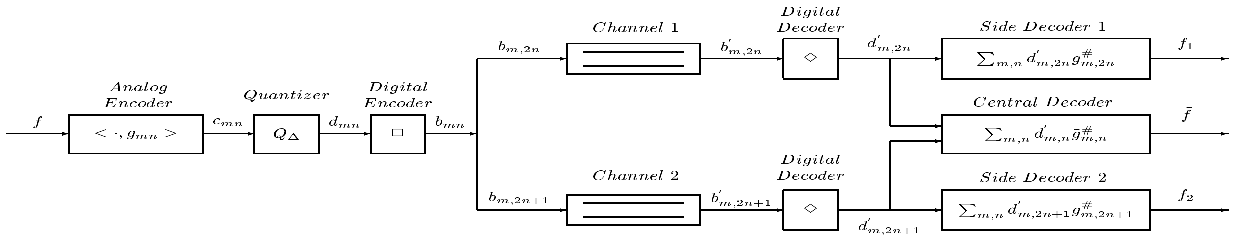


Figure 4:

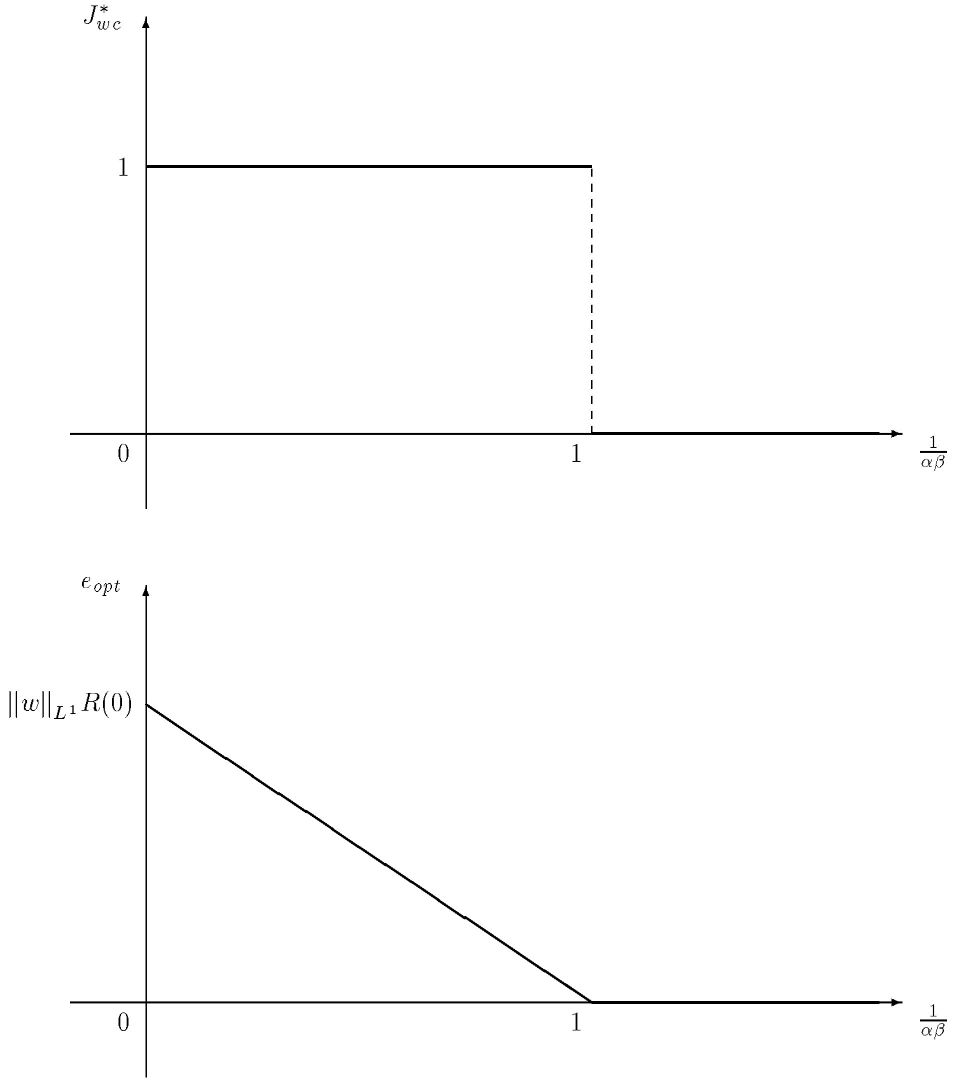


Figure 5:

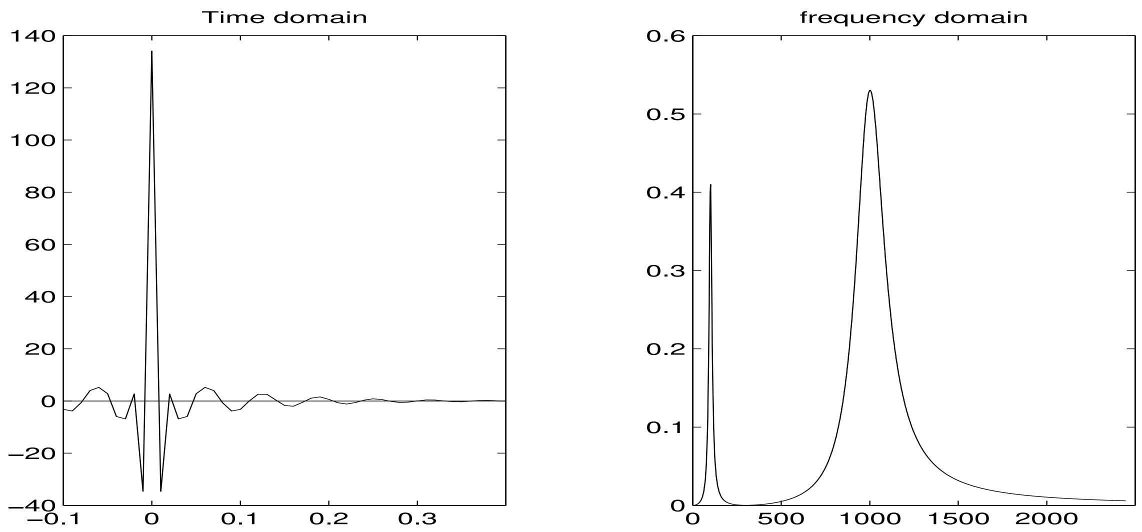


Figure 6:

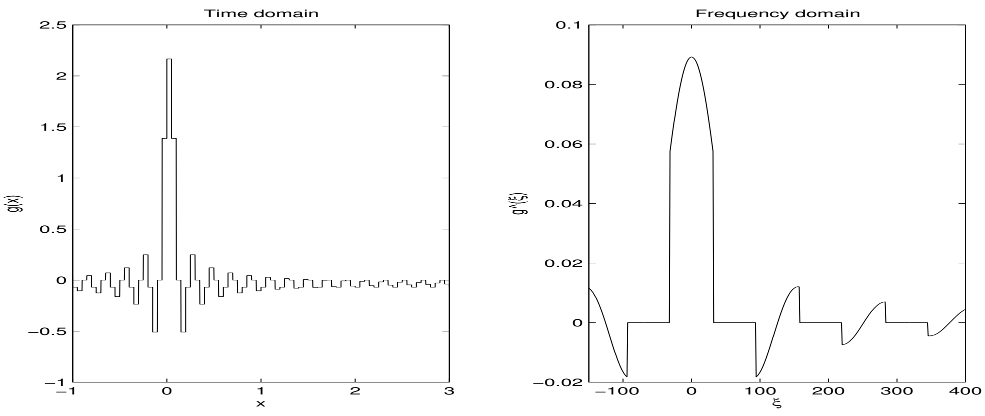


Figure 7:

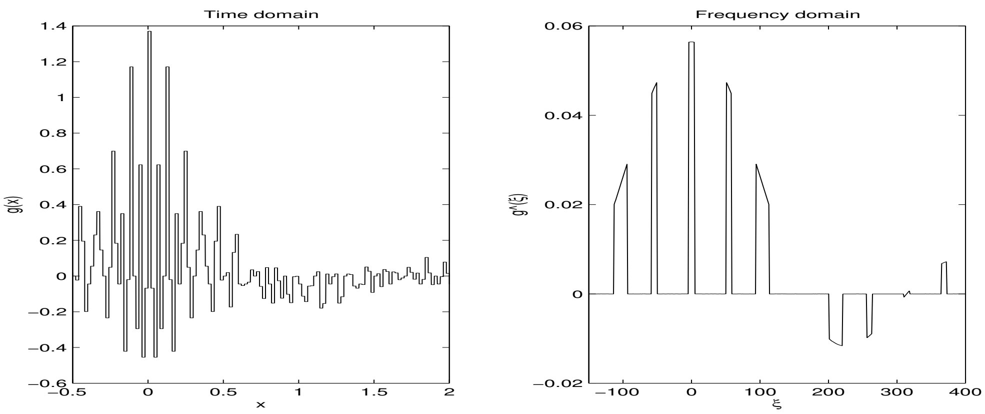


Figure 8:

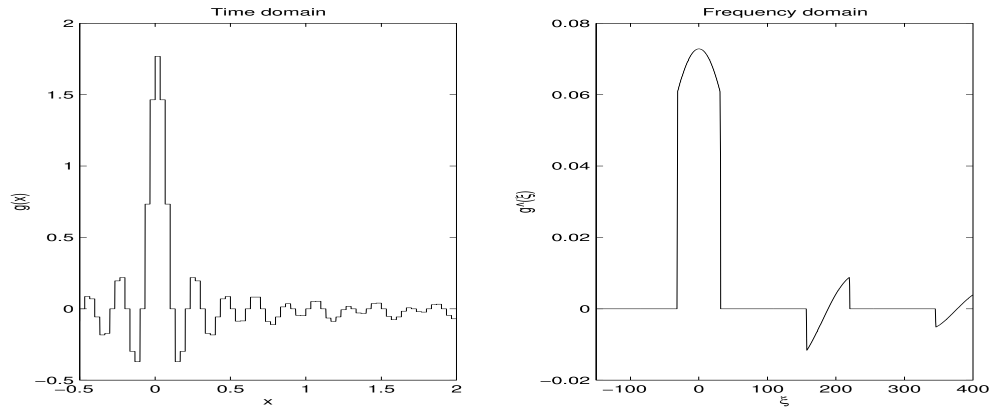


Figure 9:

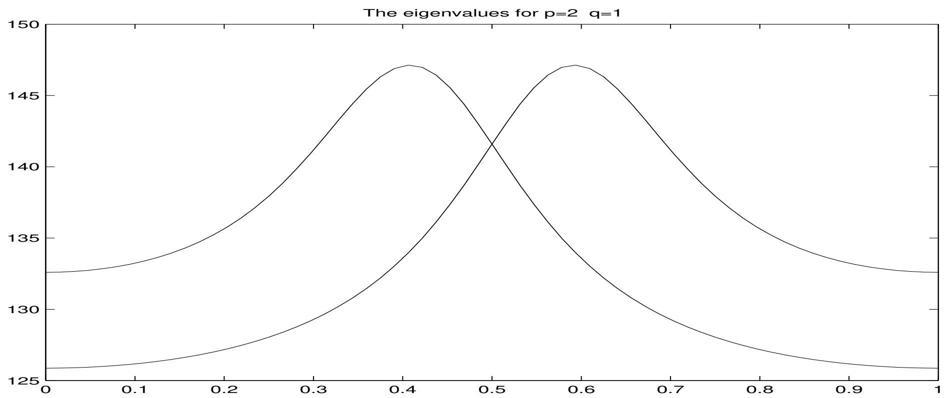


Figure 10:

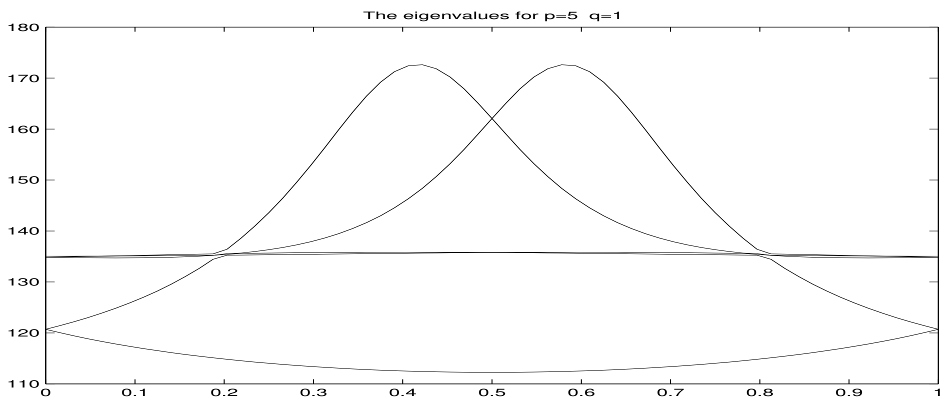


Figure 11:

p/q	$A_{Riesz} \cdot 10^2$	$B_{Riesz} \cdot 10^2$	$A_{frame} \cdot 10^2$	$B_{frame} \cdot 10^2$	$\frac{A_{frame}}{B_{frame}}$	$e^{1+2;TSDED}_{st,so}$	$e^{1+2;TSDED}_{st,opt}$	ε
$7/6$	8.5714	17.143	17.143	25.714	1.5	3.7534	3.679	2.02%
$6/5$	8.333	16.667	16.667	25	1.5	4.406	4.308	2.27%

Table 2:

		Gaussian window			Step window		
p/q	α	M_Δ^q	M_Δ	$Rate_{gauss}$	M_Δ^q	M_Δ	$Rate_{step}$
$7/6$	11.66	96	16	119.7	102	17	115.4
$6/5$	12	80	16	117.6	85	17	111.6
$5/4$	12.5	64	16	113.2	68	17	107.3
$4/3$	13.33	42	14	108.3	37	12.3	97
$7/5$	14	70	14	103.4	60	12	93
$3/2$	15	24	12	98	20	10	85.4
$5/3$	16.66	36	12	90.4	30	10	78.6
$7/4$	17.5	40	10	84	40	10	73
$2/1$	20	10	10	73	9	9	64.46

Table 3:

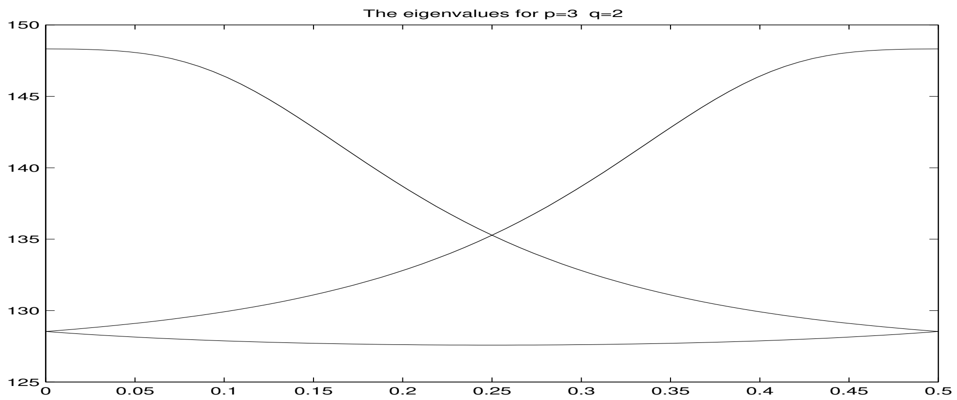

The eigenvalues for p=3 q=2
150
145
140
135
130
125
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Figure 12:

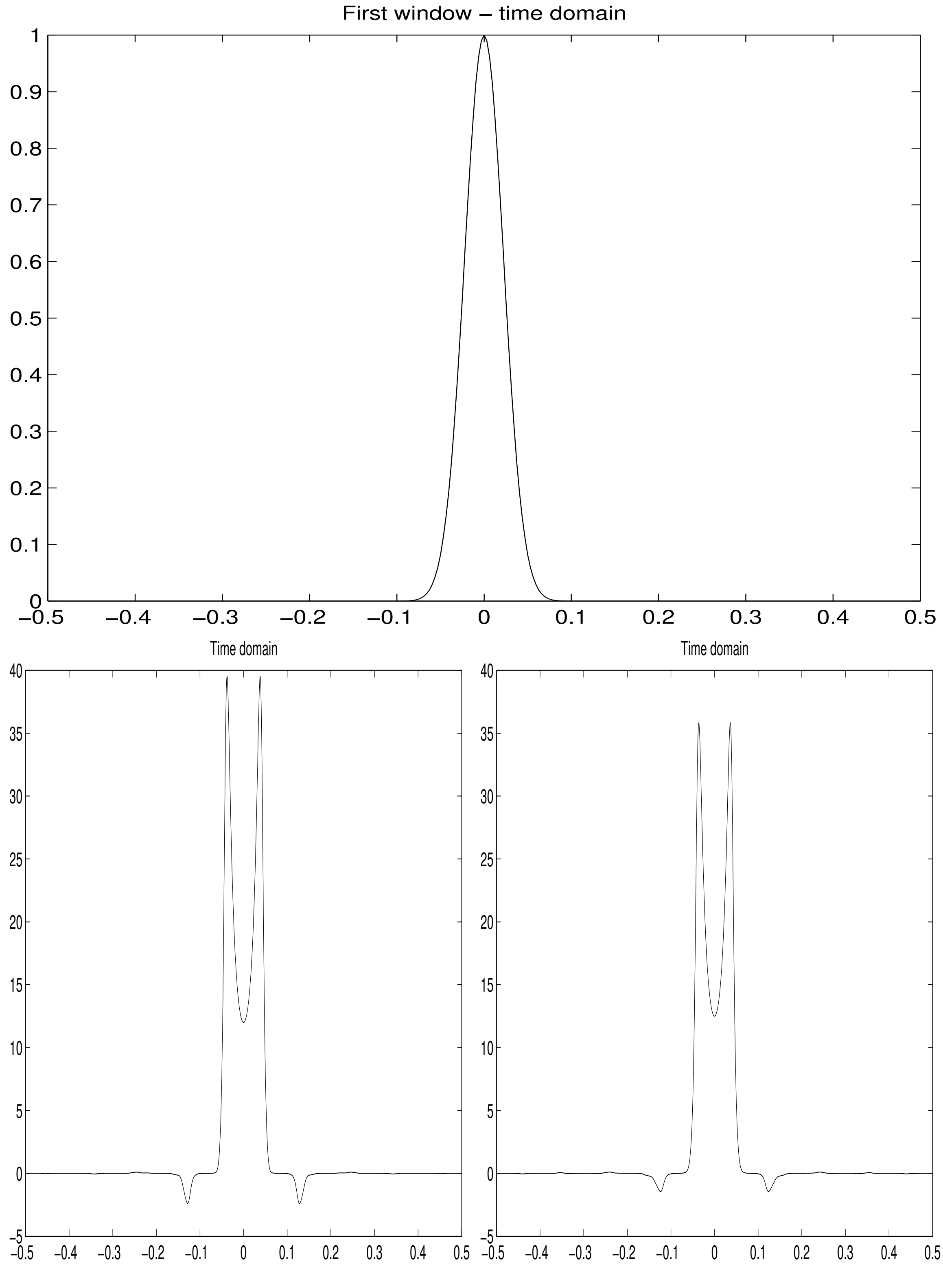


Figure 13:

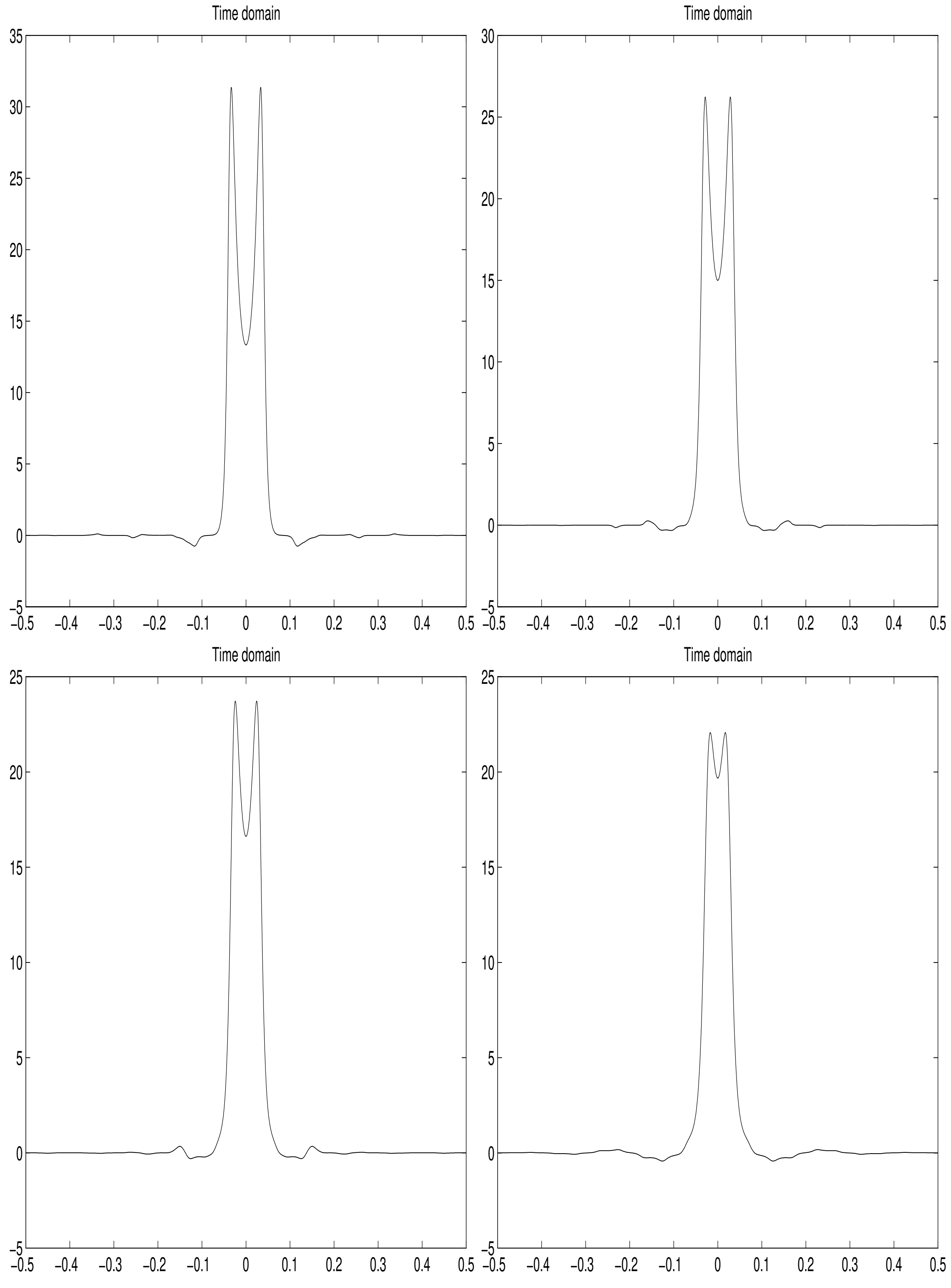

Time domain
Time domain
Time domain
Time domain
35
30
25
20
15
10
5
0
−5
30
25
20
15
10
5
0
−5
25
20
15
10
5
0
−5
−0.5 −0.4 −0.3 −0.2 −0.1 0 0.1 0.2 0.3 0.4 0.5

Figure 14:

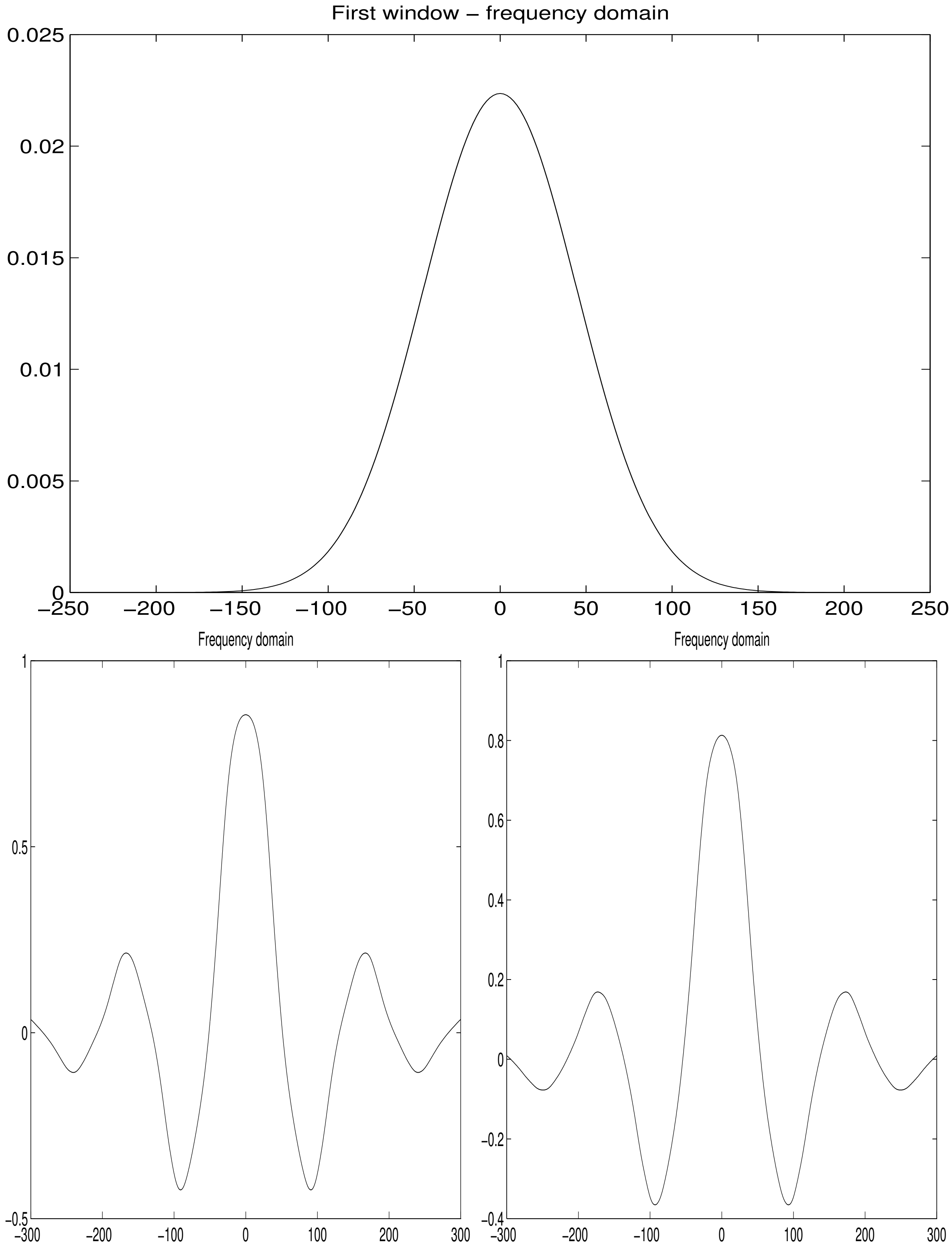

First window – frequency domain
Frequency domain
Frequency domain

Figure 15:

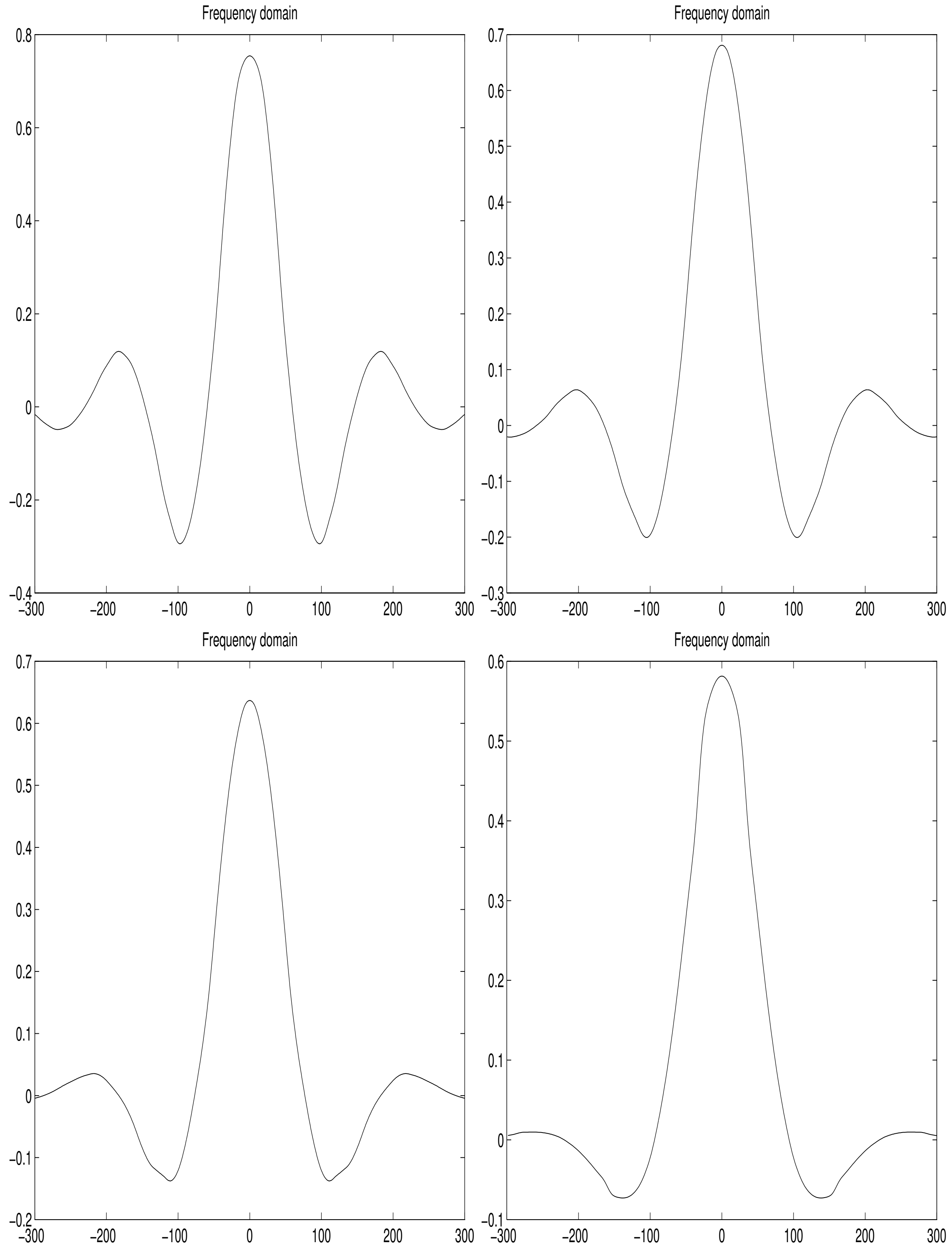
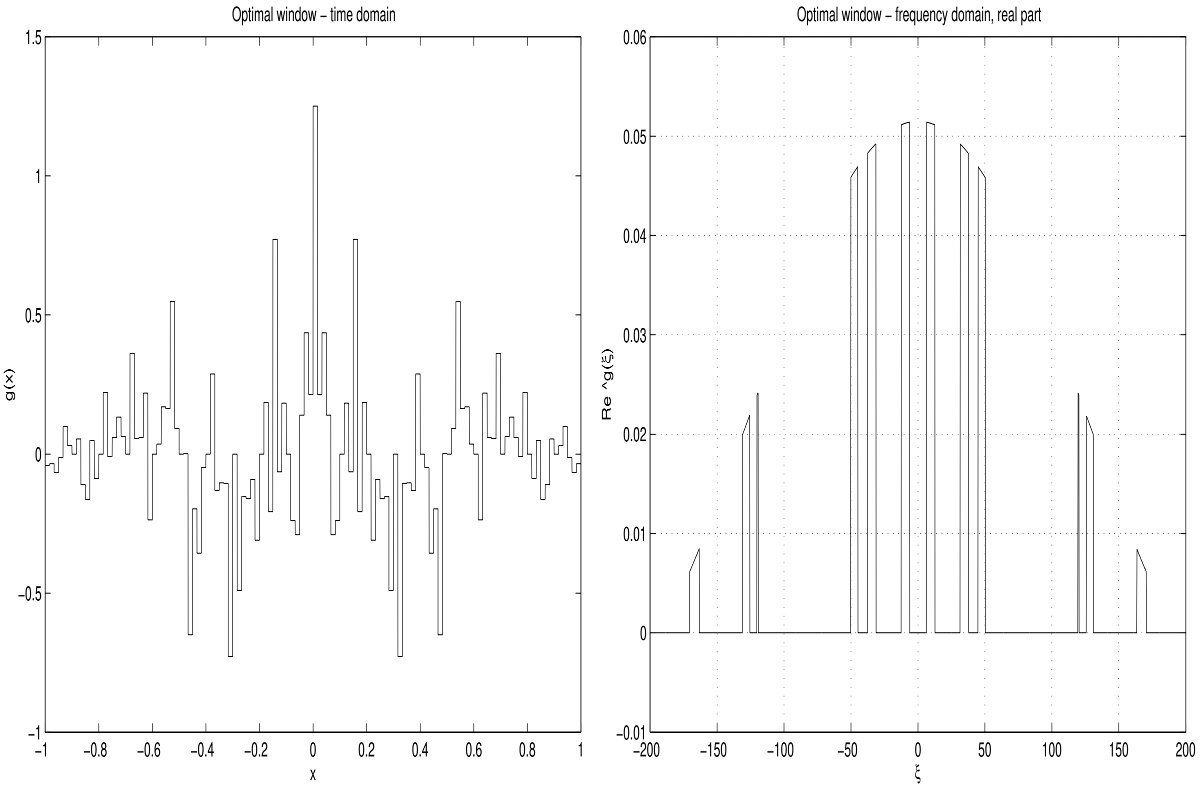
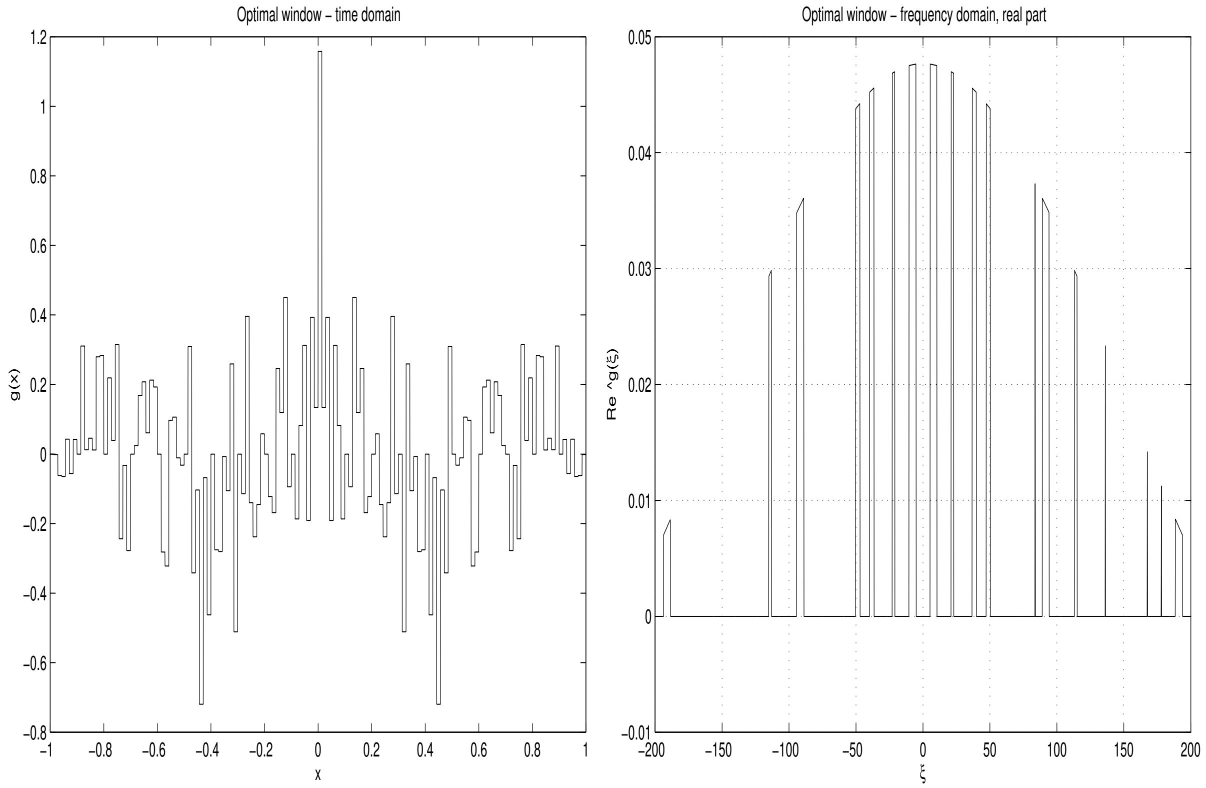

Frequency domain
Frequency domain
Frequency domain
Frequency domain
0.8
0.6
0.4
0.2
0
−0.2
−0.4
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−300
−200
−100
0
100
200
300
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1

Figure 16:



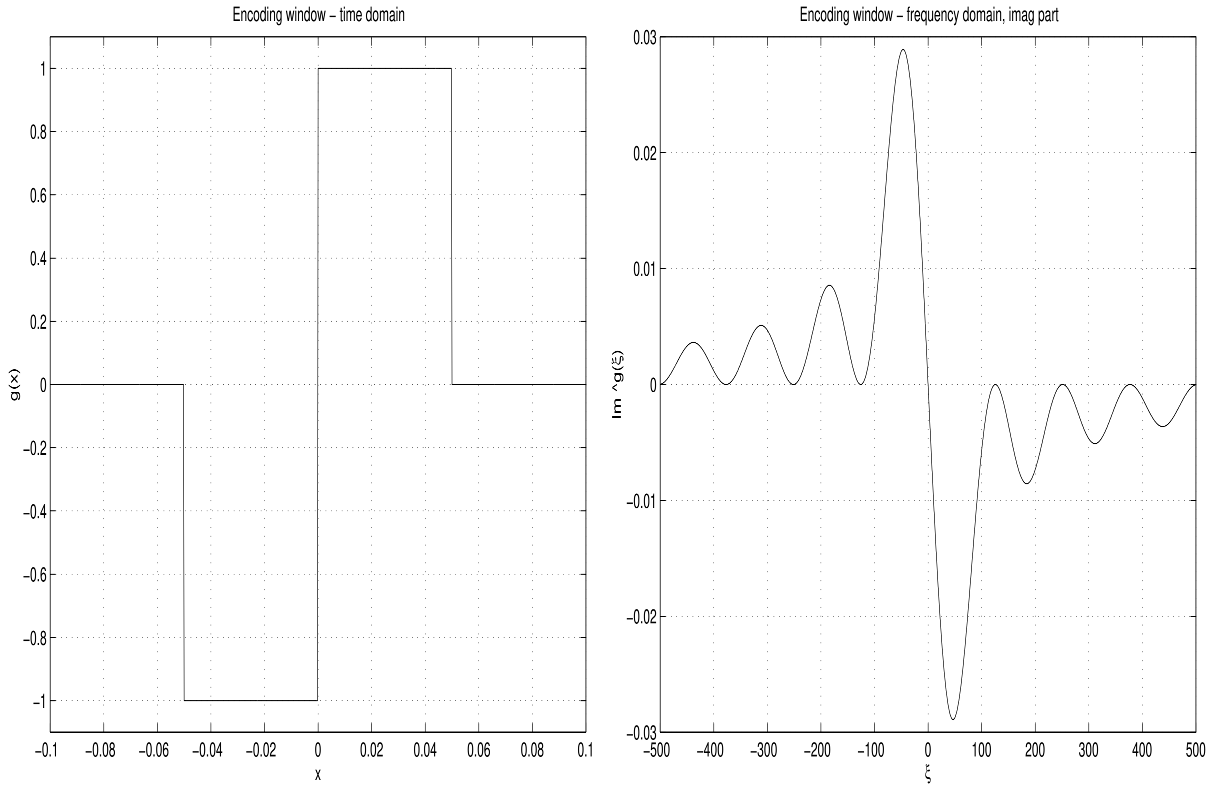
Optimal window – time domain
g(x)
x
Optimal window – frequency domain, real part
Re ^g(ξ)
ξ

Figure 17:



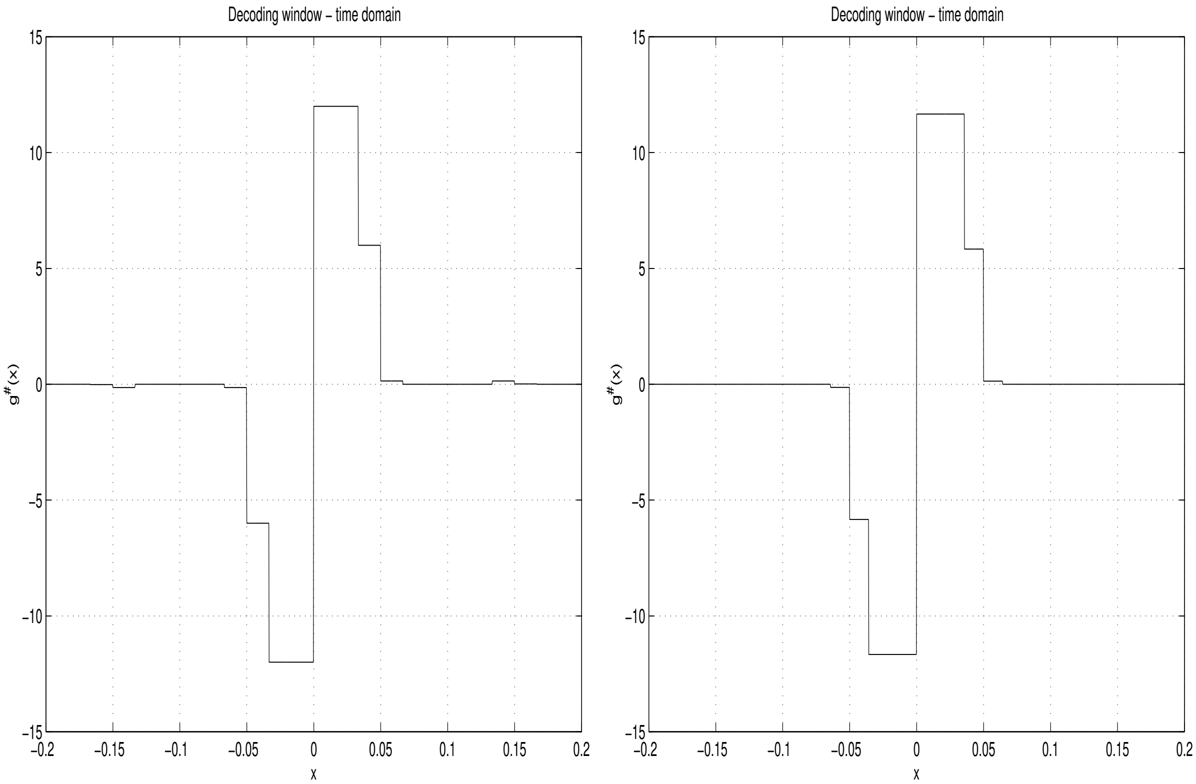
Optimal window – time domain
g(x)
x
Optimal window – frequency domain, real part
Re ^g(ξ)
ξ

Figure 18:



Encoding window – time domain
g(x)
x
Encoding window – frequency domain, imag part
Im ^g(ξ)
ξ

Figure 19:



Decoding window – time domain
15
10
5
0
–5
–10
–15
$g^{\#}(x)$
–0.2
–0.15
–0.1
–0.05
0
0.05
0.1
0.15
0.2
x
Decoding window – time domain
15
10
5
0
–5
–10
–15
$g^{\#}(x)$
–0.2
–0.15
–0.1
–0.05
0
0.05
0.1
0.15
0.2
x

Figure 20:

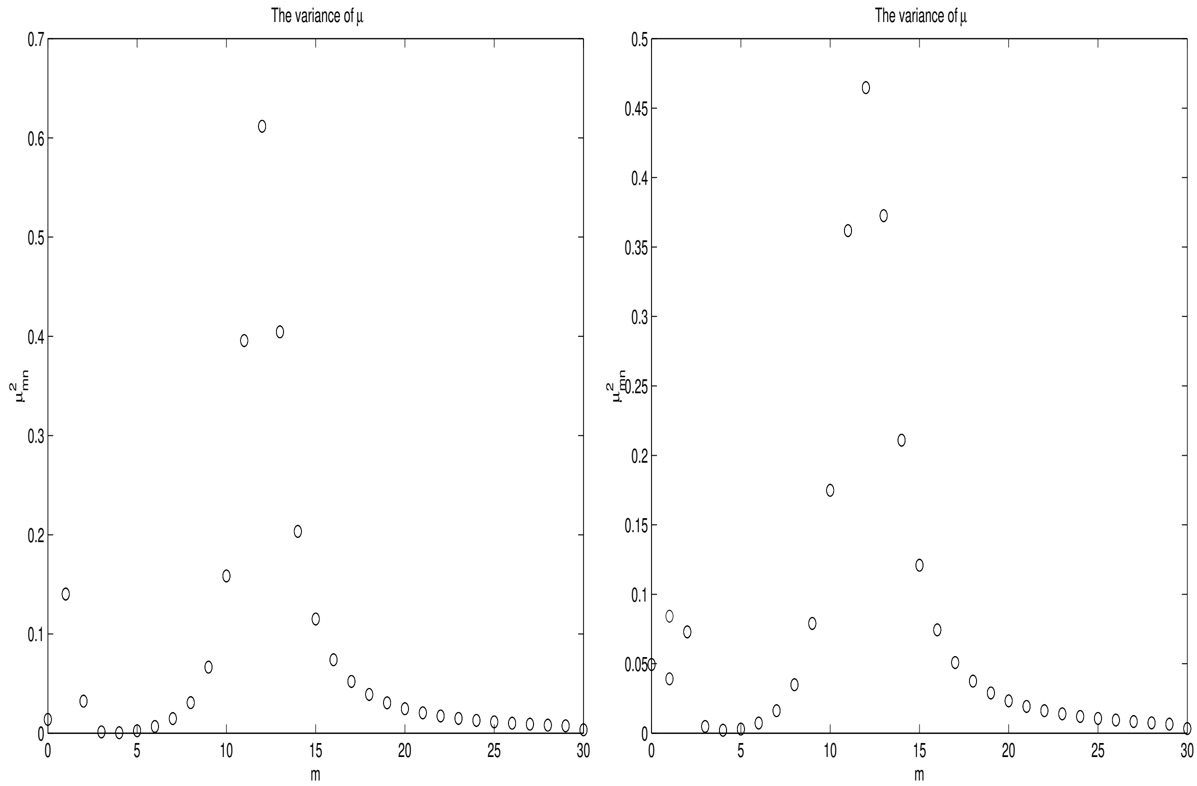

The variance of μ
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
μ^2_{mn}
0
5
10
15
20
25
30
m
The variance of μ
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
μ^2_{mn}
0
5
10
15
20
25
30
m

Figure 21:

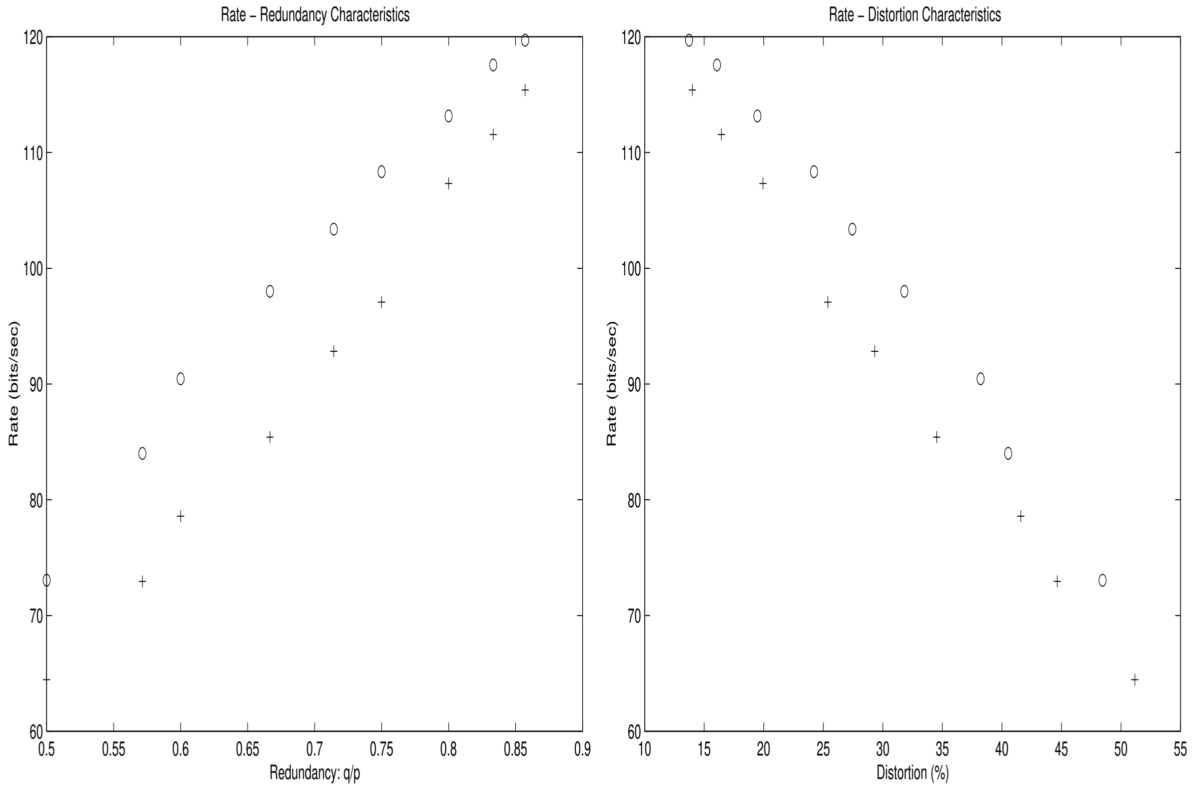

Rate – Redundancy Characteristics
Rate (bits/sec)
Redundancy: q/p
Rate – Distortion Characteristics
Rate (bits/sec)
Distortion (%)

Figure 22: